PEDL Research Papers

This research was partly or entirely supported by funding from the research initiative Private Enterprise Development in Low-Income Countries (PEDL), a Department for International Development funded programme run by the Centre for Economic Policy Research (CEPR).

Searching for Trade Partners in Developing Countries: Testing Firms in the 'Fast Fashion' Industry.*

Julia Cajal-Grossi

Department of Economics, The Graduate Institute

September 2016

First Version: February 2015

Abstract

An integral part of global supply chains is the selection by international buyers of trading partners in developing countries. However, our understanding of how buyers find a suitable long term supplier is limited. I use unique buyer-seller customs data to directly observe experimentation activity in a large market - the "fast fashion" industry in Bangladesh. I study how buyers of ready-made garments conduct trials of suppliers at the order-product level before settling into sustained sourcing relationships. To illustrate this process, I use a model of idiosyncratic search costs where the buyer's costs of testing a manufacturer are determined by the heterogeneity of potential suppliers. The model shows that (1) higher supplier heterogeneity is associated with lower experimentation, (2) as heterogeneity increases, search activity falls more markedly for larger buyers than for their smaller counterparts, and (3) while buyer-seller matches are positively assortative, more heterogeneous settings see all buyers -and more markedly, large buyers- willing to accept relationships with (weakly) worse suppliers. These implications are strongly supported by the data, and hold in terms of within-buyer, cross-market differences in experimentation behavior. Finally I show that these information frictions, rooted in supplier heterogeneity, matter for the distribution of rents in these relationships: price-cost margins for suppliers are positively related to the degree of heterogeneity in the environment.

*This paper has benefited from valuable feedback from Chris Woodruff, Guillermo Noguera, Mirko Draca, Greg Crawford, Rocco Macchiavello, Carlos Noton, Kala Krishna, Jim Tybout, Ken Teshima, Robert Akerlof, Jan Eeckhout Wiji Arulampalam and Nicola Pavanini. All errors are mine. Financial support from the Department of Economics at the University of Warwick and PEDL/Exploratory Grant is gratefully acknowledged. Part of the research work in this paper was carried out during a visiting period in Penn State University.

"When assessing a potential new partner [...] our commercial sourcing teams often start with a compliance screening process, supported by dedicated tools. If this first assessment is positive, our auditors conduct an in-depth head audit to make the final decision as to whether a supplier [...] has the potential for further performance improvements in order to become a long-term partner of H&M." [1]

1 Introduction

Trade in many markets involves a process of information acquisition, by the buyer, the seller or both. This is particularly true in international transactions, where foreign buyers purchase goods in unfamiliar environments, or where manufacturers need to introduce their products in markets with unknown demand structures. These information frictions have been shown to explain salient empirical regularities in trade flows and price dispersion.[2]

Evidence suggests that, in developing economies, those frictions take a characteristic shape and pervasiveness. The information the buyer needs to acquire is, usually, not readily available. While in many consumers' search problems the buyer is after price quotations, the search for a suitable partner in a developing country often entails screening the reliability, trustworthiness and general quality of the supplier. This in turn has two immediate implications. First, acquiring the relevant information involves costly experimentation or testing of some kind.[3] Second, the cost of undertaking such testing itself depends on the heterogeneity of the pool of potential trading partners. While obtaining a price quotation from a high-price manufacturer can be assumed as costly as acquiring this information from a low-price manufacturer, I will argue that experimenting with an unreliable supplier is likely to involve higher costs than doing so with "better" manufacturers.

This paper uses matched exporter-importer customs data to *directly observe* experimentation behavior and test the micro-level implications of a model of search that encompasses those specificities of developing countries. I consider a setup in which a heterogeneous pool of alternative trading partners is available in a developing economy. Buyers need to search through them to optimally choose a seller. The supplier's overall quality and reliability is unobservable to the buyer, who then invests resources in experimenting and testing. When the unobserved characteristic of the manufacturers is very dispersed (so suppliers are very heterogeneous), buyers may increase experimentation for two reasons. First, they might be "drawing" very low types of manufacturers, rejecting them, and moving onto subsequent search instances. Second, buyers might be searching for an exceptionally good supplier. In this sense, higher heterogeneity in the pool of available manufacturers can be positively associated with experimentation activity.[4] I argue that each search instance is costly for the buyer and

[1] Extracted from H&M's website, accessed on 20^{th} October 2015: http://sustainability.hm.com/en/sustainability /commitments/choose-and-reward-responsible-partners/about.html.

[2] In Aker (2010), Allen (2014), Goyal (2010) and Eaton et al. (2014), among others, the sellers or exporters lack full information on the demand for their products. In this context, they search and acquire information on prices and the appeal for their products in foreign environments. Related theoretical work includes the models proposed in Albornoz et al. (2012), Eaton et al. (2008), Allen (2014), Goyal (1999) and Chaney (2011), all focused on manufacturers learning about the demand for their products, their productivity in exporting activities, costs or other payoff-relevant conditions in foreign markets. In Monarch (2014), Alessandria (2009), Drozd and Nosal (2012) and Rauch and Watson (2003), uncertainty lies on the buyer's side, and the buyer needs to gather information either on the prices offered by available manufacturers or on the sellers' ability or quality.

[3] See Beil (2010) on stages of supplier selection.

[4] In their related framework, Rauch and Watson (2003) show that when the distribution of unobservables is such that the alternatives become "riskier" (in the sense of First Order Stochastic Dominance), the probability of testing or experimenting with suppliers - instead of starting a relationship - goes up.

that a more heterogeneous pool of suppliers can increase the expected cost of undertaking a search. The overall effect of suppliers' heterogeneity on buyers' experimentation behavior depends, then, on the interaction of these two forces: the direct response of search effort to increased dispersion, and its reaction to a change in the distribution of experimentation costs. In addition, buyers of different sizes can face different costs of searching the pool, conditional on the distribution of sellers' qualities. This heterogeneity across buyers induces different responses to dispersion in the environment.

To characterize this process, I exploit a detailed dataset with customs records of all garment-related shipments into and out of Bangladesh over the period 2005-2012. We observe the identity of players at either end of each export transaction, its volume, value and characteristics, the order to which each shipment belongs and the imported inputs required for its manufacturing. This dataset allows for a characterization of buyer-seller relationships from an unprecedented angle and level of disaggregation.

The setting, with regard to the industry and country, is well suited for the purposes of this study. The dimension of the sector, relative to the country's economy, and its potential as a driving force for development make the garment industry in Bangladesh worthy of study in its own right. Moreover, Bangladesh is possibly one of the leading examples of export-led takeoffs of low-income countries in the last couple of decades. The garment sector has grown 300% from 1990 to 2013, multiplying its workforce by a factor of 10 over that period, and turning the country into the second-largest garment exporter in the world.

Focusing on the main woven categories, we observe in our panel approximately $2,000$ international buyers every year placing garment orders to local manufacturers. Of these buyers, the top 10% in the distribution of sizes explains more than 80% of the traded volumes.[5] Importers of garments from Bangladesh need to select their trading partners carefully: fast-fashion markets are particularly sensitive to lead times, quality standards and delivery deadlines. Moreover, large international retailers have repeatedly been subject to media reports accusing them of having dealt with local factories with deficient social compliance practices, and poor health and safety standards. In this context, choosing a reliable, high-quality, long run trading partner is a recognized bottleneck in the industry.[6]

The formation and survival of buyer-seller relationships follow different patterns across product categories in our data. For example, in the Silk Female Blouses segment, relationships are almost 50% more persistent than those observed in Industrial/Occupational Male Cotton Ensembles. The probability a buyer allocates an order to one of its existing suppliers - rather than experimenting with a new manufacturer - is 0.74 in the first product category and 0.48 in the second one.[7] This empirical observation is reasonably linked to product-specific characteristics. However, we find that beyond product specificities, the dispersion in quality across the available manufacturers is strongly associated to the linking behavior in the data.[8]

[5]These larger players comprise a host of international brands. Among the renowned names, brands like Inditex, H&M, Marks & Spencer, Gap and Benetton are some of the larger players in Bangladesh.

[6]The larger buyers have fully dedicated offices located in Bangladesh for the purpose of selecting and monitoring suppliers. H&M's office employs approximately 200 people.

[7]The number of available suppliers and the ratio of buyers to sellers in these two categories are surprisingly similar and stable over the whole of the panel.

[8]This observation is somehow comparable to existing evidence at a higher aggregation level. In across-products comparisons, a well-established set of empirical facts corroborates the existence of higher search barriers to trade in more differentiated product categories (Goyal (1999), Feenstra and Hanson (2004)). Moreover, following Rauch and Watson's framework, Besedes and Prusa (2006) establish that highly differentiated product categories are likely to require higher search costs and stronger supplier-specific investments. They then proceed to show that the differentiation of the products traded in a relationship is positively associated with the duration of the relationship, and negatively

This evidence serves as the starting point of the central idea in this paper: search and matching rules depend crucially on how heterogeneous (how dispersed) the available suppliers are, through two channels - directly, via the distribution of manufacturers' unobservable characteristic and, indirectly, via the distribution of experimentation costs.

In the last decade, the availability of relationship- or transaction-level data has contributed to the proliferation of empirical studies on search in buyer-seller relationships. Besedes (2008), introducing search costs in a trade model, shows that the uncertainty present at the start of relationships -at the country-product level- accounts for the start of multiple short lived links, while long lasting trade partnerships are connected to higher reliability and low search costs. Using data detailing the identities of firms in Colombian-U.S. bilateral trade, Eaton et al. (2014) estimate a search-and-matching model where suppliers learn about the appeal of their product in a foreign market. They show the existence of significant costs (to the seller) of sustaining active searches for trade partners abroad, decreasing in the number of relations the supplier manages to establish. They also document high variability across suppliers in the success rates of the matching process. Highlighting the importance of long-lasting *B2B* relationships and informational frictions, Drozd and Nosal (2012) study a setup in which exporters need to match with consumers to build market shares in each destination, and they do so by investing in marketing (as a form of undirected search). Alessandria (2009) offers a model in which consumers search for price quotations in markets, incurring heterogeneous costs and obtaining different "amounts" of information. As a result, exporters price-discriminate across destinations. Exploiting data on trade flows between farmers in the Philippines, Allen (2014) quantifies the size of information frictions using trade flows from exporters unaware of prices in alternative destination markets. Using disaggregated customs data, Monarch (2014) finds that in the presence of search costs, international buyers tend to keep their Chinese suppliers. Switching away from them in the lookout for cheaper manufacturers is hindered by such costs, inducing persistency in relations and forcing switches to manufacturers within a small distance to the current base of suppliers. Also with a search perspective, Benguria (2014) uses French - Colombian data in a general equilibrium model to prove the existence of positive assortativeness in productivity in buyer - seller relations. Dragusanu (2014) studies the sorting patterns between U.S. importers and Indian exporters, under the light of a model of search where buyers seek suppliers to undertake tasks in a production chain.

With a specific focus in development, Egan and Mody (1992) and Tewari (1999) show that U.S. and European buyers start trading with suppliers in developing countries by allocating small orders, in the face of the buyer's uncertainty on the ability of the manufacturer to meet the required standards. Rauch and Watson (2003) base their model on this observation.[9] Outside of the trade literature, a growing line of work shows copious evidence on the uncertainty around the reliability, quality and profitability of matches with firms in developing countries. Examples of these are Macchiavello and Morjaria (2010) in the context of the Kenyan flowers' market, Macchiavello (2010) on the Chilean wine-for-export industry, McMillan and Woodruff (1999) focused in trade credit in Vietnam and Banerjee and Duflo (2000) on software contractors in India, among others.

This paper builds on those two strands of literature to offer four contributions. First, unlike previous studies that have exploited aggregated trade flows and price distributions to infer the role (and mag-

related to the size of the initial trade. Using Chilean firm-level data, Alvarez and Lopez (2008) show that entry into and exit from exporting activities are -respectively- negatively and positively correlated to the underlying heterogeneity across manufacturers, in terms of total factor productivity.

[9]Somewhat connected to these contributions, there is a growing stream of literature on intermediation in trade and uncertainty, within which Ahn et al. (2011) has closer connections to this paper.

nitude) of information frictions, this paper uses the direct observation of experimentation activity in the trade context. Using very rich data, I uncover new patterns in micro aspects of buyer-seller relationships in developing countries. Second, following a standard definition of suppliers' heterogeneity, I "import" an established method used in the labor literature to exploit the higher disaggregation in the data, and obtain scalar measures of suppliers' quality. This is, to my knowledge, the first application of such methods to matched exporter-importer data. Third, the paper studies buyers' experimentation with suppliers in developing countries, endogenizing the costs of search to the distribution of manufacturers available to the buyer. This mechanism offers a novel explanation for the observed negative relation between exploration and heterogeneity in the search environment. Fourth, the paper links changes in the heterogeneity across manufacturers to two market outcomes that are key to the development process: the sorting patterns between buyers and sellers and the determination of price-cost margins. In particular, dispersion in the quality of available suppliers can compromise the assortativeness of relationships, and it can affect the distribution of the gains from trade between large international buyers and their suppliers in low income countries.

Section 2 describes the main features of our dataset and the institutional environment relevant to this empirical study. In Section 3, I present a set of patterns in the data describing the choice mechanisms behind the buyers' allocation of orders to alternative suppliers. First, buyers test suppliers by allocating small orders. On average, it takes buyers three testing instances with unknown manufacturers before establishing a sustained relationship with a supplier. The number of suppliers tested before settling down into a relationship increases with the size of the buyer. Second, I describe the use of a two-way fixed effects regression to identify quality as a demand shifter for each supplier, conditional on prices and other controls. This is conceptually in line with Hottman et al. (2014) and it reproduces technically the exercise in the literature in labor economics, following Abowd et al. (1999). Using these measures, I observe that the heterogeneity in the quality of suppliers varies substantially across product markets, and that the more heterogeneous segments are those that are typically more fashion sensitive. Fourth, I present evidence suggesting that the propensity of the buyers to allocate orders to already known suppliers is higher in more heterogeneous segments. Finally, the data show a higher probability of trading with unknown suppliers when the average quality in the pool of manufacturers known to the buyer from past trading experience is poor.

Overall, the empirical regularities presented in Section 3 are consistent with a setup in which buyers need to test manufacturers to find a suitable partner, facing a cost of searching that is increasing in the heterogeneity across qualities of potential partners. I illustrate this idea in Section 4, presenting a search model of buyer-specific costs in the spirit of Wolinsky (1986), and following more recent contributions by Moraga-Gonzalez et al. (2014), Wilson (2012) and Janssen et al. (2005). In this framework, I introduce a specific search mechanism that links the distribution of types of suppliers and the costs of searching. From this formulation, I derive four testable implications. Contrary to the observation that higher heterogeneity in the environment intensifies search efforts, the model shows that (1) higher dispersion in the types of suppliers can be associated with lower overall experimentation. Moreover, (2) the effect of suppliers' heterogeneity in search behavior depends on the buyer's size: for larger buyers, experimentation is lower in more heterogeneous environments. For buyers with large search costs (small buyers), higher heterogeneity might imply higher or lower search activity, depending on whether the dispersion or cost effects prevail. Also, (3) in more heterogeneous environments buyers are willing to match with (weakly) lower suppliers. For small buyers, with high experimentation costs, the effect of dispersion on the threshold supplier to accept is negligible, while (4) for large buyers, small changes in the underlying dispersion bring the threshold supplier down significantly.

Section 5 is devoted to the empirical scrutiny of these predictions, which find strong and robust support in the data. In Section 6 I explore further effects of the heterogeneity mechanism on two relevant market outcomes: the 'extensive margin' of search behavior and the price-cost margins suppliers can charge for their orders. Section 7 revisits the main implications of these empirical findings.

2 The Empirical Context

2.1 The Ready Made Garment Sector in Bangladesh

Bangladesh is one of the salient examples of low income countries that have a pillar industry driving, via exports, the country's growth. With the garment industry accounting for more than 10% of the country's GDP, the sector has pushed annual per-capita income from USD280 in 1990 to USD838 in 2013, an increase of almost 300%.[10] Expanding international demand for garments, wage levels that are among the lowest in the region, and a very elastic labor supply have spurred the growth of the sector over the past two decades to a degree unprecedented for the country. In these decades, the country's exports of ready-made garments have grown by more than $2,000\%$ to about USD20 billion. From its take-off point in 1990, this has implied an average annual growth of some 16%, with the garments' share in export volumes expanding from 50% in 1990 to 83% in 2012. According to industry figures, before the nearly 70% increase in the minimum wage in 2013, the hourly wage for unskilled industrial workers in Bangladesh averaged USD0.24, far below comparable countries in the region. By comparison, Cambodia pays USD0.45 and Vietnam pays USD0.53.[11]

The expansion of the sector has translated into an increase in the number of garment factories, which grew from 830 in 1990 to $5,600$ in 2013, and a sharp increase in garment plant employment, which grew from 0.4 million workers to 4 million workers over the same period (BGMEA). This figure represents more than 45% of the employment in the industrial sector and it is mainly formed by young female workers, in a large proportion migrant from rural areas.

Together, the United States and European Union account for 84% of the country's garment exports. This demand is highly concentrated among a couple hundred international buyers, comprising both large non-specialized, mass retailers, like Walmart, Tesco, Kohl's, etc., and clothing retailers of varying qualities, such as H&M, GAP and Polo Ralph Lauren. Overall, the top 10% of buyers in the distribution of sizes account for more than 80% of the traded volumes every year. In our panel, 0.2% of the buyers account for approximately 40 percent of all the demand for woven products. The average size of these very large buyers, in terms of quarterly volumes (this is, aggregating all export orders in a quarter), is more than 50 times the average volumes of all other buyers. The top 10% buyers allocate about 50 exports orders per quarter; individually, these orders are twice as large as those allocated by smaller players, who on average place three orders per quarter.

While companies in Europe and the United States can leverage Bangladesh's cheap production costs to source ready-made garments, managing the supply chain in the country has been recognized as the

[10] Figures corresponding to 2014.

[11] With an eight-hour day baseline, for 25 productive days, the Bangladesh hourly wage results in a monthly wage of USD48, slightly higher than the statutory minimum USD38.

main bottleneck in their sourcing strategy. Contracts between buyers and sellers in export markets are often incomplete and ensuring the quality and timely delivery of orders tends to be a major concern for international buyers (see Monarch (2014), for example). The underlying uncertainty is usually connected to the quality of the goods, the reliability of the seller (in terms of lead times, for example) and its productivity in a broader sense. While some of these can successfully be tested and assessed within the course of a trading relationship, ex-ante, buyers do not have full information about their suppliers. The garment sector in Bangladesh is infamous for its lack of compliance with minimum health and safety requirements and human rights, even when firms hold all the necessary credentials. Governmental and official controls for these are known to be weak, and episodes of extensive coverage in the media have shown the difficulties buyers face, even after engaging in costly screening processes, to identify suppliers that might secretly break their compliance agreements. Table 8 summarizes the most salient media controversies of this nature since 2006.[12]

2.2 The Data

The empirical analysis in this paper exploits a comprehensive dataset recording all export transactions between ready-made garment (RMG) manufacturers in Bangladesh and buyers in the rest of the world. The primary source of this dataset is the compilation of mandatory export and import records in the main custom stations in Bangladesh, from 2005 through 2012. Each record constitutes a product (Harmonized Codes disaggregated to the sixth digit) within a shipment from a supplier to a buyer, taking place on a given date. The *real-time* data include details on the statistical values, quantities, destinations and specifics of the terms of trade. Importantly, they include identifiers for all buyers and sellers.

Exports of garments in Bangladesh are split almost evenly between knitted and woven products. The focus here is in woven products, whose manufacturing technologies are known to be relatively homogeneous - across products and across firms - and whose main material input, woven fabric, is imported. Our source of primary data includes, in addition to exports, all the imports by RMG manufacturers into Bangladesh, with records as detailed as those in the export side of the data. Exploiting an administrative procedure used for claiming duty exceptions for inputs to garment export orders, we can match specific orders to the material inputs used for producing them. Because the RMG sector in Bangladesh is almost exclusively export oriented, exported volumes coincide with virtually the whole of the manufacturers' supply. Therefore, we can claim to observe the firms' output in its entirety and, for the sample here, the relevant material inputs as well. We are also able to group the transactions or shipments between buyers and sellers in their corresponding orders.[13]

[12]The tragic Rana Plaza factory collapse in Savar in April 2013 killed 1,100 people. The building housed factories that were producing orders for large buyers, including Primark, Benetton and Walmart. The media have reported several (smaller) cases in which manufacturers serving large buyers experienced explosions, fires, sexual harassment accusations, forced overtime work, illegal subcontracting, etc.. In 2006 alone, large buyers including Inditex-Zara, Carrefour, Kmart, H&M and PVH were involved in 14 episodes of these kinds. These episodes have proven costly for the buyers in that, first, they needed to put in place compensation schemes on occasion and, second, (and most importantly), they needed to deal with media reports potentially damaging to their reputation. Buyers have taken both individual and collective initiatives to minimize the depth and breadth of this source of social compliance uncertainty: more extensive monitoring, frequent evaluations, "surprise" fire drills, collaborative efforts on fire safety education (a consortium with 18 large buyers), a buyers' forum sharing information on "bad" suppliers (*black lists*) and different agreements with governmental and human rights organizations (PVH, for example).

[13]We can distinguish orders from isolated shipments, using information on the Export Procedures in our dataset. The difference between the two modes of trade is not merely administrative. Orders span over time, can entail multiple shipments, can involve multiple products, imply an ex-ante specification of quantities, input requirements and quality of materials and, notably, allow for import duty exemptions if fabric is imported for the purpose of fulfilling the order. Isolated shipments, on the other hand, do not entitle manufacturers to claim for import tax reimbursements and,

In the main woven categories, we observe approximately $7,000$ buyers operating at some point in the panel (Table 11). Approximately $2,000$ of them are active every year, importing an average of five woven products (HS6) supplied by three to four local manufacturers through eight or nine orders, on average. The top 10% of the buyers each year concentrate more than 80% of exports of garments from Bangladesh and trade annually with more than 13 sellers, allocating some 45 orders per year.

At the micro level trade flows are inherently lumpy, making the characterization of the evolution of importer-exporter relationships tricky. I will refer to relationships between suppliers and buyers as either *sustained* -or long lasting- or *scattered*. A relationship is sustained if, from the wake of the first trade between the two parties, exports involving the same buyer-seller pair are recorded at least in one product category once a year, for as long as the buyer has non-zero demand in the product categories in which the supplier is active.[14] All other relations are considered scattered and comprise one-off interactions as well as trade relationships in which the buyer places orders with the seller infrequently. This classification distinguishes those manufacturers that are core recurrent suppliers from those manufacturers that, having interacted with a buyer, receive no steady subsequent demand from that buyer.

According to this classification, two-thirds of the relationships identified in the panel are scattered and one-third of them are sustained. Reasonably in line with recent findings in the literature, sustained relationships account for more than two-thirds of the traded value every year. At the level of the supplier, of the almost $3,000$ manufacturers in the subsample, 9% never form a sustained relationship in the terms defined above. And this seems to have a strong impact on the growth profile of the exporter. The horizontal axis in Figure 1 indicates the share of the seller's exports that correspond to sustained relationships, through the first three years of export activity. The vertical axis depicts the end-to-end growth rate of the seller's exported values in the three-year period.[15] The polynomial fit in the graph shows that firms with higher engagement in sustained relationships exhibit significantly higher growth profiles.[16] In other words, manufacturers that manage to establish themselves as recurrent suppliers to their buyers expand more quickly than those whose exports are largely scattered.

Importantly, now looking at all the buyer-seller relationships available in our panel, relationships tend to grow over time in terms of traded volumes (Column (2) in Table 12). This trend is accompanied by an increase in the number of orders the buyer allocates to its supplier over the course of the relationship (Column (1)). At the same time, there is no evidence of a positive trend over time in terms of the size of the orders the buyer places with the manufacturer, in established relations (Column (3)). The panel regressions in Table 12 support this characterization.[17]

---

obviously, stand alone as isolated shipments. In our panel, 93% of the exported values in our panel correspond to orders.

[14]To allow for season displacements, I consider 14 months periods and refer to them, for the sake of brevity and simplicity, as a year. For the purpose of classifying relationships into sustained or scattered categories, I exclude from this analysis relationships that are younger than a year, and those that correspond to firms appearing for the first time within the last two years of the panel. Relationships whose terminations coincide with the permanent exit of the buyer from the product category primarily served by the supplier, or an equivalent shift in the specialization of the supplier, still constitute sustained relations if trade satisfies the periodicity requirement up until the break-up event. The classification is adjusted for censoring on both ends of the panel.

[15]This is, *(exports third year - exports first year)/exports first year*, such that 1 means that exports in the third year of activity of the firm are twice as large as those in the first year. On the horizontal axis, for instance 0.7 means that the 70% of the exports of the seller in the first three years of activity corresponded to (perhaps ex-post) sustained relationships.

[16]This holds true in the fourth-to-first, fifth-to-first and sixth-to-first year growth rates; but, given that the implementation of the sustained vs. scattered relationship substantially restricts the sample, the confidence intervals are wider as a result of the smaller sets of data points for firms surviving four or more years. The unconditional probability of a relationship in the panel surviving three years is 0.32 and falls to 0.27 and 0.21 for four and five years.

[17]The generic estimating equation (for all three outcomes as reported in the relevant Table) is $y_{ijt} = \alpha_{ij} + \gamma_a t +$

The panel then shows the evolution of buyer-seller relationships in which the seller benefits mostly when establishing a sustained relationship with the buyer. The attention in the rest of this paper is centered on understanding how buyers choose with whom to form relationships.

3 Patterns in the Data

When selecting a trade partner in a developing country, international buyers face potential suppliers of varying characteristics. While idiosyncrasies such as size, productive capacity, or product specificities are more-readily observable to the buyer, the reliability and trustworthiness of the supplier are often unobserved. This dimension is particularly important in contexts where unreliable suppliers might (secretly) renege of social compliance requirements, fail to deliver shipments on time, or force the buyer to engage in costly adjustments of designs, products or processes.

Below, I present evidence on four salient facts of the context I study. First, buyers search for suppliers by allocating small orders. On average, it takes buyers three testing instances with suppliers before they find a manufacturer with whom they establish a sustained relationship. The evidence below suggests that larger buyers search more intensely -experiment more- than their smaller counterparts. Second, suppliers are heterogeneous beyond observables. After we account for several dimensions of potential idiosyncrasies (inputs used, overall size, trade partners, price of output, etc.), trade patterns in our data show that manufacturers are different in a *quality* dimension. Moreover, the distribution of suppliers along this characteristic varies across product markets. Third, buyers' decisions regarding who to trade with internalize this heterogeneity in two ways: (i) buyers decide to allocate orders to unknown suppliers more often whenever the portfolio of their known suppliers is of low *quality*, and they tend to stick to their known partners when their pool of suppliers is good relative to the distribution of types of available manufacturers; and (ii) buyers' choices of suppliers vary depending on how heterogeneous the environment is. We observe finally that buyers are less willing to experiment with unknown suppliers when the underlying heterogeneity in the market is high. This is the case both when the buyer is already active in the market, and when the buyer is new to trade in the product category.

Fact 1. *Buyers search for suppliers and this search process has two main characteristics: (i) buyers "start small", allocating testing orders to suppliers; and (ii) large buyers search more intensely than small buyers.*

Within a buyer-seller relationship in a given product category, we can distinguish the order corresponding to the players' first interaction from all subsequent (non-first) orders. The group of first orders includes both the interactions that did not fructify into further trade between the buyer and the seller, and those first orders that were later followed by other orders in the relationship. Table 13 compares first and non-first orders and shows that the average size of first orders is less than 40% of the mean size of non-first orders.[18] Volumes are spread on three shipments in the first group and on average nine in the second group. The table reveals that prices of testing orders are slightly smaller than non-first orders.[19]

$\gamma_b t^2 + \iota_t + \epsilon_{ijt}$. The cross-sectional units are buyer-seller pairs interacting for at least four quarters in the panel, and the time dimension is given by quarters of effective interaction. The main regressors account for the existence of linear and quadratic time trends -t and t^2-, controlling for buyer-seller specific intercepts (α_{ij}) and seasonal corrections (ι_t).

[18] When the sample is restricted to the top 10% buyers, the relation between first and non-first orders is 1/20.

[19] This characterization holds also when comparisons are carried over between first orders and non-first orders, separately for each product category.

In the panel, there are approximately 18,000 instances of trade that can be categorized as the entry of a buyer into an HS6 product code.[20] Of these, just over 16,000 buyer-product combinations correspond to cases in which sooner or later the buyer establishes a long-lasting relationship with some supplier. Almost all - 99% - of these entries involve the buyer placing at least one testing order with a supplier different from the one with whom they ultimately form their first sustained relationship. On average, finding a long-lasting partner takes buyers (almost) three attempts; that is, first orders to three suppliers that don't evolve into further trade. On the top end of the distribution, the 90^{th} percentile corresponds to buyers going through seven or more short-lived interactions before settling down with a supplier.

The bottom panel in Table 13 compares the average price (per kilo of garment) paid by the buyer to its first established partner, and the price paid to suppliers that the (same) buyer discarded.[21] The last row in the table tests the difference between the price of the first order the buyer placed with its first sustained supplier, and the average price of first orders previously allocated to other suppliers. Both comparisons show that "rejected" suppliers offered first orders at lower prices. This suggests that low prices are not the only sorting criterion for buyers.

Figure 2 plots, via a polynomial regression, the relation between the number of "exploratory" orders entrant buyers in a product category place to suppliers before establishing their first long-lasting partnership, and the size of the buyer.[22] I will refer to the count of short-lived interactions the buyer goes through before forging a sustained relationship the *search intensity* or experimentation effort.[23] Both buyers' sizes and search intensities are nearly Pareto distributed, with a large mass in the low values and a thin top tail. Figure 2 shows that search intensity is increasing in the buyer's size: large buyers initiate a large number of one-off interactions before establishing a recurrent relationship with a supplier.

While the buyer's size seems to drive, at least to some extent, this observed search behavior, a large proportion of the variability in search intensity cannot be accounted for by buyer, product or time effects.

Fact 2. *Suppliers are heterogeneous and fashion-sensitive segments show higher dispersion in suppliers' heterogeneity.*

In the recent literature looking at firm-level trade patterns, firm idiosyncrasy has been associated with cost-related heterogeneity, quality, markups or the horizontal scope of the firm (Manova and Zhang (2012), Hottman et al. (2014), Bernard et al. (2012)).[24] Studies exploiting matched exporter-importer data have used different measures to act as the empirical counterpart to such heterogeneity, either focusing on productivity or on quality. Despite the richness of newly available custom datasets, direct measures of either of these have proved difficult to construct. Most papers have used traded volumes (or an alternative measure of size) as a proxy for the unobserved dimension differentiating manu-

[20] Accounting for 18 months of potential censoring at the start of the panel, I call "entry" the first shipment in the HS6 code that goes to the buyer. This might discard true entries happening in the first year and a half of the panel.

[21] For this comparison, the price for the established buyer is computed over all orders except the first one during the first year of the relationship.

[22] Size of buyer calculated as the log of its imported volumes in the HS6 code.

[23] Note that search intensity here departs marginally from its standard use in the search literature. The count of experimentation instances is taken over a period of time that varies across search spells, which are all delimited by the formation of a sustained relationship.

[24] Egan and Mody referred to these elements as the "inseparable triad" -price, quality and delivery- buyers study when assessing a supplier in a developing country.

facturers.[25] Such is the path followed by Sugita et al. (2015), Bernard et al. (2013) and Dragusanu (2014), among others.[26]

The roles that efficiency, quality, market power and product differentiation play in explaining size or performance differentials across firms is studied in a large panel in Hottman et al. (2014). Disentangling these structurally, they find that quality differences explain the majority of the variation in firms' performances. Instead of using observed exports as a measure for such idiosyncratic heterogeneity, I follow Hottman et al., assuming that marginal costs affect firm volumes only via the manufacturer's price, and interpreting quality or reliability as a demand shifter that pushes volumes up or down, given such price.

I exploit the fact that we observe both volumes and prices at a disaggregated level to obtain (scalar) measures of heterogeneity of manufacturers from a linear regression of volumes on buyer and seller fixed effects, using relatively light assumptions. The operational definition for heterogeneity is that, conditional on the product and buyer with whom a seller is trading, sellers who can sell higher volumes at a given price are recognized by the demand as better suppliers. Prices of inputs are included as an additional control for product-specific quality. This is an innocuous assumption in the context of garment production, where high-quality pieces are produced with better fabric, which, in turn, constitutes not only the bulk of the weight of the garment, but also the largest component in the per-unit cost. Suppliers' heterogeneity is recovered using an approach similar to that introduced by Abowd et al. (1999), as detailed in the Appendix.

The empirical distribution of suppliers' heterogeneity exhibits, as expected, thin long tails on either side, consistent with similar findings in the literature (see Figure 4 in Appendix, vis-a-vis Figure 3 in Hottman et al. (2014)). A parametrization of the distribution of types of suppliers present in every product category was obtained by fitting, via maximum likelihood, four parameters of a Generalized Extreme Distribution. The data fits bell-shaped Type I and II EV distributions well in all product categories. Aggregating HS6 codes by broad product categories that identify the gender and general class of garment, shows that products that are typically more fashion sensitive exhibit higher dispersion in sellers' quality (Table 14), which also varies across time, within a product category. Fashion-sensitive products are usually supplied through shorter orders, with quicker lead times and with a higher proportion of the order delivered in the first shipment.

The heterogeneity across suppliers present in a given market is well associated with measures describing the fashion sensitivity of the segment. Consider a *market* to be the interaction of a product category (HS6) and a season.[27] Following the evolution of these markets over time (calendar quarters), we can characterize the distribution of available suppliers by using the standard deviation of the types of sellers active in each market quarter. Regressing this as an outcome over quarter dummies and measures of the fashion sensitivity of the market confirms the positive association between dispersion in suppliers' types and fashion sensitivity.[28]

The first of those measures is the average, across all export orders in a given market, of the *lead*

[25] Standing evidence on the high correlation between size and productivity can be found, for example, in Bartelsman et al. (2013).

[26] In a related exercise, Besedes (2008) uses the per capita GDP of a country exporting to the United States as a proxy for the reliability of the supplier.

[27] Here four seasons per year are admitted, corresponding to - northern hemisphere's- high winter, spring-summer, high summer, autumn-winter. Results do not change when allowing for six seasons.

[28] The main estimating equation is $StDev(\theta_j)_{mt} = \pi + \pi_t + \pi_1 LeadTime_m + \pi_2 OrderTurnover_m + \pi_3 Fem_m + \pi_4 StDev(p^f)_m + \epsilon_{mt}$.

time computed as the time span in days between the first incoming shipment of fabric and the first outgoing shipment of garment. Fast fashion segments, in general, exhibit shorter lead times because the life cycle for these products tend to be shorter as well. We observe then that longer lead times are associated to markets with lower heterogeneity across suppliers. Likewise, orders of fashion-sensitive products tend to be shorter, with all shipments in the order concentrated in a small window of time. The second regressor in the specification above is the average, across all export orders in a given market, of the time gap between the first and the last shipment within an order. The larger this gap, the lower the heterogeneity across suppliers. Finally, female product categories and segments with large variability in the qualities of the fabric used are associated with highly heterogeneous markets. All results are presented in Table 15 in the Appendix.

Fact 3. *The sourcing strategy of the buyer differs across different environments.*

As shown above, markets (product-season combinations) can be characterized by the heterogeneity across the available suppliers. Organising markets in ascending order according to this feature, we can construct quartiles of such distribution so the bottom "bin" (first quartile) contains markets in which the heterogeneity across suppliers is low. The top quartile, then, contains those markets that exhibit high dispersion of suppliers. This is reflected in the horizontal axis of the Figure 5.

Within a given market, we can consider the share of exports the buyer channels through already known suppliers as a measure of persistency in these relationships. Figure 5 shows the average of this measure, across all the buyers and markets in each quartile. Note that the shares have been de-meaned for each buyer, so they are centered around zero at the buyer level.

In markets with low heterogeneity across the available sellers (first quartiles) buyers tend to place a lower proportion of their export orders with known suppliers (again, relative to their own means). That is, a higher share of a buyer's demand goes to unknown suppliers. At the other end of the spectrum, in the markets that exhibit high dispersion across sellers's types, buyers allocate a larger proportion of their trade with partners they already know.

The - unconditional - average share of exports sourced to known suppliers in the sample is 0.71. The difference from the bottom to the top quartile in Figure 5 represents a change in persistency of approximately 13% relative to the average.

Fact 4. *The cost of experimenting with a supplier depends on the supplier's type.*

Unlike other consumer search problems, the cost of search in the Bangladeshi context depends on the individual option (supplier) the buyer is screening. It is often assumed in search problems that acquiring the relevant information from an option in the top tail of the relevant distribution is as costly as doing so with an option in its lower tail. That is, for instance, obtaining a price quotation from a high price firm costs the searcher the same as obtaining a quotation from a low price firm. The explorative nature of buyers' search in the context of this paper is such that this symmetry doesn't hold: placing testing orders with suppliers of low reliability and quality can cost dearly to the buyer, while marginally *better* suppliers do not necessarily reduce the costs the buyer faces in the testing stage.[29] Three examples of this asymmetry are in order.

[29] The standard search process would includes in general four stages. First, the buyer assesses like the supplier's size, its machinery and in-site safety measures, payment of minimum wage, freedom of association, the non-use of banned chemicals and absence of forced or child labor, among others. This stage is conducted in site, often through scheduled and surprise inspections. Second, 'head audits' take place, focused on management investigations, worker interviews and the examination of company files and records (timesheets, wage bills and contracts). For instance, this process

First, having a supplier ready to ship samples in advance of the arranged delivery time is likely to have no value to the buyer. However, a supplier that cannot meet the arranged delivery times can induce high costs to the buyer. In the process of testing, having samples later than scheduled can induce additional managerial costs to the buyers' assessment teams. In the data we observe realised shipment dates -and not scheduled shipments- so we cannot directly measure failures to timely delivery. However, we observe the mode of transport of each shipment. The industry, whose demand is mainly located in Europe and U.S., ships mainly by sea (98% of the transactions in our data) and air shipments tend to be reserved to urgent or last minute deliveries.[30] Splitting the set of suppliers above and below the mean quality of supplier, we observe that those below the mean have a probability of shipping by air that is twice as high as that of those above the mean (1.1% and 0.5%, respectively).[31]

Second, in testing stages, most of the quality standards are established in terms of thresholds: a given number of faulty pieces per sample lot, a certain "length" of loose threads in the finishing of the garment pieces, etc.. For rejection purposes, once the supplier surpasses the minimum threshold, there is no benefit to the buyer the further away (and up) the manufacturer's performance is. Product rejections are not observable to us. However, given that Bangladesh does not import apparel from its main export destinations, inflows of garments (ready-made, finished products) from Europe or U.S. into Bangladesh are often returns.[32] Again, splitting all manufacturers into those above and below the average of the quality measure, we observe that these inflows are 25% more frequent for below-average suppliers, relative to their better-performing counterparts. Moreover, if we compare garment producers below and above the 25^{th} percentile in the quality measure, those below are 56% more likely to receive these inflows.

Third, the material and reputation costs the buyer faces when dealing with suppliers involved in social compliance issues are sizeable. Beyond the observable credentials, reliable, high-quality manufacturers satisfy in practice the safety and human rights standards required by the buyer. In particular, manufacturers work hard to attest their non-engagement in subcontracting practices to non-compliant plants. However, buyers are often caught in media controversies exposing their dealings with plants of poor standards. In most cases, these episodes are ex-post found to be connected with unauthorised subcontracting by existing suppliers or small scale interactions at the start of relationships (see Table 8 for a summary).

The evidence here suggests that when starting interactions with suppliers, the buyer bares little (or no) cost if the supplier being tested exceeds a reliability threshold, but the loss the buyer might incur increase with the distance to the threshold if the manufacturer is below it.

---

can take between one and six person-days to $H\&M$ auditors, depending on the supplier's managerial practices. Third, specific designs, styles and technical aspects of various products are sampled. This stage can be burdensome for both parts when the local design team does not interpret buyer's requirements correctly, lead times for the development of samples extend beyond agreed schedules or raw materials are rejected, requiring re-sourcing. After all necessary adjustments to individual prototypes, a small scale order is placed to assess the back-to-back production process, lead times, consistency across pieces and demand response.

[30] The average transport time -port to port- from Dhaka to European destinations is 22 days by sea and 2 days by air.

[31] Looking at all the shipments in the data and arranging them by weight, already that on the 5^{th} percentile is large enough for the sea freight to be cheaper than the air freight. For example, shipping 50 kg. of garment from Dhaka to London in the most commonly used packaging in our data would range from USD483 and USD700 by air and USD321 by sea (quotations obtained from independent freight services: InterDirect, InterDepAir, InterCargo).

[32] In most import shipments with this characteristic, this interpretation can be confirmed in the formal Export/Import procedure, as the transaction is recorded as a "Re-import" or "Temporary Import for Return". However, this variable is missing or has miscoded information in a large number of these flows, so it cannot be used for the purpose of the characterisation here.

Fact 5. *Buyers trade with already known suppliers, depending on the quality of the buyer's portfolio of manufacturers.*

Each new order the buyer places in the market, constitutes an opportunity to return to a known supplier or to allocate the order to a new manufacturer. In the whole of the data, persistency in the choice of existing suppliers is very high. This remains the case when choice is measured in count of orders[33]: almost 70% of the orders placed between 2006 and 2012 are allocated to manufacturers with which the buyer has interacted in the past.[34]

This choice is made in markets that are thick.For each of the observed orders, I identify, based on observables, the set of available suppliers that could act as a supplier for the order. I select the suppliers who are active in the corresponding product category in the time period, who work with "similar" inputs (in terms of the average price of the fabric), and who have enough capacity to satisfy the order, based on productive capacity inferred from past exported volumes. Constructing these hypothetical choice sets shows that for every order placed to a given manufacturer, there are on average over 600 other potential suppliers.

Figure 6 plots the relationship between the persistency in the buyer's choice of suppliers and the type of seller with whom she has been interacting. The vertical axis contains the propensity of the buyer to allocate a new order to a supplier she already knows: that is, of all the new orders the buyer allocates at a given point in time, the proportion that goes to suppliers with whom the buyer has a recurrent trade relationship. The horizontal axis shows the (running) average over the types of suppliers with whom the buyer has interacted in the last four quarters. On the left bottom corner of the graph, we see that buyers who, for the last year, have traded with suppliers of an average (normalized) type of -2 have a propensity to allocate new orders to already known suppliers as low as 0.1.[35] On the opposite corner of the graph, buyers who have interacted with, on average, exceedingly good suppliers, have a very high probability of sticking with them. The probability of allocating orders to known manufacturers grows quickly as the quality of the current suppliers improves from very low types. The red dotted lines indicate the bounds for 50% of the data on the distribution of sellers' heterogeneity; that is, half of the suppliers in the data are of types enclosed by these limits. Within these, we see that changes in the quality of the portfolio of existing suppliers, are not associated to relevant shifts in the unconditional persistency of around 0.7. In other words, this pattern suggests that when the gain in its partner's quality the buyer can make by experimenting with new suppliers in not large (given that the buyer is already matched with suppliers that "look like" the majority of the alternatives), the buyer doesn't increase its exploration of unknown suppliers. If buyers are effectively sorting suppliers on quality as measured here, the observation of constant persistency for a large range of the support of suppliers' quality suggest that frictions are present in the exploration or switching processes. In a frictionless context, any non-zero marginal benefit from searching for an alternative partner should induce trade with unknown suppliers.

Overall, this evidence suggests that the heterogeneity in the pool of manufacturers with whom the buyer interacts, and the distribution of types of available suppliers matter for understanding the

[33]Instead of overall volumes as in the Fact above.

[34]Only interactions in the past 24 months are taken into account for the purpose of this definition. There are in the panel a few instances of trade between buyers and sellers whose relations have been inactive for more than two years. The re-engagement in trade here is treated as the start of a relationship with an unknown supplier, for the purpose of this classification.

[35]The normalization of types is such that here values are to be interpreted as distances to the mean in terms of standards deviations: a type equal to -1 here is a supplier whose idiosyncratic heterogeneity, as measured in this section, is one standard deviation below the average, for the basket of products it offers.

patterns of persistency in buyer-seller relationships. When the portfolio of suppliers with whom the buyer interacts is comparable to the bulk of the manufacturers available in the market, relationships have the average persistence (0.7). Potential marginal improvements don't affect such persistency. When the suppliers the buyer knows are of very low type, the buyer buys from new suppliers with a higher probability (below the 0.5 grey line in the graph). On the contrary, when the buyer has a good portfolio of manufacturers, she stays with them with a probability close to one.[36]

4 Buyers' Experimentation in Search for a Partner

A lay description of the model is as follows. International buyers of different sizes access a product market where a number of manufacturers can supply the product they are after. Manufacturers differ in their type, and the value produced in a buyer-seller relationship is positively related to both the size of the buyer and the type of the supplier. In this sense, we can think of the quality of the supplier as a scalar that ranks manufacturers according to their ability to fulfil orders in time, to interpret the buyer's requirements, to abide by social compliance norms, etc. Ex-ante, buyers don't know the specific type of each seller, but they know the distribution of types. This distribution captures the heterogeneity across suppliers available in the market. Every time the buyer needs to start a trading relationship, it can sequentially search for a suitable partner. Searching involves accessing the pool of manufacturers, randomly drawing one of them and placing a test order. Testing the supplier requires the buyer to allocate resources to the task, and smaller buyers devote more resources to this, relative to their own size. The testing stage yields no profits to the buyer, but informs her of the type of the supplier.

Quality types and value Consider one specific product market, with infinitely many buyers indexed by b and sellers or manufacturers denoted with s. Buyers differ in their size, denoted ϕ_b. For simplicity, buyers are located in the $[0, 1]$ space and the count of buyers per seller is normalised to one. Manufacturers offer horizontally differentiated products, at a marginal cost m and they differ in their quality, θ_s. This is drawn i.i.d. from a distribution $F(\cdot)$ over $[0, \overline{\theta}]$, whose pdf is differentiable and be denoted by $f(\cdot)$. In addition, $(1 - F)$ is assumed to be log concave. When matched, the pair (ϕ_b, θ_s) produce value according to a function $q(\phi_b, \theta_s) = \phi_b\theta_s$.[37] Denoting the price of s's product with p_s, buyers derive profits from buying the product from s following $\pi_{sb} = \phi_b\theta_s - p_s$. It is assumed that suppliers cannot price discriminate across buyers.

Search Costs On the occasion of starting a trade relationship, the buyer decides whether to trade with its current partner -which, again for simplicity, I restrict to one supplier- or to sequentially search for an alternative seller. Search is costly and occurs through test orders. The buyer draws a potential supplier from $F(\cdot)$ and allocates it a project of overall cost proportional to the buyer's size: $\phi_b c(\phi_b)$, with marginal effort $c(\phi_b)$ continuous and monotonically decreasing in ϕ_b.[38]

[36] This positive relation is consistent with the findings, at a higher aggregation level, in Besedes (2008), where switching to new suppliers is negatively associated to the efficiency of the incumbent partner.

[37] Alternative specifications of the production function are of course possible. This multiplicative form coincides the descriptive exercise in Shimer and Smith (2000).

[38] And $\partial\phi_b c(\phi_b)/\partial\phi_b > 0$.

The test order renders no value to the buyer if successfully completed, but uncovers the realization of θ for the drawn supplier.[39] We can define then $\underline{\theta}_s^{\phi_b}$ as the lowest type of supplier that can complete a test project $\phi_b c(\phi_b)$. [40]

If the supplier is of a standard higher than $\underline{\theta}_s^{\phi_b}$, the buyer derives a value from the interaction that compensates for the invested resources. If the supplier is of a lower standard, the buyer loses a proportion of the invested resources, paying a cost decreasing in the type of the supplier. The cost for buyer ϕ_b of testing a supplier of type θ_s is then:

$$c_{bs} = \begin{cases} \phi_b c(\phi_b)(\underline{\theta}_s^{\phi_b} - \theta_s)^2, & \text{if } \theta_s < \underline{\theta}_s^{\phi_b} \\ 0, & \text{if } \theta_s \geq \underline{\theta}_s^{\phi_b} \end{cases} \tag{1}$$

Note that the low type of unsuccessful suppliers matters in terms of its distance to the lowest successful supplier, given the buyer's investment.[41] From here, the expected search cost for ϕ_b is given by:

$$c_b = \phi_b c(\phi_b) \int_0^{\underline{\theta}_s^{\phi_b}} (\underline{\theta}_s^{\phi_b} - y)^2 f(y) dy \tag{2}$$

For notation simplicity, c_b denotes the marginal cost of buyer ϕ_b when undertaking an instance of search. Let $\overline{c}$ be the highest possible of these search costs: $\overline{c} = c(0)$ and define analogously $\underline{c}$ as the lowest search cost $\underline{c} = c(1)$. Then, the search cost function denotes the mapping from $[0, 1] \rightarrow [\underline{c}, \overline{c}]$ where each c_b is the marginal search cost for a buyer of size ϕ_b. The distribution of the resulting search costs is denoted with $G(\cdot)$ and $g(\cdot)$ pdf.

While $F(\cdot)$ characterises the types of potential partners, $G(\cdot)$ captures the differential difficulties in finding a suitable partner for buyers of different sizes.

Search Behavior Consider the market's equilibrium price to be p^*. Ex-ante, buyers do not observe deviations $p_s \neq p^*$, so the search behavior of a buyer with search cost generically denoted as c and whose current best alternative is of type Θ, follows the optimal search rule:[42]

$$h(\Theta) = \int_{\Theta}^{\overline{\theta}} (\theta - \Theta) f(\theta) d\theta = c \tag{3}$$

[39] Note that the assumption of a successful testing order having no other yield than the information on the type of the supplier is in essence equivalent to the idea behind starting relationships with small pilot projects in Rauch and Watson (2003).

[40] It is assumed that $\underline{\theta}_s^{\phi_b} > 0$ for all buyers, so the measure of non-capable suppliers is non-zero for all buyers.

[41] The intuition behind this cost structure is as follows. When assessing a supplier, the buyer spends time and other resources in getting to know the manufacturer, agreeing on quality standards, adjusting production processes, auditing social compliance issues, etc.. At this screening stage, the focus is on minimum thresholds. For instance, if the buyer is evaluating the finishing of a test garment piece, and would accept x faulty stitches, a manufacturer producing a garment with $x - 1$ faults is as good as one with exactly x faults. However, a supplier whose test piece was below the quality threshold will require the buyer to scrutinise the production process, eventually request a second attempt, etc.. In the same spirit, when evaluating timely delivery, if the requirement for shipment specifies a lead time of y days, suppliers that are ready to deliver in $y - 1$ days don't add value to the buyer at this stage. But manufacturers that do not meet the deadline impose additional costs to the buyer.

[42] Technically, Θ corresponds to the best option the buyer has encountered so far. As with standard search problems of this nature (buyer facing i.i.d. alternatives and no switching costs, continuing to search if its highest researched option renders payoffs below the reservation utility), the choice problem follows stationary reservation utility strategies, as shown in Kohn and Shavel (1974).

The $\hat{\Theta}(c)$ that solves the above equation represents the threshold type of seller for a buyer facing search cost c, such that if its current match is of a θ type higher than $\hat{\Theta}(c)$, the buyer does not search for another supplier.[43] The $h(\Theta)$ function above is monotonically decreasing and has a unique solution over the interval $[\underline{c}, min\{\overline{c}, E[\theta]\}]$. Taking first and second derivatives of equation 3 shows that $\hat{\Theta}(c)$ is decreasing and convex in c over the relevant interval and $\hat{\Theta}(E[\theta]) = 0$ and $\hat{\Theta}(\underline{c}) = \overline{\theta}$.[44]

In this context, upon meeting supplier s, the probability that b retains s as its supplier is given by:

$$Prob[\phi_b\theta_s - p_s > \phi_b\hat{\Theta}(c) - p^*] = 1 - F(\hat{\Theta}(c) + \frac{p_s - p^*}{\phi_b}) \tag{4}$$

The buyer will only search for an alternative if the expected gains are greater than those under the arrangement with s. For completeness, I present the main pricing equations for symmetric Nash equilibria.

Prices To construct supplier s's demand, we need to integrate over all the buyers that would search the market and eventually trade with s. First, I will assume that the highest cost of searching is low enough, such that buyers of all sizes can in principle consider searching. Second, the probability of trade happening between a buyer with cost c and supplier s, irrespective of the previous history of the buyer (who the best known supplier is thus far), is the unconditional probability given by $\frac{1-F(\hat{\Theta}(c)+\frac{p_s-p^*}{\phi_b})}{1-F(\hat{\Theta}(c))}$.[45] Integrating over all c, the demand for s is:

$$Q(p_s, p^*) = \int_{\underline{c}}^{\overline{c}} \frac{1 - F(\hat{\Theta}(c) + \frac{p_s - p^*}{\phi_b})}{1 - F(\hat{\Theta}(c))} g(c)dc \text{ } if \text{ } p_s < p^* \tag{5}$$

The profits for s are given by $\pi_s = (p_s - m)Q(p_s, p^*)$. Taking first order conditions from the maximisation of profits and exploiting symmetry:[46]

$$p_s = p^* = m + \frac{1}{\int_{\underline{c}}^{\overline{c}} \frac{f(\hat{\Theta}(c))}{1-F(\hat{\Theta}(c))} g(c)dc} \tag{6}$$

Conditions for existence and uniqueness of a symmetric equilibrium following equation 6 are fully developed in Moraga-Gonzalez et al. (2014).[47] In equation 6, the numerator in the second summand comes from the fact that $G(\overline{c}) = 1$. Denoting $p^*(F)$ the equilibrium price for a given F, it will be useful to re-write the price equation in the usual monopolistic competition form:

[43] For the derivation of this rule, note that the decision to search is a comparison between the highest option the buyer currently holds and the benefits, net of search costs, of searching -once- for an alternative. If that alternative gives higher payoffs, it will be preferred to the original highest option. If, instead, it gives lower payoffs, free recall guarantees that the buyer keeps its original partner.

[44] As in Moraga-Gonzalez et al. (2014), these derivatives render $\hat{\Theta}'(c) = \frac{-1}{1-F(\hat{\Theta}(c))} < 0$ and $\hat{\Theta}''(c) = \frac{f(\hat{\Theta}(c))[\hat{\Theta}'(c)]^2}{1-F(\hat{\Theta}(c))} > 0$.

[45] This is, the buyer might have arrived at the exploration of supplier s having searched no suppliers before or having searched many alternatives prior to meeting s. Therefore, the buyer starts to trade with s if s is above the cutoff and all suppliers -none or many- previously explored are below the cutoff. The expression here follows from the i.i.d. assumption on the draws of θ.

[46] $p_s = p^*$.

[47] Strictly speaking, proving existence of this unique equilibrium requires an additional step that guarantees that $g(\cdot)$ is log concave. A formal proof of this is beyond the scope of this section, but examination of the cost mapping that gives G should suffice.

$$p^*(F) = \frac{m}{1 - \frac{1}{e(p^*,F)}} \tag{7}$$

where, $e(p^*, F)$ is the elasticity of demand when the distribution of types (and then the search costs) is characterised by F and the equilibrium price is p^*, so:

$$e(p^*, F) = p^* \frac{Q^{'}(p^*, F)}{Q(p^*, F)} = p^* \int_{\underline{c}}^{\overline{c}} \frac{f(\hat{\Theta}(c))}{1 - F(\hat{\Theta}(c))} g(c) dc \tag{8}$$

Dispersion of types I now study changes in the dispersion of sellers' types. For simplicity, consider mean preserving spreads of F. Let j denote the ordering of successive mean preserving spreads, such that for $F_{j''}$ is a mean preserving spread of $F_{j'}$ whenever $j'' > j'$, such that $\int_0^z [F_{j''}(\theta) - F_{j'}(\theta)]d\theta \geq 0 \; \forall z$ and > 0 for some z.

An increase in the dispersion of suppliers' types, characterized by a mean-preserving spread of the initial distribution, alters experimentation behavior. For buyers with high search costs and, then, a relatively low threshold of acceptable sellers, a greater spread of types increases their probability of drawing a supplier below the minimum threshold (see the increase from $F_1(\Theta(c1))$ to $F_2(\Theta(c1))$ in Figure 7 in the Appendix).[48] This, in turn, increases the intensity of search (again, in terms of the count of experimentation instances).

After the spread, the cost of search for every buyer is weakly higher relative to cost of searching in the less dispersed distribution. This lowers the acceptance threshold for all buyers (this is $\Theta(c2)$ inducing a low $F_2(\Theta(c2))$ in Figure 7). This second effect, operating via the cost of searching, reduces experimentation efforts further for buyers with high thresholds and can offset the increase in search in the case of buyers with low thresholds. As a result, an increase in the heterogeneity in the environment characterized by a mean-preserving spread of the distribution of suppliers' types can reduce search for all buyers.

Focusing first on equation 4, an increase in the dispersion of types that is mean preserving, other things equal, "flattens" f in the sense of second order stochastic dominance. Buyers that exhibit high costs of search (and that then have lower $\hat{\Theta}(c)$ thresholds), have that $F_{j''}(\hat{\Theta}(c)) \geq F_{j'}(\hat{\Theta}(c))$, implying that the probability of stopping the search process and settling with a drawn supplier is lower for these buyers. The intuition behind this is that the higher dispersion implies a weakly greater probability of the drawn partner being below the type the buyer is willing to keep. On the contrary, buyers whose search cost is low, and that then have higher $\hat{\Theta}(c)$ thresholds, reduce their search intensity.

However, the change in F shifts the whole distribution of costs G, because $\int_0^{\underline{\theta}_s^{\phi_b}} (\underline{\theta}_s^{\phi_b} - y)^2 f_{j''}(y) dy \geq \int_0^{\underline{\theta}_s^{\phi_b}} (\underline{\theta}_s^{\phi_b} - y)^2 f_{j'}(y) dy$. For every buyer, the expected cost of undertaking a new search instance is (weakly) larger (see equation 2). As $\hat{\Theta}(c)$ is decreasing in c, given F, the probability of staying with the known supplier increases. The intuition behind this is that the higher dispersion in types makes searching more "risky", as the likelihood of drawing a low supplier that would destroy the buyer's testing investment is high. This increases search costs, bringing the reservation value of the buyer down and increasing the probability of staying with its known supplier.

[48] Note that in the lower regions of the support of the distribution of types, a mean preserving spread has a steeper CDF relative to the original distribution.

5 Market Outcomes

The positive relationship between within-industry heterogeneity and the buyers' need for additional search and exploration has been highlighted both theoretically and empirically.[49] However, if such heterogeneity affects the costs of undertaking exploratory searches, the relationship between uncertainty (or unfamiliarity) and search behavior is unclear. The set up above highlights that heterogeneity across alternative suppliers affects the expected payoffs of a relationship both directly, via the probability of finding a suitable partner, and indirectly, via the cost of testing alternative partners.

The model here leads to four implications that can be tested in the data. Contrary to the observation that higher heterogeneity in the environment intensifies search efforts, the model here implies that *higher dispersion in the types of suppliers can be associated with lower overall search activity.* Moreover, *the effect of suppliers' heterogeneity in exploration behavior depends on the buyer's cost of search and, with this, on the buyer's size.* For larger buyers, search activity is lower in more heterogeneous environments. For buyers with large search costs (small buyers), higher heterogeneity might imply higher or lower exploration efforts, depending on whether the dispersion or cost effects prevail. Finally, *in more heterogeneous environments buyers are willing to match with (weakly) lower suppliers.* For small buyers, with high search costs, the effect of dispersion on the threshold acceptable supplier is negligible, while *for large buyers, small changes in the underlying dispersion bring the threshold supplier down significantly.*

The exercises that follow test those implications in our data. To this purpose, I introduce here some operational definitions.

At every point in time, it is possible to observe specific pairs of buyers and sellers trading. This market outcome can be conceptualized as the result of multiple decisions by the buyer to allocate each of its orders to a supplier, out of the choice set of available manufacturers. As in many discrete- choice problems, the data reflects consummated trade, that is, ex-post decisions. Then, the set of suppliers available to the buyer constitutes a hypothetical choice set for the relevant decision. Actual choice sets are unobserved and presumably exhibit large variation across decision makers. The baseline definition will be that when allocating an order, buyers face all suppliers that are active in a window of time in the relevant product category. For each of the observed orders, I identify the set of available suppliers that could potentially act as suppliers for the order, based on this definition and offer later on a refinement that restricts choice sets associated to an order to include only suppliers that are large enough to fulfil it. These hypothetical choice sets show that for every order placed, there are, on average, just over 660 potential suppliers.[50] Ranking all the orders according to the size of their

[49] *Unfamiliarity* of the buyer with the environment has been shown to induce exploratory search and testing. In the form of allocating small test projects, this is suggested in Rauch and Watson (2003) and empirically shown in Besedes and Prusa (2006). The mechanism behind their prediction is somewhat different to the one presented here. In their context, when the environment is more unfamiliar (and then search costs are higher), the mass of suppliers with which it is optimal to do testing increases. The higher cost of search brings down the expected value of drawing a supplier from the pool. With this, the cost of delaying the start of a large project (a sustained relationship) goes down, making test orders more appealing. Then, the more unfamiliar the environment is (and the higher the search costs) the more likely the buyer is to test the supplier with a small order.

[50] The technical construction of the choice sets is as follows. I date each order in the panel using the date of the first shipment related to the order. The order corresponds to one HS6 code in 87% of the cases. The remaining 13% corresponds to multi-product orders, but of these, there is a unique HS6 code in the order that explains more than 90% of the order's volume. I consider this to be the relevant HS6 code for the order. I first consider all the suppliers that are active in the HS6 code 18 months prior to and three months after the order allocation. I discard all the suppliers whose exit from the panel takes place before the date of the order. I also discard all the suppliers that don't show sufficient capacity to fulfil the corresponding order. For this, I take the ratio of the volume of the order to the duration (in days) of the order. For each seller, I compute this ratio for all past orders. If the maximum value of these ratios for

constructed choice set, the 10^{th} percentile has 182 potential suppliers and the 90^{th} percentile has $1,313$.

Each manufacturer is characterized by a scalar measure of its overall quality and appeal, as explained in Section 3 and detailed in the Appendix. I will use measures of dispersion across the types of suppliers in the buyer's choice set as a proxy for the heterogeneity in the buyer's decision environment. Similarly, I use the mean and median of types in the choice set as measures of central tendency in the distribution of available suppliers.

There are other possible definitions for the choice sets. In particular, as a robustness check, I will use the information about the actual allocation to weight the types of the suppliers when constructing the standard deviation as a proxy for the heterogeneity the buyer is facing. Of all the suppliers that are active in the corresponding product-time combination, I give higher weight to the subset of manufacturers that are "closer" to the manufacturer that is selected ex-post by minimizing a score based on three observable characteristics. For each pair of observations, formed by the actual supplier of the order and another manufacturer available in the market, the measure is constructed as the weighted product of distances between the realizations of the three variables, where weights are given by the corresponding covariance matrix [51]. The selected variables collect the quality of the input used by the seller (measured as the average price of the fabric used by the manufacturer up to the corresponding date), the experience of the supplier (measured in the number of quarters it has been producing the item), and an approximation of the *segment* of the market in which the seller operates (measured as the median buyer he is used to serving, ranked by its location in the normalized distribution of prices). This alternative is presented in the Appendix.

For some of the explorations below, I will distinguish between suppliers who are known to the buyer and those who are unknown. Known or existing suppliers are manufacturers with whom the buyer has traded in the last 18 months in any product category.[52]

5.1 Experimentation Activity

The search cost function in the previous section implies that in more heterogeneous environments, the costs of experimenting with unknown suppliers are higher. Therefore, highly heterogeneous environments may then exhibit lower search activity.[53]

I consider that the search for a suitable partner in a product market is successful when the buyer establishes a relationship that lasts for at least a year.[54] The relationships that are in this sense unsuccessful involve, in 50% of the cases a one-off shipment and in 95% of the cases interactions that last for less than one quarter. If search is costly in the form proposed in the previous section (there is a marginal cost to searching sequentially over partners), we should observe that buyers minimize

the seller is lower than 0.8 times the ratio of the relevant order, I consider that the seller doesn't have the productive capacity to be a potential supplier, and I eliminate it from the choice set.

[51] For pair (i,j), for instance, $\Delta'_{ij}W\Delta_{ij}$, where Δ_{ij} is a vector whose k^{th} entry is defined as $x_i^k - x_j^k$ and W is the covariance matrix over all k's. Note this is the square of Mahalanobis score.

[52] All results are robust to alternative definitions using cutoffs of 24 and 36 months.

[53] Previous studies have been able to uncover important patterns in buyer-seller relations using firm-to-firm trade flows. Focused on information frictions, Eaton et al. (2014), Monarch (2014), Dragusanu (2014) and Bernard et al. (2013) have "backed out" search costs exploiting observed prices and volumes traded within or across exporting relationships. The richness of the data used in this paper, allows for the direct observation of experimentation behavior. By identifying individual export orders, I can characterize search activity directly from the data.

[54] Of course, conditional on both parties being active in the panel in the corresponding period.

the search instances in environments where this cost is higher. We can then consider the number of unsuccessful relationships the buyer enters in after ending a long-lasting relationship and before starting a new one. High realizations of this measure correspond to high experimentation, meaning that the buyer attempted multiple instances of trade with different partners without these turning into long-lasting relationships, before succeeding with one.

For the empirical exploration, the search period is the time between the end of one long relationship and the commencement of a new long relationship. Two types of cases (at least) could be excluded from this definition. First, the buyer might decide not to search and to reallocate its orders to other (existing) suppliers instead. I consider this is potentially the case in instances in which we observe that within a year after the end of a long relationship the equivalent to the volume traded with the dismissed supplier is reallocated to the buyer's pre-existing suppliers at a level that exceeds the expected growth of those relationships for that particular buyer-product combination.[55] Second, there are cases in which the buyer starts searching before it effectively ends the relationship with a supplier. I consider the buyer to have replaced the broken relationship with anticipatory search whenever there is a new long-lasting relation that (a) starts at most six months before the end of the broken long-lasting relationship and (b) the traded volume with the new partner grows at a rate above the average for the buyer-product after the break-up of the first partnership.

There are over $21,500$ search spells in the data, reducing to $16,000$ after removing those attributed to re-allocation to existing suppliers or anticipated search.[56] These search periods have a median duration of 92 days and the 95^{th} percentile is 168 so search spells generally take between one and two quarters. For each of these search spells, we are interested in the search intensity, measured as the number of unsuccessful (short-lived) relationships started in the spell. Considering all the observed search spells, the average intensity is 2.4 and the 90^{th} percentile is 5.

Conditional on the buyer and the product category, low heterogeneity across suppliers should decrease the marginal cost of searching for an additional partner. Then, search intensity is expected to be negatively related to suppliers' heterogeneity. I use the following specification to test this hypothesis:

$$SI_{bms} = \alpha_M + \alpha_{t(s)} + \alpha_{bt(s)} + \alpha_{bM} + \beta_1 \bar{\hat{\theta}}_{ms} + \beta_2 St.Dev.(\hat{\theta})_{ms} + \mathbf{X}_{bms}\gamma + \epsilon_{bms} \tag{9}$$

where the outcome variable, SI_{bms} is the search intensity of buyer b when undergoing search spell s in product m, measured as the (log) of the count of short lived relationships during s. Fixed effects for the product category, α_M, and for the calendar quarter of the spell, $\alpha_{t(s)}$ are included in some specifications.[57] Buyer-product and buyer-quarter specific intercepts are also allowed for and denoted with $\alpha_{bt(s)}$ and α_{bM}. The two regressors of interest are $\bar{\hat{\theta}}_{ms}$ and $St.Dev.(\hat{\theta})_{ms}$, the mean and standard deviation of qualities of available suppliers in spell s in product m.[58] Other controls of interest are

[55]In practice, for each buyer-product combination I compute the average inter-annual growth rate of long lasting relations. With this, I generate the hypothetical volumes of trade between a buyer and all its suppliers over the year in a product category. After a breakup, I compute the volume traded with existing suppliers in excess of this projection. If the combined volume exceeds the annual volume traded with the dismissed supplier, I consider that the search period might be characterized by re-allocation.

[56]For the purpose of these counts and the estimations presented in this section, the first and last 14 months of the panel are excluded.

[57]Here, $t(s)$ is the quarter in which the majority of the spell occurred. When the quarter with the largest share has less than 51% of the spell, the quarter of the end of the spell is assigned to $t(s)$.

[58]The average and standard deviation are computed over the types of sellers that are available to the buyer, following the definitions in the introductory paragraphs of this Section. In particular, I consider all suppliers that are active in

$q_{mt(s)}$, the (log) size of the market demand in the quarter of the spell, $N^s_{mt(s)}$ the number of available suppliers in the relevant quarter, the quality of the supplier in the broken relationship, θ^x_{bms}, and the share of the buyer's demand this supplier met in the last year of activity in the relationship, sh^x_{bmT-1}.[59]

The results of linear regressions following the specification above are presented in Table 16 .

Column (1) considers all the spells we observe in the panel, while columns (2) to (7) exclude those spells that satisfy the criteria that define them as instances of potential re-allocation to existing suppliers or anticipatory search. Across all specifications, conditional on the number and average quality of available suppliers, the dispersion across them brings the search intensity down. Figure 8 shows the distribution of the heterogeneity measure over all search spells included in these regressions. Sets of suppliers that are more dispersed by one standard deviation are associated with a 32% decrease in the count of search instances. For concreteness, consider a buyer seeking a supplier in an environment as heterogeneous as that of female woven dresses of artificial fibers for spring/summer. If that heterogeneity were to drop to the levels seen in summer cotton male shirts, other things equal, search intensity would increase by one additional instance, relative to the median of four search instances in the product category. This is robust to the inclusion of buyer effects disaggregated at the buyer-product and buyer-season level. Most of the variation is absorbed in the fixed effect once buyer-quarter specific intercepts are allowed for, as there are virtually no instances in the data in which a buyer ends multiple relationships in a given calendar quarter. In terms of the formalization of the previous section, less heterogeneous environments with lower search costs, imply a higher threshold type of supplier the buyer is willing to take, making it more likely that the buyer will continue to search before settling on a supplier. In this sense, the results in Table 16 refer to the intensive margin of the search process: conditional on ending a relationship and searching, buyers undertake more search instances, the less heterogeneous the pool of available suppliers is. I explore the extensive margin separately later in the paper.

Columns (6) and (7) show that after including buyer-product effects, the characteristics of the supplier that is being replaced do not affect search intensity. The quality of the supplier and the share of the buyer's demand that was met by the supplier have no significant effects on the amount of search the buyer undertakes.

5.2 Buyer's Size and Experimentation Behavior

The model above assumed that search costs were decreasing in the buyer's size. Therefore, small buyers' responses to an increase in the heterogeneity were due to the net effect of an increase in the intensity of search due to higher rejection rates and a decrease in search due to a downward shift in the threshold of acceptable suppliers.

According to this, the negative relation between higher heterogeneity and search intensity should be

the product category in the 18 months prior to the search spell and the three months following it. When imposed, the capacity restriction is implemented using the average size of the median orders placed by the buyer during the relevant spell.

[59] The importance of within-industry heterogeneity as a driver of firm turnover in international markets has been studied in a somewhat similar way in Alvarez and Lopez (2008). After recovering firm-level measures of TFP as residuals in a production function regression (a la Levinshon and Petrin) in a survey panel of 5,000 Chilean plants, Alvarez and Lopez proxy within-industry heterogeneity as the median difference in TFP between the firms in the top and bottom of the distribution.

stronger for relatively large buyers, and be moderately negative or reverse sign for small buyers. To test this hypothesis, I augment the specification in equation 9 to include indicators of the decile of the size of the buyer. The results of this exercise are presented in Table 17 and are comparable, in terms of the included regressors, to Columns (2) and (6) in Table 16.

Column (1) in this table includes intercepts specific to each decile of the distribution of buyers' sizes. With the smallest buyers (Decile 1) as the base category, search intensity significantly increases with the size of the buyer. According to the model presented above, conditional on searching, larger buyers have higher acceptance thresholds in the search process. This means that the lowest type of supplier they are willing to accept is, other things equal, higher than that of their smaller counterparts. Given this, the probability of rejecting a drawn supplier in the search process is higher for larger buyers. The first nine entries in column (1) support this. Columns (2) and (3) include buyer and buyer-product fixed effects, so the decile intercepts are not included in these regressions. Across all specifications, the sign and magnitude of the effect of sellers' dispersion remains similar to those reported in the analysis of Table 16.

However, in the new specifications, allowing for decile-specific slopes shows that increases in suppliers' heterogeneity increase search intensity for the smaller buyers. Moving up in the distribution of buyers' sizes, the upward shift in search intensity shrinks and from the median-sized buyer onward, higher heterogeneity induces lower search intensity. This pattern is compatible with the representation in Figure 7: buyers whose threshold supplier is high enough decrease their search intensity in the face of a more spread-out distribution of suppliers, both due to the downward shift in the distribution of types and to the drop in the threshold; buyers whose threshold is low experience an upward shift in the distribution, which may or may not be compensated by leftward shifts in the threshold manufacturer.

5.3 Sorting and Matching Thresholds

Search costs were specified to be increasing in the heterogeneity across suppliers. This in turn implied that, for a given buyer, the threshold supplier she would be willing to accept would be lower in highly heterogeneous environments. This negative relation can be tested using the specification:

$$\bar{\bar{\theta}}_{bmt_{|\theta>\Theta_{bmt}}} = \alpha_b + \alpha_t + \alpha_m + \beta_1 StDev(\hat{\theta}_{mt}) + \beta_2 q_{bmt} + \epsilon_{bmt} \tag{10}$$

The dependent variable, $\bar{\bar{\theta}}_{bmt_{|\theta>\Theta_{bmt}}}$, is the average type of supplier the buyer has accepted (is trading with beyond a one-off interaction) in the product-year combination. It is expected that the average supplier the buyer accepts is positively related to the buyer's size, q_{bmt}. Other things equal, large buyers experience low costs of search and are then willing to search until a higher type supplier is found. The dispersion in suppliers' types in the $-mt$ combination, $StDev(\hat{\theta}_{mt})$, is meant to shift the cost of search upwards for all buyers participating in the market. This, in turn, should bring down the threshold supplier the buyer is willing to accept. Thus, we expect β_1 to be negative. Moreover, the convexity in Θ with respect to search costs implies that increases in the search cost will bring the threshold down significantly for large buyers, whose marginal cost of search is relatively low. At the other end of the distribution, for small buyers with high search costs, an increase in the dispersion of types should have a small effect on the threshold supplier they are willing to take. To explore this, a different slope for $StDev(\hat{\theta}_{mt})$ is allowed for each quartile in the distribution of buyers' size.

The results are presented Table 18. Across all specifications, the relation between the buyer's size and the average type of accepted supplier is positive and strongly significant, as expected. This result confirms the positive assortativeness between size of the buyer and quality of the supplier, suggested in other explorations in the literature.[60] Conditional on the size of the buyer, dispersion always brings down the average type of supplier with whom the buyer matches. This result relates to the findings in Dragusanu (2014), where buyers would search more intensely (invest more) when operating in product categories that are less differentiated (more homogeneous). In Dragusanu's setting, this relation arises as the result of the marginal benefits of search being increasing in the elasticity of the demand the buyer is facing. In my set up, search intensity increases when the choice set for the buyer is more homogeneous, via the reduction of the costs of undertaking exploratory search. While in Dragusanu's analysis positive assortativity worsens with product differentiation (in the buyer's end market), in the setup here positive assortativity worsens with suppliers' dispersion (in the buyer's upstream sourcing market).

Focussing on columns (3) to (5), accounting for buyer (or buyer-year) specific intercepts, an increase in the standard deviation across the types of the available manufacturers pulls down the average accepted supplier. As an example, consider that the dispersion across suppliers present in (an average $-mt$ combination within) summer cotton male shirts was to increase, such that the standard deviation goes up by one and assume a buyer was accepting, before the increase in dispersion, suppliers whose average type coincides with the median type in the market. Other things equal, the increase in dispersion would imply a shift of the average type of suppliers to the 25^{th} percentile of the distribution.

Columns (6) and (7) confirm the convexity of this effect. Small buyers, whose search costs are high, don't change significantly the average type of accepted supplier in more heterogeneous environments. In the context of the model in the previous section, Θ is less responsive to changes in search costs, for high values of these costs. By contrast, large buyers located in the top quartiles of the distribution, exhibit significant drops in the type of accepted suppliers in the presence of higher heterogeneity in the pool of available manufacturers.

6 Additional Implications and Extensions

The focus of the analysis here has been on the intensive margin of search defined as the count of experimenting instances the buyer undertakes. In other words, conditional on the buyer's participation in search activities, how much does the buyer search. The extensive margin of this process, whether the buyers search at all or not, is also relevant. In Subsection 6.1 I show that in environments that exhibit higher heterogeneity, the probability of allocating an order to an unknown supplier is lower.

The relation between heterogeneity and search intensity, through its effect on the demand, can affect prices. In particular, for a given supplier (and marginal cost), when higher search costs due to increased heterogeneity bring search intensity down, the demand for the supplier becomes more

[60] In particular, Sugita et al. (2015) and Dragusanu (2014) use firms' sizes to proxy their "capabilities" and corroborate empirically the -positive- sign of the sorting. The exercise in this paper involved recovering a measure of the supplier's quality, free from other factors that might affect the manufacturer's size.

inelastic. This, in turn, could bring price-cost margins up. This hypothesis is explored in Section 6.2.

6.1 The Extensive Margin: Persistency in Buyer - Seller Relations

Conditional on the buyer's size and characteristics, as search costs are increasing in the heterogeneity in the environment, we should observe that the probability of retaining a known supplier (rather than searching for an alternative) goes up in settings in which suppliers' types are more dispersed.

Every allocated order in the panel can be interpreted as the result of a binary decision, made by the buyer, of allocating the order to one of its existing suppliers or to a new one. A positive outcome in this binary choice signals persistency in the choice of sellers, favoring already known manufacturers.

The baseline specification of interest describes the probability of an order o for product m being allocated at time t by buyer i to any existing supplier. Note that in our panel, an order identifies m, i and t uniquely, so part of the sub-indexing below is redundant, but, hopefully, clarifying.

$$Pr(a^K_{oimt} = 1|X, \hat{\theta}) = \Phi(\alpha + \beta_1 \bar{\hat{\theta}}^k_{oimt} + \beta_2 \bar{\hat{\theta}}^u_{oimt} + \beta_3 StDev(\hat{\theta}^u_{oimt}) + X'_{oimt}\gamma) \tag{11}$$

The outcome variable a^K_{oimt} takes value one if order o, in product category m at time t, is allocated to a supplier that is known by buyer i. Recall θ_j constituted the the seller fixed effect obtained in the previous section, as a proxy for the type of the supplier. $\bar{\hat{\theta}}^k_{oimt}$ is the average type of the known or existing suppliers to buyer i, relevant to the current order. Similarly, $\bar{\hat{\theta}}^u_{oimt}$ denotes the average type of all available suppliers that are unknown to the buyer.[61] $StDev(\hat{\theta}^u_{oimt})$ is the standard deviation across these unknown suppliers. X_{oimt} contains other covariates, including buyer, product and quarter fixed effects, counts of known and unknown players on each side of the market, the size of the order and the size of the overall demand in the product-quarter combination. Table 19 presents the results from a Maximum Likelihood Probit estimation following the equation above.

Across all specifications, we corroborate that the probability of re-allocating an order to the pool of known manufacturers is increasing in the average type in the pool. However, this effect is not dramatically large. A unit increase of the corresponding covariate would imply shifting that average, from the median to the 95^{th} percentile. Such a jump would induce an increase in the probability of allocating orders to the known suppliers of approximately 0.01. With an unconditional probability of choosing a known supplier of 0.69 over the sample included in this exercise, such an increase represents a 1.4% change.

While the median (and average) type of the unknown suppliers does not have a significant effect on the outcome, increases in the deviation of types of the unknown suppliers increase (0.03 to 0.04) the probability of allocating orders to existing suppliers. Relative to the unconditional probability of reselecting an existing supplier, such changes imply 4.5% to 5.7% shifts. For concreteness, consider the following example. When allocating an order in a market such as female woven dresses of artificial fibers for spring/summer an average buyer faces relatively high heterogeneity and sticks to its known suppliers with high probability. If the dispersion in suppliers' types dropped to a level typical (median)

[61] Again, in this baseline specification, the set of unknown available suppliers is formed by all the suppliers active in the product category at the time of allocation of the order and that the buyer has not traded within the last 18 months.

of categories such as summer cotton male shirts, the probability of choosing a known partner would decrease by 5.3% relative to the unconditional average persistence in the segment.[62]

6.2 Price-Cost Margins

To study the effects of dispersion on profitability, I construct the price-cost margin at the order level as the difference between revenues and costs, as a proportion of the costs in the order (i.e., (PQ-C)/C). This constitutes the outcome variable of the estimation equation 12 below, denoted with μ_{ijoms}, varying at the level of the order o, placed by a buyer i to a seller j, in product m in sequencing time s.[63]

$$\mu_{ijoms} = \alpha_{ij} + \delta_m + \iota_{t(o)} + \gamma_1 s + X_{ijoms}\beta + \epsilon_{ijoms} \tag{12}$$

As explained above, orders span over time and for the purpose of these regressions we consider them to be dated by the date (aggregated in quarters) of the first shipment in the order. Orders in a trading pair are then arranged by date and numbered subsequently. This 'linear trend' is represented by s. $\iota_{t(o)}$ introduces seasonal corrections based on calendar times t of order o. Dummies for the buyer-seller pair, products and seasons are kept in all regressions.[64]

The results of these regressions are presented in Table 21. Across all specifications, we observe a small positive effect of every additional order in the relationship. Across all pairs in the data, the average number of orders in the relationship is 3.6, although a large share of the trade takes place in the top tail of the distribution of number of orders in the relationship, which, on the 95^{th} percentile is 12. The average price of the fabric used for producing the garment and the size of the order are both negatively related to the margin over costs, as expected. The number of buyers allocating orders in the relevant product - quarter combination shows a positive effect in the price - cost margins. An expansion of the demand of 1% measured via the count of buyers is associated with margins 0.085 higher.

Finally, conditional on the type of the supplier, an increase in the dispersion of types the buyer is facing leads to a strong positive effect on markups, though the effect is significant only at the 0.10 level due to large standard errors. An increase of one unit in the standard deviation measure raises price-cost margins by 0.10. It is useful to compare this effect with the shift in price-cost margins induced by trading with a "better" supplier. The results in Table 21 imply that if a buyer were to change sellers - changing from trade with a supplier in the median of the overall distribution of types (this is, pulling all products together), to trade with a supplier in the 75^{th} percentile of the distribution-, the price-cost margin for an otherwise equivalent order would go up by 0.11.[65]

[62] This result is compatible with Besedes and Prusa (2006) who at a higher aggregation level (comparisons across industries) find that switching to new or unknown suppliers is more likely in markets for homogenous products.

[63] Note that, m is specific to o and the triplet ijs fully defines o so notation here, is again technically redundant.

[64] I exclude all relations that last for less than a year.

[65] A change of 0.5 in the regressor.

7 Conclusions

This paper contributes to the understanding of the decisions firms in developed countries make regarding cooperation with potential suppliers from less-developed economies. The direct observation of buyers? experimentation with suppliers as documented by the data offers a window into these relationships with an unprecedented level of detail. Previous studies have been able to uncover important patterns in buyer-seller relationships using firm-to-firm trade flows. Those focused on information frictions have inferred the existence of search costs exploiting observed prices and volumes traded within or across exporting relationships. By contrast, the identification of individual export orders in our data, uncovers search activity directly - rather than ?backing it out? from aggregated flows.

Based on the observed patterns, I outline a model that describes the main trade-off between the benefits of finding a high-quality trading partner and the costs of undertaking the necessary exploratory search. I argue that these costs, in the context of many developing countries, are largely driven by the variability in quality across potential partners. I formalize this idea by endogenizing the costs of search to the distribution of manufacturers available to the buyer. Both the model and the empirical exercises uncovered novel facts of the formation of buyer-seller relationships in developing countries.

First, I show that higher heterogeneity in the environment is associated with lower search . Contrary to the predictions of standard search models, the more dispersed the characteristic buyers are uninformed about, the lower the number of exploratory trials they undertake. Related work showing a similar negative correlation between heterogeneity and search efforts has argued that this pattern is driven by product differentiation in the buyer's end market. Less differentiation downstream increases the marginal benefit of search for suppliers. In this paper, the relevant dimension of heterogeneity for buyers' experimentation decisions lies upstream. More heterogeneity across suppliers increases the marginal cost of trying out unknown manufacturers.

Second, the results suggest that larger buyers form lasting relationships with higher quality suppliers. This is compatible with similar findings on positive assortativity between exporters and importers. Whereas previous empirical work has used firms' sizes as a proxy for individual productivity or "capability", I employ an arguably preferable measure. I recover a scalar describing seller's quality, free from other factors that might affect the manufacturer's size. In particular, the quality measure here is obtained by 'removing' any buyer-specific effects in traded volumes.

Third, this paper shows that experimentation and sorting behavior varies with the size of the buyer. In a highly concentrated market where a dozen large buyers account for the majority of the exports from the sector, understanding how these players deal with information frictions when forming trade partnerships is key. Beyond the specificities of the Bangladeshi context, the presence of large, international corporations in low income countries is a distinctive feature of many development processes. The sourcing decisions of these buyers directly affect the survival patterns of exporters, their growth prospects, and the split of the gains from trade.

Finally, It is interesting to reflect on the policy implications arising from the analysis. The search frictions lead to inefficiencies in the match quality: buyers settle for lower quality suppliers as search costs increase. As has often been pointed out, there are global welfare gains from reducing search costs. But the data also suggest that these same search frictions leave domestic producers with a larger share of the rents from the relationship: price-cost margins increase with supplier heterogeneity. Thus, from the perspective of policy makers in Bangladesh, it is unclear whether reducing search frictions would be welfare increasing. Further understanding of these issues is important for development in countries whose growth is heavily reliant on international trade.

References

J. Abowd, F. Kramarz, and D. Margolis. High wage workers and high wage firms. *Econometrica*, 67 (2):251–333, February 1999.

J. M. Abowd, R. Creecy, and F. Kramarz. Computing person and firm effects using linked longitudinal employer-employee data. Mimeo, 2002a.

J. M. Abowd, R. H. Creecy, and F. Kramarz. Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau, Mar. 2002b. URL `https://ideas.repec.org/p/cen/tpaper/2002-06.html`.

J. Ahn, A. Khandewal, and S. Wei. The Role of Intermediates in Facilitating Trade. *Journal of International Economics*, 84(1):73–85, 2011.

J. Aker. Information from markets near and far: Mobile phones and agricultural markets in niger. *American Economic Journal: Applied Economics*, 2(3):46?59, 2010.

F. Albornoz, H. Calvo, G. Corcos, and E. Ornelas. Sequential Exporting. *Journal of International Economics*, 88:17–31, 2012.

G. Alessandria. Consumer Search, Price Dispersion, and International Relative Price Fluctuations. *International Economic Review*, 50(3):803–829, Aug. 2009.

T. Allen. Information frictions in trade. *Econometrica*, 82(6):2041–2083, 2014.

R. Alvarez and R. Lopez. Entry and exit in international markets: Evidence from chilean data. *Review of International Economics*, 16:692–708, 2008.

T. W. Anderson and C. Hsiao. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82, 1982.

A. Banerjee and E. Duflo. Reputation effects and the limits of contracting: a study of the indian software industry. *Quarterly Journal of Economics*, 115:989 to 1017, 2000.

E. Bartelsman, J. Haltiwanger, and S. Scarpetta. Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review*, 103(1):305–34, February 2013.

E. Barth and H. Dale-Olsen. Assortative matching in the labour market: Stylised facts about workers and plants. Mimeo, Institute for Social Research, Oslo, February 2003.

G. S. Becker. Job turnover and the returns to seniority. *Journal of Business and Economic Statistics*, 23(2):192–199, 2005.

D. Beil. Supplier Selection. in ?wiley encyclopedia of operations research and management science?, 2010.

F. Benguria. Production and distribution in international trade: Evidence from matched exporter-importer data. Mimeo, 2014.

A. Bernard, J. Jensen, S. Redding, and P. Schott. The Empirics of Firm Heterogeneity and International Trade. *Annual Review of Economics*, 4:283–313, 2012.

A. Bernard, A. Moxnes, and K. Ulltveit-Moe. Two-Sided Heterogeneity and Trade. Mimeo 9681, CEPR Discussion Paper, 2013.

T. Besedes. A search cost perspective on formation and duration of trade. *Review of International Economics*, 16:835?849, 2008.

T. Besedes and T. Prusa. Product differentiation and duration of US import trade. *Journal of International Economics*, 70:339?358, 2006.

D. Card, J. Heining, and P. Kline. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, 128(3):967–1015, 2013.

T. Chaney. The network structure of international trade. *American Economic Review*, 104(11): 3600–34, 2011.

L. De Wulf and J. Sokol. Customs modernization handbook. Technical report, World Bank, 2005.

R. Dragusanu. Firm-to-Firm Matching Along the Global Supply Chain. Mimeo, 2014.

L. Drozd and J. Nosal. Understanding International Prices: Consumers as Capital. *American Economic Review*, 102(1):364–395, Feb. 2012.

J. Eaton, M. Eslava, M. Kugler, and J. Tybout. *Export Dynamics in Colombia: Firm-level evidence.* The Organisation of Firms in a Global Economy. Harvard University Press, Cambridge, MA, 2008.

J. Eaton, M. Eslava, D. Jinkins, C. Krizan, M. Kugler, and J. Tybout. A search and learning model of export dynamics. Mimeo, 2014.

M. L. Egan and A. Mody. Buyer-seller links in export development. *World Development*, 20(3):321–334, March 1992. URL `http://ideas.repec.org/a/eee/wdevel/v20y1992i3p321-334.html`.

R. Feenstra and G. Hanson. Intermediaries in entreport trade: Hong Kong re-exports of Chinese goods. *Journal of Economics and Management Strategy*, 13:3–35, 2004.

A. Goyal. Networks versus markets in international trade. *Journal of International Economics*, 48 (1):7–35, 1999.

A. Goyal. Information, direct access to farmers, and rural market performance in central india. *American Economic Journal: Applied Economics*, 2(3):22?45, 2010.

C. Hottman, S. J. Redding, and D. E. Weinstein. What is 'Firm Heterogeneity' in Trade Models? The Role of Quality, Scope, Markups and Cost. CEP Discussion Papers dp1294, Centre for Economic Performance, LSE, Sept. 2014. URL `http://ideas.repec.org/p/cep/cepdps/dp1294.html`.

M. C. Janssen, J. L. Moraga-Gonzalez, and M. R. Wildenbeest. Truly costly sequential search and oligopolistic pricing. *International Journal of Industrial Organization*, 23(5-6):451–466, June 2005. URL `http://ideas.repec.org/a/eee/indorg/v23y2005i5-6p451-466.html`.

M. G. Kohn and S. Shavel. The theory of search. *Journal of Economic Theory*, 9:93–123, 1974.

R. Macchiavello. Development uncorked: reputation acquisition in the new market for chilean wines in the uk. Working Paper, University of Warwick, 2010.

R. Macchiavello and A. Morjaria. The value of relationships: evidence from a supply shock to kenyan flower exports. Working Paper, University of Warwick, 2010.

K. Manova and Z. Zhang. Export prices across firms and destinations. *Quarterly Journal of Economics*, 127:379 to 436, 2012.

J. McMillan and C. Woodruff. Interfirm relations and informal credit in vietnam. *The Quarterly Journal of Economics*, 114(4):1285 to 1320, 1999.

R. Monarch. It's not you, it's me: Breakups in u.s.-china trade relationships. Mimeo, November 2014.

J. L. Moraga-Gonzalez, Z. Sándor, and M. R. Wildenbeest. Prices, Product Differentiation, and Heterogeneous Search Costs. CESifo Working Paper Series 4889, CESifo Group Munich, 2014. URL `http://ideas.repec.org/p/ces/ceswps/_4889.html`.

J. E. Rauch and J. Watson. Starting small in an unfamiliar environment. *International Journal of Industrial Organization*, 21(7):1021–1042, September 2003. URL `http://ideas.repec.org/a/eee/indorg/v21y2003i7p1021-1042.html`.

R. Shimer and L. Smith. Assortative Matching and Search. *Econometrica*, 68(2):343–370, March 2000. URL `http://ideas.repec.org/a/ecm/emetrp/v68y2000i2p343-370.html`.

Y. Sugita, K. Teshima, and E. Seira. Assortative Matching of Exporters and Importers. Mimeo, March 2015.

M. Tewari. Successful Adjustment in Indian Industry: the Case of Ludhiana's Woolen Knitwear Cluster. *World Development*, 27(9):1651–1671, September 1999. URL `http://ideas.repec.org/a/eee/wdevel/v27y1999i9p1651-1671.html`.

C. M. Wilson. Market frictions: A unified model of search costs and switching costs. *European Economic Review*, 56(6):1070–1086, 2012. URL `http://ideas.repec.org/a/eee/eecrev/v56y2012i6p1070-1086.html`.

A. Wolinsky. True Monopolistic Competition as a Result of Imperfect Competition. *Quarterly Journal of Economics*, 101:493?511, 1986.

S. Woodcock. Heterogeneity and learning in labor markets. in: Essays on Labor Market Dynamics and Longitudinal Linked Data, Cornell University Ph.D. Thesis, 2003.

Appendix A: Data Construction

Coverage and Quality of Data

Exports Data

This dataset contains all the records that correspond to exports. A line in this dataset can be read as an item (product as classified using the Harmonized Codes to the 6th digit) within a shipment from a seller in Bangladesh to a buyer elsewhere on a given date. As shipments can be multi-product, the dataset is more disaggregated than the level of the transaction.

Together with other relevant information, the most salient variables in this data contain identifiers for the buyer and the seller (see the corresponding section for our work on *cleaning* the identities of the players), a classification and description for the product, the statistical value of the product, its net mass in kilograms and characteristics of the shipment (mode of transport, terms of delivery, ports, countries, currency of invoice, exchange rate conversions, etc.).

The National Board of Revenue compiled the records coming from the different Custom Stations. The data before 2005 was considered of low quality, as comparisons with UN Comtrade sources and reports from BGMEA showed poor coverage of the universe of trade, both on the exports and imports side. This coincides with the migration into the Asycuda system in the main Custom Stations. Discarding the records before 2005 and restricting the attention to the garment sector only, including both woven and knitwear products, the Exports Dataset contains $3,059,844$ observations. The distribution of these over years and custom offices look as follows:

Table 1: Frequencies of Observations over Years, Exports Data

Year	Freq.	Percent	Cum.
2005	250,749	8.19	8.19
2006	321,318	10.5	18.7
2007	319,456	10.44	29.14
2008	388,744	12.7	41.84
2009	352,715	11.53	53.37
2010	507,459	16.58	69.95
2011	486,569	15.9	85.85
2012	432,834	14.15	100
Total	3,059,844	100	

Table 2: Frequencies of Observations over Custom Stations, Exports Data

Station	Code	Freq.	Percent	Cum.
Dhaka	101	288,470	9.43	9.43
Dhaka-K	102	25,724	0.84	10.27
Chittagong House	301	510,978	16.7	26.97
Chittagong - EPZ	303	320,340	10.47	37.44
Chottagong Main	305	1,912,337	62.5	99.93
Benapole	601	1,995	0.07	100
	Total	3,059,844	100	

These correspond to five Customs Stations in Bangladesh: Dhaka Custom House (101) and Dhaka Export Processing Zone (101/1073), Dhaka Kamalapur (102), Chittagong Custom House (301), Chittagong Export Processing Zone (303), Main Chittagong (305), Benapole (601, land).[66]

[66]Note that 301 records as 305 from 2007 onwards.

As the table above shows, the non-EPZ Stations in Chittagong concentrate the vast majority of the observations. Unfortunately, the *raw* data we obtained from Offices other than Chittagong exhibited some limitations. In particular: (i) there is no data from Custom Offices 101 and 102 available for year 2009 or after September 2010; (ii) the information sent from Benepole only covers years 2011 and 2012; (iii) the identities of the exporters was missing for a large proportion of the observations across al Custom Offices, for years 2011 and 2012.

Using aggregated data, we verified that in any given year between 2005 and 2011, the selected Custom Stations process more than 94% of the total exports (in volumes and in values) in garments from the country. Of these customs offices, between 2005 and 2011 the non-EPZ Chittagong (305 / 301) accounts for an average of 90% of the exports we observe. After September 2010, due to a change in the system used to record import and export bills, we only have records collected in Chittagong, the main custom station. Still, for the period September 2010 to September 2012, our data accounts for more than 87% of the exported values in garment in Bangladesh.

Using the years in which data from both Dhaka and Chittagong is available, we corroborated that at the firm level, manufacturers tend to use one or other (set of) Custom Station. For this restricted sample, the proportion of transactions that the firm operates via Chittagong is above 91% already in the 25^{th} percentile and it is zero (meaning all the exports circulate via Dhaka offices) in the 10^{th}. The intermediate percentiles mostly exhibit proportions between 80% and 90%. This implies that each Custom Office seems to be self-consisted when transactions are aggregated at the seller level. Equivalent conclusions were reached when aggregating at the buyer level and at the buyer-seller level. These exercises were performed by quarter, by year and over the whole of the panel, excluding the years for which data from any one of the stations was missing.

Benapole is indeed a very small station, dealing with trade transported via land, which is a negligible choice of mode of transport for the bulk of the trade in the sector we are considering.

Imports Data

This dataset contains all the records that correspond to imports. Again, a line in this dataset can be read as an item (product as classified using the Harmonized Codes to the 6th digit) within a shipment from a *supplier* somewhere in the world and an importer or *manufacturer* in Bangladesh. As many shipments are multi-product, the dataset is more disaggregated than the level of the transaction.

Together with other relevant information, the most salient variables in this data contain identifiers for the importing firm and the country of origin of the shipment, a classification and description for the product, the statistical value of the product, its net mass in kilograms and characteristics of the shipment (mode of transport, terms of delivery, ports, countries, currency of invoice, exchange rate conversions, etc.).

The National Board of Revenue compiled the records coming from the different Custom Stations. After appending the data coming from the different sources, our dataset contains $6,546,504$ observations, of which 0.45% (29,309 lines) constitute partial duplications that are left in the base dataset. The treatment of these differs with the different uses we gave to the dataset, but in all the cases, our calculations are free from distortions induced by these duplications.

As with the exports dataset, the data before 2005 was considered of low quality, after poor comparisons with UN Comtrade sources and reports from BGMEA. However, the observations corresponding to 2003 - 2005 were left in the dataset, for the purpose of cross-checking some of our assumptions in the matching inputs-to-outputs procedure. The distribution of the observations over years and custom offices look as follows:

Table 3: Frequencies of Observations over Years, Imports Data

Year	Freq.	Percent	Cum.
2002	23	0	0
2003	6,257	0.1	0.1
2004	195,618	3	3.09
2005	412,461	6.32	9.42
2006	488,640	7.49	16.91
2007	520,804	7.98	24.89
2008	797,125	12.22	37.11
2009	1,044,918	16.02	53.13
2010	950,635	14.57	67.7
2011	1,279,179	19.61	87.31
2012	827,898	12.69	100
Total	6,523,558	100	

Table 4: Frequencies of Observations over Custom Stations, Imports Data

Station	Code	Freq.	Percent	Cum.
Dhaka	101	1,003,614	15.34	15.34
Dhaka-K	102	412,145	6.3	21.64
Chittagong House	301	3,744,752	57.23	78.87
Chittagong - EPZ	303	184,107	2.81	81.68
Chottagong Main	305	641,487	9.8	91.48
Mongla 1	501	22,480	0.34	91.83
Mongla 2	502	54	0	91.83
Benapole	601	534,691	8.17	100
Total		6,543,330	100	

As in the case of the exports, the non-EPZ Stations in Chittagong concentrate the vast majority of the observations. Data from these Custom Stations is available for the whole of the period of the panel. Unfortunately, for the rest of the Offices we face the following restrictions: (i) Dhaka Custom Offices (101 and 102) only report information from 2008 onwards; (ii) information coming from the EPZ in Chittagong is as well only available from 2008 until the end of the panel; (iii) like on the exports side, data coming from small custom offices is only available for part of the period covered (2008 and 2009 for 601 and 2010 - 2012 for Mongla).

Like on the exports side of the data, the most worrying restriction is that of the missing data from Dhaka before 2008, for the purpose of doing firm-level or relationship-level analysis over time exploiting the whole of the 2005 - 2012 period. Again, the garment manufacturers that we observe on the exports dataset tend to use one or the other (set of) Custom Station almost exclusively for their imported inputs. Exercises equivalent to those performed with the exports data at the importer level, by quarter, by year and over the whole of the panel, excluding the years for which data from any one of the stations was missing, confirmed this conclusion.

A potential issue of concern would be the missing data from Benapole. This is a Custom Station that deals with in-land commerce and it is almost fully dedicated to imports coming from India. The missing data before 2008 and after 2009 could be a problem if significant volumes of garment-relevant inputs were coming through this custom office. The product codes imported through Benapole over the period we observe correspond 87% of the times to categories that are not related to garment.

The remaining 13% could potentially be related to garment (mainly chemicals and dying products) and half of the times, these imports correspond to firms we identify as garment exporters. For this reason, when working with the import-export matched data, we have accounted for the fact that some manufacturers might be sourcing via Benapole outside the observed 2008 - 2009.

As we are interested in the imports that are related to the RMG sector only, the universe of the imports into the country is not as relevant. To select the right product categories , all the imports undertaken by garment exporters (whose identities are obtained from the exports dataset) were analysed. All the product categories at two digits of aggregation that were imported by garment manufacturers were kept in the data, irrespective of the identity of the importer.

Of the 6.5 million observations in our data, less than 27% correspond to imports performed by our garment exporters. However, considering the universe of transactions in the product categories that the garment exporters import, we have almost 90% of the observations in the original dataset. For completeness, we keep all the product categories and flag the non-garment-relevant 10%+.

Data Management and Transformations

Prices and Quantities

Statistical values for the shipments, both for exports and imports, are already present in the data. According to the information we obtained from the NBR, these statistical values are calculated using the data in the bill of entry or export directly: taking the FOB price, converted into BD Takas at an exchange rate that the Central Banks provided every month or daily, depending on the year. If no insurance is specified in the bill, it is computed as 1% of the FOB. Similarly, if freight is not included, it's computed as 1% over the (FOB + insurance). Landing charges are computed as 1% of (FOB + insurance + freight).

In the data, we always observe the *mode* of the transaction -i.e., FOB, CIF, CNF, etc.-, the value in the invoice and the currency of the invoice, the exchange rate and the statistical value. Using the details above, we were able to reconstruct one or other record (invoice or statistical value) consistently in the best part of the data.

For many of the calculations in this project, quantities and prices were winsorized transforming the top and bottom 0.5% of the quantities and values within each HS4 product category.

Firm Identifiers

After the procedures described below, the data was anonymized to mask the identities of both international buyers and local manufacturers.

Sellers' Identities

The dataset, as constructed from the records in the Custom Offices, identify the exporters using the Business Identification Number (BIN) of the firm. This constitutes a 10 (or 11, in the new system) digits number. The first digit corresponds to the Commissioner to which the productive activity is settled (not the administrative location). Firms that have productive activities in two different locations corresponding to different Commissioners are assigned one of the two by the National Board of Revenue, according to the size of the business in each location. Each Commissioner is divided into circle offices (for example, in Dhaka there are around 30 circles) and the second and third digits in the code correspond to the circle in which the productive activity of the firm is located. The fourth digit corresponds to the tax category of the firm, according to its yearly turnover. This is re-assessed at the end on each fiscal year, which might lead to changes in the BIN number for the firm (the whole number changes, not only this digit). The main categories are 1: VAT, 2: Turnover, 3: Small Cottage Industry, 4: Others. Digits five to nine are the actual firm identifier within the National Board of Revenue (NBR) and it is assigned by the circle processing the application. The tenth digit is a number coming from a random numbers generator to avoid duplications.

The main complication of using BINs as firm identifiers was that the firm code (digits 5 to 9) is not necessarily unique across plants under the same ownership structure. One ownership structure might register different divisions within the same firm under different BINs, for tax purposes, inducing misidentification of the sellers. Moreover, the same plant could potentially have - completely - different BIN codes over time, if its turnover bracket or location change. Therefore, over time a firm whose essential characteristics remain unchanged might change BINs to obtain tax incentives or fall under special subsidies schemes offered by the government. The information in our *raw* data didn't allow us to spot one or the other misidentification issue.

We dealt with these two concerns in five ways, generating an alternative (to BIN codes) firm identifier that was used in the study for robustness checks.

First, using data (up until 2010 only) from the VAT Office within NBR, each BIN number in our dataset was matched with the name of the firm, its address and contact details. Whenever two different BIN codes were matched with the same firm name and address, these were unified to be considered the same firm.

Second, we matched the BINs in our dataset with the database that Bangladesh Customs holds on updates of BIN codes for all the exporters and importers. In this dataset, each entry contains an *old* BIN code, a *new* BIN code, the name and address of the firm. Most of the code migrations are associated to switches from 10 digits to 11 digits codes. We crosscheck the information in this database with our dataset and there's an overlap of 110 firms, whose identities are unified as appropriate. However, these coincide with unifications done in the previous step.

Third, we were able to crosscheck our data with the lists of Members and Associate Members of the Association of RMG Exporters. The Association is the competent authority for processing the necessary applications for export / import permissions and tax reliefs, so they maintain accurate records of the identities of the exporters, including ownership data in some cases. Woven sellers outside Export Processing Zones are bound to be registered with the Association, while knitwear-only exporters and the small fraction of firms in Export Processing Zones can also use other channels for exporting / importing (BKMEA and BEPZA). Using the data from the Association, the original number of 7033 distinct sellers observed in the panel before September 2010 was brought down to 6027 firms. The identification of firms was done in stages using the correspondence between BINs and internal codes of the Association, matches in the names and addresses and coincidences in the BINs and documents produced in the applications for exports permissions. This procedure was found to unify within the same identifier, different BINs exporting at the same time and different BINs over time.

Fourth, we used the Bond Licence Numbers in the dataset we obtained from the BOND Commissionaire to unify BIN numbers that held the same Bond Licence, as plants that share bonded warehouse facilities under the same licence are typically part of the same ownership structure.

Finally, we explored the trajectories of all the firms appearing in the panel within a suitable time-window after a seller drop out [67] . The idea of the exercise was to check whether the characteristics and trading patterns of a new firm were *similar enough* to those of a dropping seller, to suggest they could actually be the same firm. The key aspects that were analyzed were the timing of the death and births, the location of the firms, the main products and volumes exported and the main buyer for each of these. Using these criteria with different weights assigned to each factor, we found no strong evidence to impute the same identity to any two firms.

As a result of the steps above, we have the BIN codes as *conservative* plant identifiers and an alternative identifier for the firms using the unifications above. Most of the exercises done using this data were carried out using both identifiers, as a robustness check. For the purpose of these project, when not specified otherwise, we will refer to a firm or a plant as units identified with their BINs.

Buyers' Identities

At the most downstream level, we have information on the firms located elsewhere buying ready-made garment from Bangladesh. Strings containing names and addresses of these firms - buyers from now on- where introduced manually in the system that originated our data. Spelling mistakes, varying criteria across Custom Offices or over time and differences in administrative procedures induced difficulties in the mapping of transactions to well delimited unique buyers. Using the names recorded as identifiers, the raw data contained nearly 340,000 different buyers pooling all the years together. After a *cleaning* procedure focussed on correcting the mistakes mentioned above, the list went down to about 7,000 different buyers and a pool of small firms collected in a broad category of firms for which the cleaning was not possible.

The *cleaning* procedure was done in stages.

First, using the uncleaned strings, the names of the (almost) 1,000 largest buyers were manually cleaned.

Second, using these strings, one relevant substring was chosen for each of them and the whole of the data was scanned to find matches in the uncleaned strings [68]. In this procedure, almost 80% of the matches were unique and, after the relevant controls, the names were corrected. the multiple matches cases were analysed one by one and, when suitable, replacements were introduced accordingly.

Third, the remaining uncleaned strings were modified to discard strange characters and unify expressions such as "INT.", "INTL.", "INTERNATIONAL", etc.. The scanning routine was performed again over these modified strings.

[67] Time windows were set at 1 month, 3 months, 6 months and 9 months and finally using the mean/median with two standard deviations / median absolute deviations of the gap between transactions for the dropping firm.

[68] When the clean name of the firm wouldn't render a compelling substring, the related uncleaned strings were leaned manually. A - fictitious - example of this would be a clean name like "THE COMPANY INC.", whose possible substrings would generate matched with clearly unrelated firms.

Fourth, from the remaining uncleaned lines, the largest (almost) 1,000 buyers were selected and the first and third steps were repeated. Fifth, with now 2,000 identified clean names, the remaining uncleaned strings were scanned now allowing for: (i) a spelling mistake involving one character only (including one missing or one extra character); (ii) one parsing mistake, involving one extra or one missing space only; (iii) two parsing mistakes of any kind; (iv) a character swapping, involving two characters; (v) combinations of (i) and (ii); (vi) combinations of (i) and (iii); combinations of (iii) and (iv); combinations of (iv) and (i). The more conservative scans performed really well, allowing for corrections in most of the remaining uncleaned lines. Criteria (v) and above produced multiple potential matches and only 15% of these were used to introduce corrections. At this stage, about 90% of the lines in the exports dataset ad a clean name for the buyer. Robustness checks of these stage were performed exploiting the buyers addresses and a soundex, to identify matches of names that "sounded" similar.

Fifth, a large number of line-by-line corrections were introduced, using a *.do* file that contains over 100,000 replacement statements.

Sixth, the denominations in the Euromonitor Data were used to unify identities that showed in our data sometimes using local denominations of a brand, global denominations of a brand, national denominations of the firm or group or the name of the parent company in cases of joint ownership.

Seventh, publicly available company reports of the 15 largest buyers were explored to correct identities of firms in the presence of mergers and acquisitions. Using this source of information, 9 relevant changes were introduced.

Eigth, for the top 100 buyers, the patterns of trade were observed, with special focus on the volumes of trade, product categories and destination of the shipment to spot miss-imputations.

As a result of these steps, 96% of the lines in the exports dataset, explaining 97% of the traded values, have a *clean* name for the buyer.

Comparison with Comtrade Data

To better assess the representativeness and robustness of the coverage of our main datasets, we compared our records to those in the UN Comtrade database[69].

In broad terms, disagreement with UN Comtrade Data is expected to occur due to a number of reasons:

- After received from the national authorities, data is standardized by the UN Statistics Division, using Comtrade standardization protocols that can induce discrepancies with the data we have from the National Board of Revenue.
- The Comtrade data might feature records coming from different sources of information.
- In the Comtrade data values of disaggregated commodities do not necessarily sum up to the total trade value at higher levels of aggregation. This is mainly due to potential restrictions in disclosure from the reporting country.

[69] DESA/UNSD, United Nations Comtrade database.

- The time-wise coverage of our data and that in Comtrade differs.
- Our data comprises records of the largest Custom Offices but not the universe of trade with Bangladesh.
- Product classification criteria might differ.

More specifically, the only overlap our panel has with the data available in the UN Comtrade database is for years 2005 to 2007[70]. Also, the Explanatory Notes in the Bangladesh section of the Comtrade search engine reports that all the data corresponding to 2005 and 2006 for Bangladesh was obtained from FAO, while that of 2007 has the Bangladesh Bureau of Statistics as ultimate source of data.

All the three years are presented by Comtrade with a note stating that "Data for this year has been re-processed to make correction to the data". All the imports are reported CIF and exports are reported FOB. While our data features the Customs daily (or monthly in some cases) exchange rates to convert foreign to local currencies or viceversa, Comtrade data uses a fixed currency conversion rate (from Bangladesh Takas to US Dollars) is used for each year, according to the following rates:

Table 5: Currency Conversion Table, Comtrade Database (BDT to USD)

	Flow	
Year	Import	Export
2005	0.015544	0.015539
2006	0.014493	0.014495
2007	0.014521	0.014521

For the purpose of the comparisons of values, we unified the exchange rate conversions to use those reported by Comtrade, to avoid discrepancies induced by currency rates.

Product classifications also show minor mismatches when data from both sources are matched at the product - time levels. When the match is performed at 6 digits of aggregation in the HS codes (2002 or as reported), there are three product categories (611512, 611520, 611691) that are present in the Comtrade Data and that we don't have in the Customs Data (for years 2005 or 2006) and there are other codes (610310, 611510, 611522, 611529 , 611530, 611594, 611595, 611596) in the Customs Data (2007 only) that don't show in the Comtrade Database.

An additional source of discrepancy in the comparison of volumes is that 98% of the data obtained from the Comtrade dataset has the traded volumes (in kilograms) estimated, possibly from the quantities traded reported in an alternative measurement unit. No details on the estimation procedure are available, but while the comparisons of volumes recorded in both Datasets looks very weak, that in values performs relatively well. This might be evidence of discrepancies of information across sources due to the manipulation of quantities in the Comtrade data.

Given that the Comtrade at different aggregation levels doesn't necessarily match, the comparisons at 4 digits are done using the reported UN Comtrade data rather than the aggregation of the subcat-

[70] As explained above, our records before 2005 were considered of low quality and not used for any part of our analysis, except when stated in the procedure of matching inputs and outputs.

egories. Only when the data at four digits is not available, the aggregation of six digits categories is used.[71]

Table 6: Ratio of Exported Values (USD, Comtrade currency conversion) Comtrade Data / Customs Data: Four Digits Codes

	Year			
HS4	2005	2006	2007	Total
6101	0.991	0.912	0.955	0.953
6102	0.863	0.957	0.976	0.932
6103	0.957	0.938	0.984	0.960
6104	0.986	0.911	1.285	1.061
6105	0.999	0.919	0.943	0.954
6106	0.996	0.909	0.883	0.929
6107	0.995	0.946	1.020	0.987
6108	0.971	0.951	2.400	1.441
6109	1.002	0.904	0.896	0.934
6110	0.988	0.975	0.990	0.984
6111	0.994	0.857	0.954	0.935
6112	1.000	0.902	0.924	0.942
6113	0.984	0.988	1.292	1.088
6114	1.001	0.888	0.800	0.896
6115	1.000	0.932	0.957	0.963
6116	0.999	1.018	1.000	1.006
6117	1.000	0.979	1.495	1.158
6201	0.932	0.944	1.088	0.988
6202	0.883	0.926	1.024	0.944
6203	0.966	0.927	0.978	0.957
6204	0.964	0.916	1.024	0.968
6205	0.992	0.936	0.907	0.945
6206	1.004	0.919	1.009	0.977
6207	0.976	0.945	0.976	0.965
6208	0.996	0.952	1.048	0.999
6209	0.965	0.956	0.912	0.945
6210	1.014	0.986	1.168	1.056
6211	0.961	0.862	1.061	0.961
6212	0.903	0.926	0.913	0.914
6213	1.000	0.953	1.370	1.108
6214	1.000	0.960	1.351	1.104
6215	1.001	0.609	1.876	1.162
6216	0.998	0.950	0.996	0.981
6217	0.978	0.953	1.346	1.092

Matching Datasets

One of the most interesting aspects of our datasets is that they allow for the matching of exports and imports at various levels of specificity. The most disaggregated one, involves matching imports and exports for a given *order*, placed by one buyer to a given manufacturer and this will be possible in a sub-sample of our data. At the other end of the spectrum, imports and exports can be matched broadly for each firm (manufacturer) in a given time period. Within these two extremes, we can perform different matching protocols.

What follows is mainly concerned with the matching procedure at the order level. I first explain the administrative device that allows for these matches and I then describe how this is expressed in our datasets. Then, I explain the procedures we followed to *clean* the relevant variables in the dataset to perform the matching. I finally compare alternative matching procedures and explain a number of refinements that we have gone through.

[71] For brevity, I omit here reporting the equivalent exercise on the imports side of the data. All files available on request.

The Utilisation Declaration Procedure

Broadly speaking, the Ready Made Garment sector in Bangladesh relies heavily on imported inputs. This is even more the case for woven garments, where the local production of suitable fabrics and input materials is still negligible. For this reason, imported inputs for the production and exports of clothing are given preferential treatment by Customs (for details, see Thomas's section in (De Wulf and Sokol, 2005)). First, in order to help reduce the lead time in export orders in the garment sector, the clearance of textiles and other garment-relevant inputs is done within two days whenever possible, much quicker than the up to seven days for other imports. Second, manufacturers can establish bonded warehouses to facilitate the import, storage during clearance and transit of inputs, including fabrics, accessories, dyes and chemicals and yarn.

The most commonly used Customs Procedure in garment exploits what is called the Special bonded warehouse (SBW) facility. In practice, raw materials used in the production of RMG export orders are imported duty-free into SBWs, manufactured into finished articles of clothing, and exported.

In the data we obtained from the Bond Commission, we identify 5,811 Bond Licences[72], out of which almost 5,000 were active in 2012, with the rest having been suspended, canceled or closed. The universe of licences we observe are matched to 5,377 firms in our dataset, identified by BINs[73]. Of the licences that correspond to garment exporters or related importers of raw materials (more on this later), 64% are associated to SBWs, 18% to Private Bonded Warehouses (PBW), 2% to EPZ Warehousing and the remainder has no specified type, as far as we could classify. These correspond to older licences (some of which are no longer active), opened during the nineties, when the Licence Number included no code identifying the type of facility. Restricting the attention to those GBO codes that correspond to garment firms, 80% of the licences are SBWs. The vast majority of the SBWs are located in the Chittagong port area, with Dhaka being the second location in importance.

In simple terms, to take advantage of the tax exemption, a firm holding a Bond Licence, after receiving an order from a buyer, opens - if needed - a Letter of Credit in the manufacturer's bank (in occasions, this will be a back-to-back Letter of Credit). To produce for that order, the manufacturer is allowed to import raw materials duty-free for a value equivalent to up-to 75% of the value of the export order. In every import shipment under the umbrella of that order, the manufacturers declares a Utilization Declaration (UD) number, issued by the Bangladesh Garments Manufacturing and Export Association (BGMEA) or corresponding alternative association and the specifics of the import transaction are recorded in the bonders and customs™ *passbooks*, which act as a record of stock going into the SBW. This procedure takes place with every import shipment within the relevant order or UD. Likewise, every export shipment that corresponds to the UD, is recorded under the same UD number for clearance. The customs modernization project that started in 1999, introduced an electronic tracking system of the goods in the SBWs facility, enabling both bonders and Custom Stations to retrieve accounts on flows into any individual SBW to reconcile this with physical movements of inputs and outputs, without relying on passbooks.

The evaluation of the suitability of the exception is almost entirely down to the industry Association,

[72] This document is required for the access to any type of Bonded warehouse. Moreover, each licence includes the specific codes the licence holder is allowed to import and any additions follow a specific request of permission from Bond, with support from the relevant Industry Association.

[73] The almost 500 licenses we don't match with our data can correspond to EPZ firms in other sectors or textile companies that are not exporting garment. A quick exploration of the firms names, show textiles and packaging as the most common activities of the licence holders that are not in garment.

after the Duty Drawback Authority outsourced this activity. The key role of the Association is to control that the input-output coefficients implied in the proposed Utilisation Declaration is adequate, to restrict abuses to the system. It is expected that the ASYCUDA++ system internalises in the future the UD formula for calculating the coefficients of utilization of raw materials to finished articles, to automatically track the goods flow. At the moment, the ex-post control is done in the clearance stage.

The UD document contains all the information needed for calculating the input-output coefficients and for their evaluation in the Association. The main body of the UD specifies the characteristics of each item in the order (size, style, etc.), the destination country, the quantity and the unit value. For each of these items, there is a chart that specifies the inputs (fabrics and accessories) required, whether imported or domestically supplied, including its value and characteristics and, in most of the cases, the firms that are going to supply these inputs. Based on this, the BGMEA experts assess the output-input consistency and approve the order or recommends amendments. The second part of the UD contains information on the LCs and local inputs. First the number and date of the Export LC opened against the corresponding order is shown, with its value and the estimated shipping date. Second, the LCs for imported inputs are described, showing the country, LC number, date and value, per LC. Third the LC for local inputs are listed. The third part of the UD contains a detailed description of the items in the order: description of the garment, quantity, measures per size for the whole size chart. Then, a table presented all the required inputs, shows the imported inputs separated from domestically sourced items and each of these are divided into fabrics and yarn on one hand and accessories on the other. Within each of these categories (for example, imported fabrics or domestic yarn), per item in the order of the UD, there is a description of the input required, the total quantity and the input requirement per unit-piece of garment in that item (fabrics only), together with the supplier's name, address and country (if appropriate).

UDs constitute the main document that allows firms to claim for the duty exception, and are issued at the discretion of the Association. BGMEA gathers the majority of the garment exporters. Exporters of knitwear only, can alternatively obtain their UDs through the Bangladesh Knitwear Manufacturers and Exporters Association (BKMEA). Garment manufacturers based in Export Processing Zones follow a similar procedure under the EPZ Association (BEPZA), to access EPZ Warehouses. A manufacturer can submit a UD application only if it is a regular member of the Association and non-members need to register before applying. BGMEA members pay a fee of 450 to 650 BDTk (slow or fast track) per application they submit. This Association receives around 500 UDs a day (plus 300 amendments), and takes from hours to a couple of days to have the process finished.

To the UD application, the manufacturer needs to attach a number of documents collected in a Scrutiny Sheet. This contains the date of submission and corresponding port, the name of the firm, the BIcode (unique identifier for the firm given by BGMEA or BKMEA), the Bond licence number, the associated UD number and the name of the buyer that placed the original order. The UD number will in general follow the structure BGMEA / DHK / YY / XXXX / ZZ. BGMEA corresponds to the Association issuing the UD. Alternatively, it can read BKMEA. The second part, DHK, corresponds to Dhaka, and can alternatively be CH for the Chittagong Office. YY corresponds to the year of submission (and must coincide with the date of submission in the header). XXXX is the BIcode that BGMEA assigns. BKMEA has its own system of codes and the two institutions have completely separated numbering systems. For this reason, the same XXXX in two UDs can be identifying two different firms, one registered with BKMEA and the other one with BGMEA. ZZ corresponds to the number of UDs the firm has placed in year YY. The first UD that the firm submits in the year will take ZZ = 01, the next one will be 02, and so on.

Given the above, a given UD number to be quoted in imports and exports under an order uniquely identifies all the transactions associated to that specific order. Then, all the export shipments and imports of inputs that correspond to an order placed from a buyer to a manufacturer are (in principle) recorded in the corresponding Customs Station with its UD number and this is the administrative device that allows for matching imported inputs to outputs, at the order level.

The Records in our Dataset and the Cleaning Procedure

The UD number that would allow for the input-to-output matchings at the order level as described above, is recorded in every custom office with various levels of detail and coding problems. In general, the issuing association is not recorded and only the numeric components of the UD number are present. This was potentially problematic in two ways. First, due to the fact that BKMEA and BGMEA do not coordinate the assignment of codes to firms, the first concern was that a given code could refer to two different firms, as explained above. Second, firms in Export Processing Zones, exporting through custom stations 101/1073 (DEPZ) and 303 (CHEPZ), have their UDs issued by BEPZA. Although the BEPZA numbers have a slightly different structure from that of the UDs, coding mistakes in the Export Procedure in non-EPZ-dedicated Offices could induce incorrect extractions of the UD numbers.

Issuing institution aside, we detected a lot of variation across Custom Offices on what they record in the UD field. In many cases not all the numeric components of the UD are included. An additional complication was that the order in which these components show might differed across records, making any simple extraction routine unsuitable. Other problems we encountered were associated with relevant numbers different from the UD number, but with similar structures (like dates, Bond Licence Numbers, Export / Import Permission Numbers) being coded by mistake in the UD field. Because in theory a UD can cross over different Custom Stations, differences in data entry protocols across Offices can also induce problems in the matching. We recognise that across-offices UDs are not very common, but this was still one of the concerns when writing our matching routine.

Finally, the field recording the UD number being a very flexible string space, various sorts of data-entry typos and mistakes were found.

In this context, the first challenge to merge the datasets was to extract the components from the string that identify a UD. These are: the year in which the UD is issued, the code for the exporter given by the issuing Association and the order number. After clearing the main string from strange characters, the transactions were split into Custom Offices and Extended Procedures combinations, to identify common patterns of recording the information. In occasions, breakdowns over time (years) were also necessary. For each sub-group, the main string was split into components using the most common parsing characters ("/" or space). This gave 1 to 10 components for each string. Then according to the observed patterns, the three components of the UD number were extracted following a protocol that was in most cases specific to the year, Procedure and Custom Office combination. The code implementing these extractions on the Imports Data and the Exports Data is extense and available upon request.

The result of that initial procedure was a first version of the cleaned UD numbers. A number of robustness checks, imputations and corrections were performed both on the imports and exports sides of the data. For the purpose of the description of these steps, we focus below on the exports side, which for various reasons was more complicated than the imports side.

Of all the observations in the exports dataset from 2005 onwards ($3,059,844$), 13% contain a missing value in the field collecting the UD number. Out of the non-missing lines of the whole of the data) we managed to recover a *proper* UD number for the vast majority of the observations (86%). The cleaned UD numbers are correctly distributed over Custom Offices, with 92.7% of these falling in the non-EPZ Chittagong Offices, 7% in Dhaka under non-EPZ Procedures and the remaining lines (6 cases) found in EPZ Stations. Similarly, over Ectended Procedures, we corroborated that 97.7% of the cleaned lines fall under the code that corresponds to the use of SBW, and the rest were distributed over Procedures associated to re-exports or direct exports.

The second stage of the cleaning procedure involved the following amendments and robustness checks.[74]

Items within the same transaction: Two different products within the same transaction should belong to the same UD. As items in the same transaction record are associated to one invoice (pro-forma and final) they need to correspond to the same UD. Therefore, we first explore the lines in the export dataset whose UD information is missing but that belong to a shipment where at least one item has a proper record for the UD number. There are $27,858$ cases under this category. In these cases, we impute the UD number of the line with UD record to all the lines in the shipment with missing information. This is one of the imputations that was not carried out on the imports side of the data, as the unique UD per transaction does not necessarily hold.

Different UD numbers within the same transaction: Similarly, it cannot be the case that within the same transaction, two different UD numbers are quoted. We have only 52 cases in which this happens and the discrepancy between UD numbers can be in one or many of the components of the UD identifier, i.e. the exporter code, the order number or the year. Discrepancies in the exporter code are solved in the following steps. Discrepancies in the year are often due to coding errors and were manually corrected. As a general rule, we make the decision of keeping, for the whole transaction, the earliest recorded year as the year of the UD. When this is not possible, we keep the year closes to that in which the transaction takes place.

Exporter Codes for BGMEA firms: Approximately 62% of the firms for which we managed to obtain at least one 'clean' UD record are present also in our complementary BGMEA dataset. The rest of the firms with UDs could be under the BKMEA orbit or might be exporting with a BIN code different to the one used to register with the Association. In fact, the vast majority of the export trasactions that didn't produce a match with the BGMEA data are classified into HS codes that fall in knitwear categories.

We use the list of firm identifiers and BGMEA internal codes to check the exporter code component of the UD numbers. For more than 92% of the sellers present in both datasets, the exporter code recovered in our routine coincides with the internal code that BGMEA provided us with. This, in

[74] Again, all .do files are available upon request.

turn, implies that less than 9% of the transactions have an exporter code in the UD that doesn't coincide with the BGMEA internal code.

The majority of these cases were connected to data entry problems (lack of parsing characters separating the components of the UD numbers, miscoding, etc.) and they were amended appropriately. In the case of a handful of sellers, the same BIN code seems to be exporting using a firm identifier in the UD coinciding with the BGMEA code and one or more additional codes that are observed systematically. Four of these companies were found to have two different codes assigned within BGMEA, corresponding, probably to two different business units. These were left unchanged.

In all the remaining cases in which the BGMEA internal code didn't coincide with the firm identifier in the UD that were not solved as explained above, were evaluated case by case. If no data coding problems were observed, the UD number was left unchanged, despite the incongruence with the BGMEA record. The exports transactions associated to these cases were mainly in knitwear categories, suggesting that these were likely to be BKMEA firms as well. All changes were done preserving the UDs that, unchanged would generate a match on the imports dataset.

Different Exporter Codes for the same seller: As mentioned above, because a manufacturer can very often use a sister company (or a specific division within the company) to open the UD process within BGMEA, many sellers, as identified by their BIN code, can have different exporter codes in the UDs. Therefore, it is not problematic to observe the firm-specific component of the UD varying for the same seller. However, around 5% of the sellers (not only associated with BGMEA now) show multiple exporter codes in their UDs that vary in a 'suspicious' way (consecutive numbers over time, is the most common pattern or codes that seem to relate to different containers in the shipment). Those cases, are corrected using the BGMEA codes as described above when appropriate and using the codes on the import dataset, whenever possible. If none of these produce a set of UD numbers consistent for the seller, the information in the UD field is discarded.

Different buyers within the same UD: As each UD is opened against a buyer's order, theoretically it cannot be the case that a given UD has two different buyers.

There were a number of UDs under the names of more than one buyer (19% of all transactions with identified UDs). The vast majority of these, corresponded to orders in which both a retailer and a trader or a logistic company showed as the buying company. Those cases were corrected substituting the identity of the trader for that of the retailer (for the purpose of the matching only). The main exception to this imputation was in the case where the trade shows in more than 50% of the transactions within the UD. After these corrections, almost 97% of the UDs have a unique buyer. The rest of the UDs, then involving more than one buyer were removed from the analysis (flagged as non-valid UDs), as it was hard to distinguish cases of data entry error in the name of the buyer of cases of data entry error in the UD number.

Non licensed firms: Using the Bond Licences data, we explored whether lines for which we had cleaned UDs corresponded to firms (BINs) that had a valid Bond Licence. The type of mistakes we wanted to rule out was the cleaning of UD numbers for the original string for cases in which the procedure was miss-coded and the firm was not bonded. Fortunately, we found no cases of this type.

Cases where a date is available: In occasions, the original string would include a date, that is presumably related to the date in which the UD procedure was opened or the date of approval of the last amendment. Whenever possible, the year extracted as part of the UD number was checked against the recorded date. A handful of year mis-imputations were corrected.

Two-components matchings: There were cases in which the only two components were extractable from the original string. In most of the cases, the missing component was the order number. Some of these missings were originated in cases the information was not present in the original string and some others corresponded to lines in which many of the extracted substrings could constitute the order number (often, when amendment numbers or dates were attached to the UD number). These cases were merged using only the two available components with the imports dataset to evaluate whether any of the orders on the imports side, for the same year and exporter, could inform the third component of the UD number. Where the matches were unique, this is the year and exporter on the imports side had only one order to match with on the export side, the order number was imputed if three conditions were satisfied: (i) the weight ratio of input to output was within product-specific reasonable bounds; (ii) the material of the inputs was not at odds with the output (i.e. synthetic fabric is not imported to produce pure cotton shirts) and (iii) the time window of the exports and imports fall within a quarter. In the cases where more than one match was produced and there were candidate substrings extracted from the original string on the exports side, scanning the candidate substring with the potential orders rendered a unique choice of number to impute as order.

These amendments had little impact in the overall matching but helped guarantee that there are no major omissions in the the datasets that we used for work on matched data, due to failing to march export lines.[75]

Consistency at the Buyer-Seller-Product level: Due to the initial condition of the variable that records the UD number and the various cleaning routines we performed to recover the codes we need for the matching, one of our concerns was to have isolated *lines* - an item in a transaction - associated to a UD that we were not able to recover from the original variable. If that was the case, when computing statistics at the level of the order from a buyer to a seller, we would not be accounting for some of the transactions within that order.

For this reason, we explored the set of transactions for which we were not able to recover the UD number. We first assumed that whenever the original variable collecting UD numbers was missing,

[75] In the whole of the exports dataset, there were 1.2 lines that didn't produce any match via UD numbers with imports. For these lines, a two-part matching was attempted, just using coincidences in the year and firm ID components of the UD number. From this matching procedure, 65% of the unmatched lines remained unmatched, meaning that the firm ID and year combination didn't have a match on the imports. Of these unmatched lines, we found that: (i) the vast majority of the unmatched lines have either of the two components of the UD number missing, and of these, 20% fall outside of the selected product categories, another 10% corresponds to lines that either belong to unsuitable sellers as defined belong or belong to Dhaka and the rest of the unmatched UDs with at least one component missing, are 99% of the time in the group of lines in which the original string for UD extraction is fully missing or the data in it has a format that suggests a document that is NOT a UD. (ii) 13% corresponds to cases in which none of the year and ID component of the UD are missing and still, a match was not found; of these, 95.5% of the cases fall into cases of: Dhaka custom offices, UD outside the set of selected categories, seller is unsuitable, no information present in the original string or the information the string is likely correspond to a procedure different from the UD. This implied that little improvement is possible over the lines that were not matched two-piece-wise. Therefore, we focussed only on the 35% of the lines that didn't have three-piece matches but that formed at least one match in the two-piece procedure (with no missings in any of the two matching components). Out of these, 20% of the lines correspond to selected product categories (88 thousand approximately), of which only 78,317 were outside Dhaka. Out of these, 68,139 corresponded to suitable sellers and finally, only 46,919 were uncensored. As a result, only 3.6% of the overall unmatched exports was material we could work with to improve upon the matching.

the transaction was not associated to any UD at all. This is consistent with what we observe in the data (missings in the relevant variable coincide with Custom Offices were no UDs are used or with -almost- one off transactions in a buyer-seller pair) and with the conversations we had with the technicians at the National Board of Revenue. Then, the lines subject to the risk we refer to above were those for which there was non-missing information in the relevant variable, but for which we still didn't manage to clean a UD number.

Aggregating the transactions at the buyer-seller-product level, we first explored the proportion of transactions that having non-missing information in the UD field didn't have a proper UD number. The ratio of transactions with uncleaned UDs to all the transactions with non-missing data is zero almost everywhere for all the buyer-seller-product triplets. Less than 12% of these triplets have a non-zero ratio and the vast majority of these, have a ratio equal to 1, meaning that for that specific buyer-seller-product triplet no information (on any transaction) was recovered at all from the UD fields. These are largely explained by the firms that operate in Export Processing Zones and export through *normal* custom offices (under the EPZ procedure), and that record a BEPZA number in the UD field.

For robustness, all the ratios that were strictly greater than zero and strictly smaller than one, led to the following robustness check. For each buyer-seller pair we ordered all the exports in each product, chronologically. For every transaction with no UD we explored whether the buyer-seller pair had a valid UD featuring the same product, active in a reasonable time-window with respect to the transaction with no UD. We studied this using five different time windows: a) a fixed window of 30 days; b) a window equal to the average gap between transactions in the candidate UD; c) same as b but allowing for one standard deviation from the mean; d) a window equal to the average gap taken over all the transactions between the buyer and the seller on that product; e) same as d but allowing from one standard deviation from the mean. Although this procedure would have induced some imputations of isolated transactions into clean UDs, we decided not to perform these corrections as miss-imputation carried the potential risk of introducing noise in genuinely clean and complete orders. The 4% of the buyer-seller-products for which the ratio of uncleaned UDs to all UDs is left unchanged.

Appendix B: Recovering Suppliers' Types

The definition of the manufacturer's quality as a shifter to its demand conditional on the price at which it sells has been used extensively in the trade literature (**?**), Monarch (2014), **?**)). However, it has been recognised that some buyers demand more that others, for manufacturers of equal quality. This suggests that the dispersion in traded quantities in a product-time combination can be driven both by a manufacturer effect and a buyer effect. I recover suppliers' shifters, as a measure for qualify, fitting a linear model with additive fixed effects for buyers and sellers, in the tradition of Abowd et al. (1999). As in comparable exercises using employer-employee matched data, the approach here relies on an assumption on strong separability of these effects and on non-endogenous mobility of suppliers across buyers. I discuss these two in turn after presenting the estimation strategy.

Denoting suppliers with s, buyers with b, quarters with t and product categories with m, we can describe trade volumes at the $-sbmt$ level, q_{sbmt}. For the vast majority (85%) of the almost $40,000$ $-st$ combinations in the data, a main product category accounts for more than 87% of the seller's exports. Likewise, a single main buyer explains at least 88% of the manufacturer's volume. For each $-st$ combination, then, I will focus in the main buyer and product, such that there are a unique $-m$ and a unique $-b$ per seller-quarter.

From variation in q_{st}, the manufacturerer's average deviation of the expected quantity, conditional on the price, the buyer, product and time, reveals the supplier's appeal. I will refer to this as the supplier's quality or type.

$$q_{st} = \theta_{b(st)} + \theta_s + x'_{st}\beta + \eta_{st}$$

The object of interest here, θ_s is then a set of seller-specific characteristics, that are appealing across all buyers. This can include the seller's reliability, its commitment to timely delivery, its managerial protocols dealing with buyers, etc.. $\theta_{b(st)}$ constitutes a fixed effect for the buyer trading with s at time t. In x_{st}, I include p^g_{bsmt}, the average price of the garment in the $-sbmt$ combination, so the θ_s shifters in the demand condition on output prices. The underlying assumption in this definition of supplier's quality, as in the relevant literature (see Hottman et al. (2014)), is that marginal costs affect firm volumes only via the manufacturer's price. An alternative specification uses p^f_{bsmt} as an instrument for the price of the output. In the context of garment production, high-quality pieces are produced with better fabric, which, in turn, constitutes not only the bulk of the weight of the garment, but also the largest component in the per-unit cost. In addition, x_{st} contains dummies for the product category (m) and the quarter (t). Other controls include the material of the fabric used (cotton, synthetics, etc.), other non-fabric imported inputs, the mode of transport, the Customs Port and the terms of trade. The cumulative count of the quarters of activity of the supplier in product category $-m$ and overall are also included. Importantly, I include a dummy indicating whether the quarter corresponds to the first interaction between the buyer and the manufacturer and the cumulative sum of previous trade between the pair, if appropriate. To account for capacity restrictions, I use the (running) maximum volume shipped by the supplier in a quarter as a proxy for its capacity constraint.

We can decompose the error term η_{st} into three components. A zero-mean random match component, μ_{sb}, that collects demand shifters specific to the $-sb$ pair, on top of the additive part $(\theta_b + \theta_s)$. A unit root component χ_{st}, again with mean zero, allowing for innovations connected to the supplier's learning process, changes in its “outside option” over time, or other time-driven effects. Finally a standard random error with mean zero, ϵ_{st}.

Stacking exported volumes sorted by s and t, q is an $\mathcal{S}\times 1$ vector of outcomes, one per $-st$ combination. With N_S the total number of sellers and N_B the total number of buyers, the design matrices of seller and buyer identifiers can be defined as the $\mathcal{S}\times N_s$ matrix $S=[s^1,\ldots,s^{N_S}]$ and the $\mathcal{S}\times N_b$ matrix $B=[b^1,\ldots,b^{N_B}]$, respectively. Calling X the $\mathcal{S}\times K$ matrix of regressors with K covariates and η the vector of errors, the model is:

$$q = B\theta^B + S\theta^S + X\beta + \eta$$

which can be estimated by standard OLS solving $Z'Z\gamma = Z'q$, with Z denoting the concatenation $[B,S,X]$ and $\gamma = [\theta^{B'},\theta^{S'},\beta']$.

The estimation procedure mimics the techniques that have been used in the labour literature in the tradition started by Abowd et al. (1999) (and subsequent papers of the same authors). They exploit employee-employer matched data to recover firm and individual fixed effects from a wage equation, to measure the unobserved productivity, ability or the *types* of the players. The underlying assumption in all these applications is that after including the appropriate controls, fixed effects recover the relevant dimension of the unobserved heterogeneity. Some applications are those in Becker (2005)) on returns to seniority, Woodcock (2003) on heterogeneity and worker-firm learning and Barth and Dale-Olsen (2003) on assortativeness. Estimating these many fixed effects with standard techniques introduces the typical problems in the computation of a generalized inverse of the estimation matrix in the normal equation, involving very sparse matrices. I follow the approach presented in Abowd et al. (2002a), who develop a method to solve exactly the least squares problem in this setting, grouping the data in the ”components” of the network, which is proved necessary and sufficient for the estimation of both fixed effects for most of the buyer-seller pairs. The procedure consists of dividing the data in the fully connected subgraphs that are not inter-connected with each other (the“components”), sweeping out one of the fixed effects using a within transformation and calculating the fixed effect of the other set of players by introducing individual dummy variables. Those components in which a buyer has only sellers that don’t trade with other buyers, the buyer fixed effect cannot be estimated. Like in the relevant literature, the successful recovery of players types in the context of our assumptions depends on the number of “movers” each player is connected to. This implies that not all fixed effects are identifiable. In particular, those buyers and sellers that have few interactions within the panel and, even more so, restricted to one trade partner only have no fixed effect estimated, which tends to select against small players. I therefore restrict the estimation to the largest component of the network (the maximal connected set).[76] With this, the specification above recovers fixed effects for sellers that account jointly for 83% of the trade in woven garments in the panel.[77] For concreteness, there are $20,855$ buyer-seller combinations in the data, and I use $59,431$ buyer-seller-product-quarter level observations to obtain the suppliers’ fixed effects. The vast majority ($59,362$) of those observations lie in the main component of the network and discarding collinear intercepts, I obtain quality measures for $2,227$ suppliers.

[76] Note that the recovery of fixed effects on non-principal components of the network is also be possible, but because the “centering” of these effects would differ across groups, comparisons of fixed effects of sellers or buyers in different components is tricky. In addition, fixed effects are only identifiable for the second largest component in the network. Including the players in this component, would only increase the coverage of the estimation by $< 1\%$ in terms of volumes accounted for by players with an identified fixed effect.

[77] Of the $4,031$ suppliers in the dataset, I obtain quality measures for $2,227$.

Assumptions and Endogeneity

The analysis of the assumptions underlying the estimation process follow closely Card et al. (2013) and Abowd et al. (2002b). As in any OLS setup, orthogonality of all vectors in B, S and X with respect to the error term needs to be assumed. The description on the decomposition of the error term (and in particular, the zero-mean conditions) guarantee exogeneity of the suppliers' fixed effects. The exogeneity of buyers' fixed effects, $E[\theta_b'\eta] = 0$ yet needs to be established and as in Card et al. (2013), it suffices to have that the assignment of sellers to buyers is strictly exogenous with respect to η: $Prob(\theta_{b(st)} = i|\eta) = Prob(\theta_{b(st)} = i) = A_{it}(\theta_s, \theta_{b_1}, \ldots, \theta_{b_{N_B}})$ for all sellers and time periods.[78] Note that the assignment probability function, $A_{it}(\theta_s, \ldots)$ clearly allows for systematic patterns of assignment of sellers to buyers that depend on θ_s and θ_b. For instance, it could be that suppliers move more frequently onto buyers with larger demands, or that lower quality suppliers switch buyers more often or that higher quality suppliers are more likely to match with high demand buyers.

The orthogonality condition imposed above would be violated if matches were driven by μ_{sb}: the demand shifter for different suppliers in their relationships with a certain buyer would depend on an unobserved characteristic of the match. As in Card et al. (2013), this can be tested in two ways in our context.

First, if suppliers are selected into relationships based on the unobserved match component in the demand, the gains in the demand for suppliers that transition from buyer b_1 to buyer b_2 should be different from the gains or losses incurred when doing the opposite move, from b_2 to b_1. For a supplier that at t trades with b_2 while in $t-1$ was trading with b_1 the change in the expected demand is given by:

$$E[q_{st} - q_{st-1}|b(st) = b_1, b(st-1) = b_2] = \theta_{b_2} - \theta_{b_1} + E[\mu_{sb_2} - \mu_{sb_1}|b(st) = b_1, b(st-1) = b_2]$$

And if the a seller was to move in the opposite direction:

$$E[q_{st} - q_{st-1}|b(st) = b_2, b(st-1) = b_1] = \theta_{b_1} - \theta_{b_2} + E[\mu_{sb_1} - \mu_{sb_2}|b(st) = b_2, b(st-1) = b_1]$$

If sorting was to be driven by the match component, the sorting bias terms would induce (positive) demand gains in both cases. If there was no sorting on the match component, then the gains from the transition in one direction would be equivalent to the losses in the transition in the opposite direction. To asses an analogous problem in their context, Card et al. (2013) study the transition patterns of the movers and find that gains from moving up to larger buyers are symmetric to losses incurred when moving down to smaller players. I find an equivalent result in my setting, indicating that match-specific components don't seem to drive the transitions. For supplier s, trading with b in quarter t, I calculate the average size, over time up to t, of the demand channelled to a single supplier by buyer b, excluding its interactions with s. This is $\overline{q}_{st}$. For all transitions the seller does -that is, moving from one buyer to another- we can compute the difference in buyers' sizes as $\Delta_b\overline{q}_{st} = \overline{q}_{b_2;st} - \overline{q}_{b_1;st-1}$. This takes a positive value if the supplier is moving onto larger buyers (where "large" is measured as explained above) and negative otherwise. In that transition, the supplier's gains (or losses) in terms

[78] From Card et al. (2013), $E[\theta_b'\eta] = E[\sum_{s,t} \theta_{b(st)}\eta_{st}] = E[\sum_{s,t} E[\theta_{b(st)}|\eta]\eta_{st}] = E[\sum_{s,t} A_{it}(\cdot)\eta_{st}] = \sum_{s,t} A_{it}(\cdot)E[\eta_{st}] = 0$.

of its demand can be written as $\Delta q_{st} = q_{b_2;st} - q_{b_1;st-1}$. It is expected that transition to larger buyers (positive $\Delta_b \overline{q}_{st}$) induce increases in the supplier's demand, generating also positive Δq_{st}. Therefore, the relation between these two should be positive. However, if those transitions are induced by match-specific components, the effect of $\Delta_b \overline{q}_{st}$ on Δq_{st} should be significantly different when the supplier is "downgrading" and moving down to smaller buyers than when it is "upgrading" and moving up to larger buyers. This means that in a regression of Δq_{st} on $\Delta_b \overline{q}_{st}$ should have a different slope when moving up or down. This is tested in the regressions presented in Table 23. We can corroborate that the effect of the difference in buyers' sizes on the gains or losses in demand for the supplier after transitions is of the expected sign and does not differ significantly with the direction of the transition.

Another way of assessing the role of the match component in the sorting patterns is to decompose the remaining variability in the outcome variable in a bs-specific intercept and other varying factors. Regressing the residuals of the model with additive buyer and seller fixed effects on buyer-product-quarter dummies, and adding afterwards buyer-seller dummies (capturing μ_{sb}), the improvement in fit is very small, in the order of 0.02 in the R^2 of the regression.[79]

Endogeneity problems could also arise if the buyer effect is correlated with the unit root component χ_{st}. This drift was mainly capturing two stories. The first one is that the supplier might be learning over time in such a way that it becomes more (or less) appealing to all buyers. In this case, endogenous mobility would arise if, for instance, a supplier that ends up revealing itself as better than expected might capture a higher demand from its current buyer and also might become more likely to transition onto larger buyers afterwards. At the other end of the spectrum, suppliers that are learnt to be of a lower quality might see their volumes shrinking with their incumbent buyer, to move later on to smaller buyers. This effect would induce a positive correlation between the χ_{st} drift and the buyer effect, overestimating the role of the latter. The second reason for correlation between the χ_{st} drift and transitions onto different buyers would be that suppliers that over time reveal themselves as good, leave their buyers as these are "outbid" by larger buyers.

One way of testing for these sources of endogenous mobility is to study the presence of systematic trends in volumes prior to the seller changing buyers. For supplier s, trading with b in quarter q, I calculate the average size, over time up to q, of b's demand to its suppliers, excluding its interactions with s. I consider transitions of supplier s from one buyer to another and characterise these with the change in the size of the buyer, measured with those averages: $\Delta_b \overline{q}_{st} = \overline{q}_{b_2;st} - \overline{q}_{b_1;st-1}$. The drift story above is compatible with those changes being driven by previous increases in volumes with the first buyer, $\Delta q_{st-1} = q_{b_1;st-1} - q_{b_1;st-2}$. I test for this regressing the $\Delta_b \overline{q}_{st}$ on Δq_{st-1} including product-quarter fixed effects. As an additional control, I include changes in prices Δp_{st-1} and re-run the regressions using subsequent lags of q_{st} (instead of first differences in the lag). The cross-sectional unit in the panel is a seller and the time dimension is given by quarters. I include only seller-quarter (st) combinations where the following conditions are satisfied: (i) s is active in the panel for at least a year before t; (ii) s trades with b_1 for at least two consecutive quarters before switching to b_2; (iii) after switching, s and b_2 trade for at least two quarters before s returning to b_1. The results are presented in Table 22 and show no systematic relation between past changes in volumes and switches to buyers of different "sizes".[80]

[79] This supports the idea that the match-specific component, μ_{sb} is a random effect uncorrelated to the additive buyer and seller effects.

[80] Note that the main regressor is not strictly speaking a lagged dependent variable, so I don't face the standard initial conditions problem. The second column of the table, including regressors in lagged levels (a la Anderson and

Excluding Buyers' Effects

Like in the labor literature, if instead of estimating $q = B\theta^B + S\theta^S + X\beta + \eta$ we were to estimate $q = S\tilde{\theta}^S + X\tilde{\beta} + \eta$, the estimated seller's effects $\tilde{\theta}^S$ would be equivalent to the sum of the θ^S effect and a weighted average of the buyer effects of those buyers that traded with the supplier, conditional on X:[81]

$$\tilde{\theta}^S = \theta^S + (S'M_X S)^{-1} S' M_X B\theta^B$$

where $M_X = I - X(X'X)^{inv}X'$ and $(\cdot)^{inv}$ is the generalized inverse of the matrix in parenthesis. Then, if X is orthogonal to S and B, the omitted variable bias is given, for each seller s, by the average of buyers fixed effects, weighted by the instances of trade between them: $\tilde{\theta}_s - \theta_s = \sum_t^{T_s} \theta_{b(st)} T_s^{-1}$.

Robustness Checks

The baseline specification in model $q_{st} = \theta_{b(st)} + \theta_s + x'_{st}\beta + \eta_{st}$ included buyer and seller fixed effects, $\theta_{b(st)}$ and θ_s, and product and time dummies. The main set of regressors includes: (i) the average price of the garment for the $-sbmt$ combination, (ii) indicators for the material of the fabric used (cotton, synthetics, etc.) and other non-fabric imported inputs, (iii) the mode of transport, the Customs Port and the terms of trade, (iv) the supplier's experience, measured as the cumulative count of the quarters of activity of the supplier in product category $-m$ and overall, (v) the experience of the relevant pair in the relationship, as informed by a dummy indicating whether the quarter corresponds to the first interaction between the buyer and the manufacturer and the cumulative sum of previous trade between the pair, (vi) a proxy for the supplier's capacity constraint - the (running) maximum volume shipped by the supplier in a quarter.

A number of alternative specifications are considered and described below. The comparisons with the baseline specification are presented in Figure **??**, and for reference, the raw correlations of the obtained measures are:

Table 7: Correlations

	(1)	(2)	(3)	(4)	(5)
(1) Baseline	1.000				
(2) IV on Prices	0.899	1.000			
(3) Aggregated over Buyers	0.703	0.715	1.000		
(4) Buyer-Product FEs	0.878	0.799	0.681	1.000	
(5) Homogeneous Products	0.843	0.800	0.738	0.829	1.000

Prices of Inputs as Instruments One possible concern with regards to the baseline specification is the simultaneity bias arising from the inclusion of contemporaneous prices in the quantity equation. An alternative specification uses the price for the fabric used by s, when serving b with product m at time t, p^f_{bsmt}, as an instrument for the price of the output. In the context of garment production, (i) fabric is the largest component in unit costs and (ii) the overall quality of the garment is driven to

Hsiao (1982)) should dissipate any concerns around this.

[81] See Abowd et al. (1999) for a detailed decomposition

a large extent by the quality of the fabric, which in turn is highly correlated to the quality of other inputs.

Homogeneous Products Only There are on average 464 firms per product category (HS6) in our sample and each firm produces on average 5.5 products over the span of the panel. For the purpose of the recovery of the fixed effects, we retain the main product the seller is exporting in a given quarter, which accounts for more than 87% of its exports for the vast majority of the $-st$ combinations. Even though we allow for product fixed effects, it is possible that the differentiation within product categories is driving the dispersion in the sellers's effects. I re-estimate the seller fixed effects restricting the sample to products that are relatively homogeneous. For each product category, I use the dispersion in the prices of fabric used as a proxy for the within-product differentiation. I keep then the ten product categories with the lowest differentiation according to this measure and re-estimate the sellers' fixed effects on this subsample. Notice that this restricted sample allows for the identification of a much smaller number of seller demand shifters.

Checking Against Buyers' Measures

In Figure 7 I match the measures produced by the approach described here to the classification the largest importer in the sample, $H\&M$ uses to rank their suppliers into Gold, Silver and Other partners. The median and interquartile range align as expected.

Appendix C: Tables and Figures

Table 8: Compliance and Reputation - Selected Episodes

Details Last Two Episodes

Date	Episode	Buyers	Press Coverage	Compensations	Other
November 2012	Fire in Tazreen Fashions: 115 deaths + 200 injuries; captive workers; child labour	C&A, Walmart, others (unauthorised production)	International. Google "C&A" + "Tazreen": 1,120 links to Int Articles	USD 5.7 million (Less than 2% effective - Caritas, Government, BGMEA) C&A compensations of < 200,000 USD. Walmart refused.	Negative Press Walmart. Tripartite National Action Plan (Consortium; No Actions so far)
April 2013	Building Collapse in Plaza Rana: 1,132 deaths + 1,800 injuries; forced labour	Benetton, Kik, Mango, Primark, Walmart (?)	International. Google "Walmart" + "Plaza Rana": 1,600 links to Int Articles	USD 71 million (government and NGOs). Primark lump sum of USD 190 to each of 3,300 families.	Accord on Fire and Building Safety (170+ international buyers, with contributions from 1,000 USD per year (volume of < 1 MillUSD) to 500,000 USD (500+ MillUSD per year)).

Overview Past Episodes

Date	Episode	Buyers
April 2005	Spectrum Factory collapse during night of forced work: 64 deaths + 75 injuries	Inditex, Carrefour, etc.
February 2006	Fire at KTS Textile, child labour, locked exits: 61 deaths + 100 injuries	ATT Enterprise, VIDA Andrew Scott
February 2006	Collapse Phoenix Building, unauthorised plant: 22 deaths + 50 injuries	Unlabeled but export to Germany, Switzerland and Denmark
February 2006	Imam Group explosion and blocked exits: 57 injuries	Kmart, Folsom Corporation.
March 2006	Fire Sayem Fashions: 3 deaths + 50 injuries	Inditex, Bershka, BSK.
February 2010	Fire in Garib and Garib, no ventilation, suffocation, blocked exits: 21 deaths + 57 injuries	H&M, El Corte Ingles.
December 2010	Fire Thats It Sportswear (Hameem Group), no exits, no drills, illegal inaccessible top floors: 29 deaths + 11 injuries	Gap, PVH Corp., VFCorporation.
December 2011	Boiler Explosion Eurotex, stampede, collapse of stairs with exits blocked: 2 deaths + 64 injuries	Tommy Hilfiger (PVH Corp.), Inditex, Gap, C&A.

Information from Clean Clothes Campaign at the International Labour Rights Forum.. Actual compensations in Plaza Rana: To date 834 families have received BDT 20,000 (USD 257) from the government to cover burials costs and the dependents of 777 dead workers have received payments ranging from BDT 100,000 (USD 1,285) to BDT 400,000 (USD 5,140), covered with donations from governmental, non-governmental and private donors. Thirty-six amputees and paralysed workers have received compensation from the government in the range of BDT 10 (USD 12,870) and BDT 15 (USD 19,305) each. Results of Accord Inspections: 80,000 safety hazards in the 1,106 factories investigated so far. 20+ immediate closures. Results of Alliance (Walmart + Gap): Investment of 22 MillUSD, 600 inspections. Episodes listed from 2005 to 2011 are examples taken from reports in Clean Clothes Campaign. Not an exhaustive list.

Table 9: Exporter Dynamics - Selected Sample, Own Data

Variable	2005	2006	2007	2008	2009	2010	2011	2012
Counts of Exporters and Survival								
Number of Exporters (N=4,301) - Selected Products	1,430	1,587	1,707	1,853	1,849	1,955	1,994	1,718
Distribution of birth dates all sellers	0.51	0.09	0.10	0.10	0.07	0.06	0.05	0.02
Number of Entrants - Active, All products	1,430	197	194	216	167	152	160	80
Number of Entrants - Active, Selected products	1,430	455	414	433	371	387	353	188
Number of Entrants - Active, Selected products, first time exporters		258	220	217	204	235	193	108
Number of Exiters - All products	100	84	105	162	147	158	222	1,718
Entry rate	1.00	0.29	0.24	0.23	0.20	0.20	0.18	0.11
Entry rate (first time exporters)		0.16	0.13	0.12	0.11	0.12	0.10	0.06
Share of first time exporters in TEV	1.00	0.06	0.03	0.04	0.02	0.03	0.01	0.01
Survival for 1 year or more, conditional on cohort	0.92	0.84	0.88	0.83	0.80	0.79	0.49	
Survival for 2 years or more, conditional on cohort	0.90	0.71	0.81	0.74	0.70	0.51		
Survival for 3 years or more, conditional on cohort	0.80	0.68	0.72	0.67	0.45			
Export Values per Exporter								
Export Value per Exporter: Mean	1,689,675	1,927,905	1,871,275	2,060,098	2,158,974	2,421,725	3,124,945	2,401,872
Export Value per Exporter: Median	515,109	531,982	400,608	375,019	382,287	357,141	385,842	426,200
Export Value per Exporter: StDev	3,303,062	3,825,839	3,999,126	4,857,570	5,305,241	5,702,518	7,395,031	5,501,074
Unit Prices per Exporter								
Unit Price per Exporter: Mean	9.84	9.94	10.29	11.18	11.22	11.65	15.53	14.80
Unit Price per Exporter: Median	9.47	9.47	9.85	10.66	10.71	11.05	14.43	13.99
Unit Price per Exporter: StDev	5.08	4.49	4.51	5.42	5.64	5.56	33.20	8.54
Competition								
Herfindahl-Hirschman Index	0.00337	0.00311	0.00326	0.00354	0.00380	0.00334	0.00331	0.00363
Share of top 1% Exporters in TEV	0.136	0.136	0.148	0.170	0.182	0.173	0.170	0.165
Share of top 5 Exporters in TEV	0.418	0.420	0.436	0.468	0.480	0.489	0.498	0.480
Share of top 25% Exporters in TEV	0.814	0.823	0.847	0.867	0.866	0.886	0.888	0.872
Product Diversification								
Products per Exporter: Mean	4.10	4.07	3.70	3.55	3.29	3.31	3.27	3.03
Products per Exporter: Median	3.00	3.00	3.00	3.00	2.00	2.00	2.00	2.00
Products per Exporter: StDev	3.31	3.22	2.98	2.89	2.58	2.63	2.64	2.34
Destination Diversification								
Destinations per Exporter: Mean	3.00	3.24	3.16	3.25	3.40	3.54	3.86	3.96
Destinations per Exporter: Median	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Destinations per Exporter: StDev	2.96	3.24	3.20	3.41	3.73	3.93	4.68	5.18
Regions per Exporter: Mean	1.51	1.57	1.62	1.65	1.69	1.75	1.79	1.86
Regions per Exporter: Median	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Regions per Exporter: StDev	0.69	0.77	0.84	0.89	0.92	0.96	0.98	0.99
Proportion Sellers in US non-Europe	0.22	0.22	0.21	0.22	0.20	0.19	0.18	0.19
Proportion Sellers in Europe non-US	0.25	0.25	0.25	0.25	0.26	0.25	0.26	0.29
Proportion of Sellers in both	0.27	0.26	0.25	0.25	0.25	0.26	0.25	0.25

A cross-sectional unit in this table is a firm producing in the selected product categories through the main Custom Office. The operational definitions for the reported variables are as follows: exporter is any firm that exports in year t; the distribution of birth dates is the proportion over the whole set of sellers (4,031) that started trading (any product) in year t, irrespective of whether they are active or not in the selected products in such year; entrant - all products uses the year of entry into exporting garment (any product at all) as the entry date and considers active firms only; entrant - selected products uses the year of entry into exporting at least one of the selected products as the entry date, irrespective of whether the exporter was active before or not; entrant - selected products, first timer defines entry as the first time exporting experience being in the selected products; exit is defined by the last date a supplier is observed in the panel (note that this is subject to above-censoring and it is also different from the World Bank's definition that takes only the one-year-ahead approach); Entry rate is defined as the ratio between the number of entrants in the selected products to the number of all exporters in selected products; Entry rate is also computed for first time exporters; the share of entrants in TEV is calculated summing over all the exports from first time exporters (entrants, selected products, first timers) in the year, relative to the overall value of annual exports; The Herfindahl-Hirschman Index is not normalised, for ease of comparison with the World Bank measures, so it ranges from $1/N$ to 1 and calculated as $HHI = \sum_{i=1}^{N}(s_i)^2$ with s_i, the share of firm i in the industry. The total number of products (HS6) in the sample is 48. Destinations are countries as recorded in the Customs Data. End Markets are groups of countries into five categories: "US and Canada", "European Union", "Garment-relevant asian countries", "Bangladesh local", "Others". All other definitions are analogous to those use in the World Bank tables.

Table 10: Across-Panel Summary Statistics - Buyers

Panel A: Buyer - Quarter Level Variables; All Non-Large Buyers					
Variable	Mean	Std. Dev.	Min.	Max.	N
Value in 10,000 USD	53.97	154.86	0.01	4,372.35	37,959
Volumes in 10,000kg	4.55	12.27	0.00	300.5	37,959
Unit Values in USD, w.a.	12.74	7.09	0.13	165.52	37,959
Location in distribution of prices	-0.47	1.13	-3.52	32.11	37959
Simultaneous active orders	3.86	6.38	1	181	34,387
Simultaneous allocation of orders	2.86	4.11	1	67	23,781
Count of products	2.44	2.25	1	23	37,959
Count of trade partners	2.54	3.16	1	58	37,385

Panel B: Buyer - Quarter Level Variables; Large Buyers Only					
Variable	Mean	Std. Dev.	Min.	Max.	N
Value in 10,000 USD	2,973.72	2,616.56	4.87	20,879.84	398
Volumes in 10,000kg	231.34	168.19	0.31	1,015.79	398
Unit Values in USD, w.a.	12.42	3.17	6.02	21.53	398
Location in distribution of prices	-0.3	0.56	-0.97	0.64	398
Simultaneous active orders	78.51	57.58	1	412	392
Simultaneous allocation of orders	51.93	36.91	1	201	216
Count of products	13.55	5.3	1	26	398
Count of trade partners	24.11	14.32	1	68	392

Panel C: Order Level Variables; All Non-Large Buyers					
Variable	Mean	Std. Dev.	Min.	Max.	N
Value in 1,000 USD	319.56	614.32	0.11	19,015.4	51,852
Volumes in 1,000 kg	25.96	49.11	0.01	1,593.39	51,852
Unit Values in USD, w.a.	14.01	7.03	0.56	168.64	51,852
Quarterly average volume in 1,000 kg	15.1	22.39	0.01	676.01	51,852
Duration in quarters	1.56	0.81	1	13	51,852
Count of products, HS6	1.58	0.99	1	16	51,852
Count of products, HS4	1.26	0.5	1	4	51,852
Price of Imported fabric, USD, w.a.	7.95	4.75	0.02	193.48	51,852

Panel D: Order Level Variables; Large Buyers Only					
Variable	Mean	Std. Dev.	Min.	Max.	N
Value in 1,000 USD	753.83	1,164.2	0.04	19,812.69	13,640
Volumes in 1,000 kg	57.7	96.1	0.02	1,363.74	13,640
Unit Values in USD, w.a.	15.66	7.49	0.15	109.96	13,640
Quarterly average volume in 1,000 kg	30.68	44.78	0.01	639.82	13,640
Duration in quarters	1.77	0.9	1	11	13,640
Count of products, HS6	1.73	1.17	1	17	13,640
Count of products, HS4	1.33	0.57	1	4	13,640
Price of Imported fabric, USD, w.a.	7.94	4.68	0.16	197.22	13,640

All values are in USD. Aggregation over all product categories. “w.a.” stands for weighted average. The location variable is generated as the weighted average over all products of the normalised (with the median absolute deviation) distance between the median price for the product and the average price the buyer pays. This average is taking over all quarters in the panel, so the summary statistics for this variable in Panels A and B describe buyers as units of analysis. For the rest of the variables, Panels A and B have buyer-quarter pairs as units. A unit in Panels C and D is an export order.

Table 11: Buyers, Sellers and Orders - 2005-2012

	2005	2006	2007	2008	2009	2010	2011	2012
Exported Values (100k USD)	23,323	29,702	30,710	36,740	38,868	46,298	60,892	47,998
Share of Top 10% Buyers	0.80	0.81	0.82	0.83	0.84	0.85	0.83	0.82
Share of Top 10% Sellers	0.54	0.54	0.57	0.60	0.61	0.63	0.64	0.68
Number Sellers / Year	1,430	1,587	1,707	1,853	1,849	1,955	1,994	1,718
Number Buyers / Year	1,642	1,829	2,000	2,238	2,340	2,422	2,247	1,944
Average Sellers per Buyer per Year	3.78	3.92	3.65	3.61	3.48	3.62	4.03	3.71
Average Sellers per Top Buyer per Year	13.52	14.13	13.55	13.73	13.13	13.63	15.49	14.59
Average Orders per Buyer per Year	8.96	9.16	8.73	8.66	8.25	8.57	9.66	9.28
Average Orders per Top Buyer per Year	44.76	46.92	45.30	45.50	42.62	44.27	49.91	47.02

Data from January 2005 to September 2012. Exports in the four main woven product categories (HS4 6203, 6204, 6205, 6206). All 48 HS6 code products within these categories are pulled together.

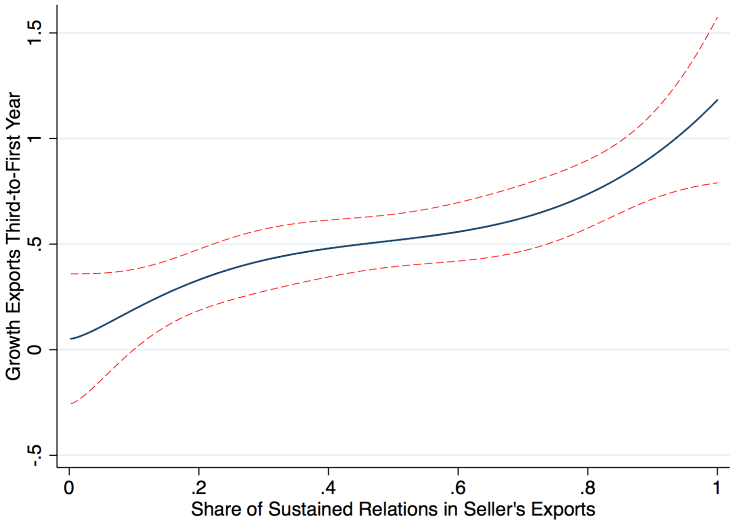


Figure 1: Sustained Relations and Sellers' Growth.
The figure shows the fitted line from a polynomial regression of the growth rate of exports in the first-to-third year of the seller's export activities on the share of sustained relations in seller's exports over that period. The dotted lines are the bounds for 90% confidence intervals. The regression is based on 986 datapoints (suppliers) after trimming the top and bottom 5% of the growth variable and discarding the suppliers that lie in the top and bottom 10% of the size distribution.

Table 12: Evolution of Relations Over time

	(1) Number of orders	(2) Traded volumes, logs	(3) Average volume orders, logs
Trend	0.044***	0.042***	0.012
	(0.01)	(0.01)	(0.02)
Squared Trend	-0.002***	-0.001***	-0.001
	(0.00)	(0.00)	(0.00)
Relation Fixed Effects	Yes	Yes	Yes
Seasonal Effects	Yes	Yes	Yes
Constant	1.817***	11.573***	2.954***
	(0.03)	(0.02)	(0.02)
Observations	62059	62059	62028
R^2	0.029	0.013	0.014

Panel regressions with fixed effects for cross sectional units (buyer-seller relationship). Standard errors are clustered at the buyer level. Trends are taken over quarters of effective interaction, which do not coincide with calendar quarters. Only relationships that survive 4 quarters of interaction (gaps allowed for) are considered.

Table 13: Comparison First Orders and Non-First Orders

Mean Comparisons Across Sample			
	All First Orders	All Non-First Orders	t-statistic
N	20474	54548	-
Size of Order (kg)	14,320	28,689	11.96
Price of Order (USD/Kg)	11.95	12.69	16.16
Number of Shipments	3.90	9.11	49.13
Price of Fabric	8.06	10.51	0.98
Mean Comparisons Paired Observations			
	First Orders - Rejected	Non-First Orders - Accepted	t-statistic
Price of Order (USD/Kg)	13.01	14.70	7.06
	First Orders - Rejected	First Order - Accepted	t-statistic
Price of Order (USD/Kg)	13.02	14.33	5.13

The table compares means of the relevant variables across two groups: all the first orders we observe in the panel and all non-first orders. The first group includes interactions that continued to constitute long-lasting relations and first orders that constitute the only interaction in the relationship. Censoring both at the start and end of the panel are accounted for. All relations are considered within the product category (HS6). The bottom panel compares the average price paid by each buyer-product to its first established partner (average computed over all orders but the first one during the first year of the relationship) and the mean price paid by the same buyer-product in first orders with one-off suppliers. The last row in the table compares the price of the first order the buyer placed with the supplier that it formed a long lasting relationship with ("Accepted") and the average price of first orders allocated to other suppliers before ("Rejected"). For reference, critical values for two sided tests at 10%, 5% and 1% are $+/-1.645$, $+/-2.326$ and $+/-3.090$ respectively.

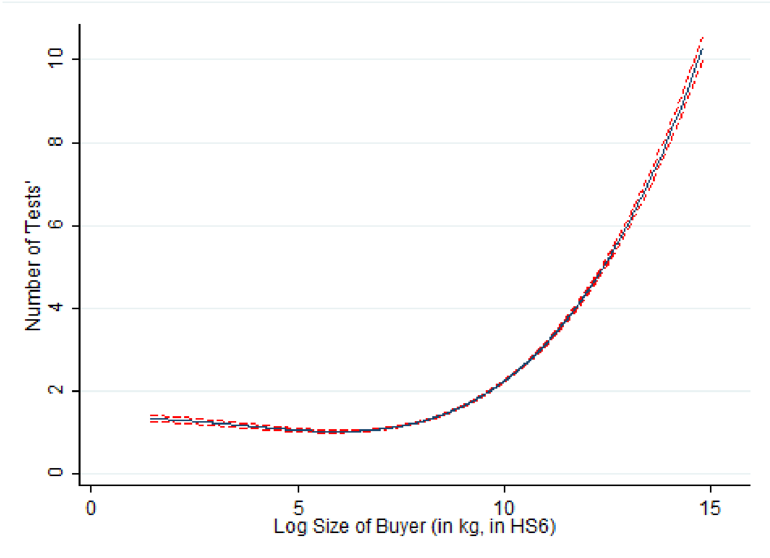


Figure 2: Search for Suppliers and Buyer's Size.
The figure shows the fitted line from a polynomial regression of the number of exploratory orders entrant buyers in a product category place to suppliers before establishing their first long-lasting relationship on a constant and the size of the buyer, calculated as the log of its imported volumes in the HS6 code. The dotted lines are the bounds for 90% confidence intervals. The regression is based on 16,072 datapoints corresponding to buyer-HS6 combinations for which we 'observe' the buyer's entry into the market and the buyer eventually forms a sustained relationship with a supplier.

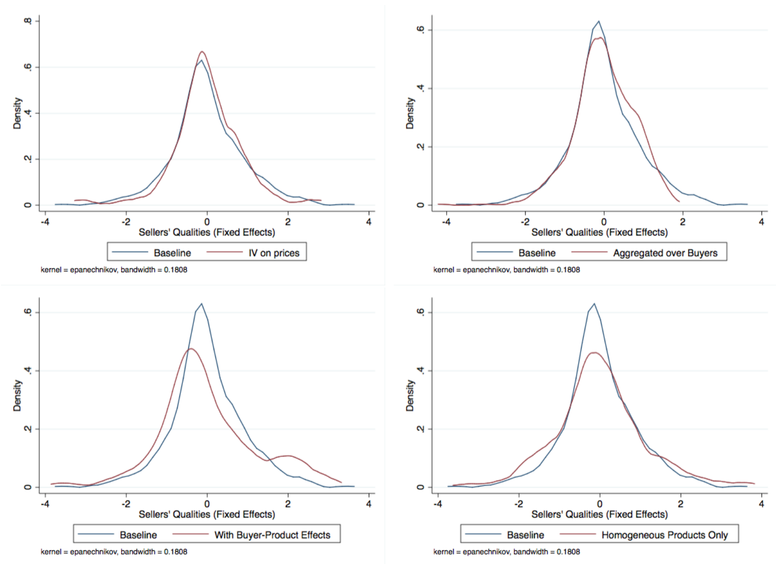


Figure 3: Comparison of Alternative Specifications of Fixed Effects Regressions.

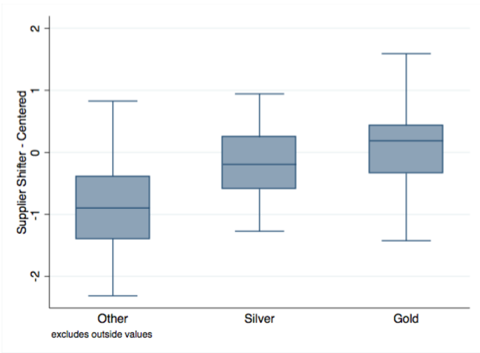


Table 14: Variance of Types of Suppliers per Broad Product Category

Broad Product Category	Variance
Female Dresses	2.3681
Female Trousers	1.7288
Male Suits	1.4781
Female Jacket	1.4219
Female Skirts	1.4104
Female Ensemble	1.2129
Male Jacket	1.1539
Male Ensemble	1.0670
Female Shirts	1.0438
Male Trousers	0.9362
Male Shirts	0.8746

The 48 HS codes disaggregated to the 6^{th} digit were grouped in 11 broad categories. Note that women's ensembles are pooled together with women's suits. The empirical variances are computed directly from the estimated coefficients, when data is de-meaned to be centred around zero.

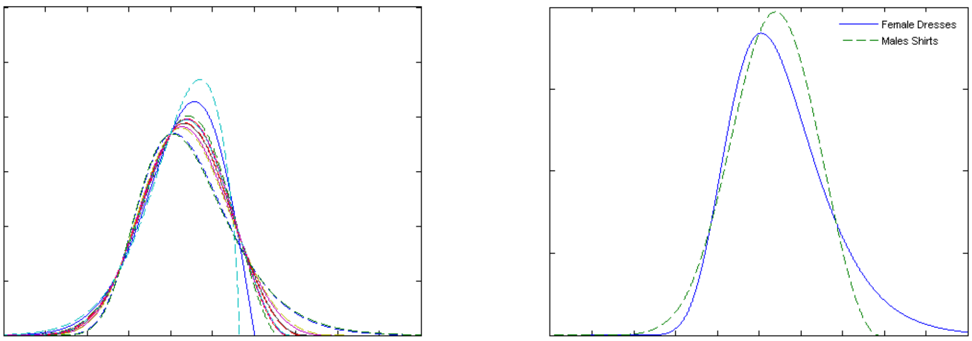


Figure 4: Distribution of Types of Suppliers per Broad Product Category.
Graphs show generated distributions over sellers' types within each broad product category. The right panel corresponds all product categories and the left panel, to two selected products for comparison. Top and bottom 3^{th} percentiles within each product are dropped from the graphs. All graphs generated in Matlab.

Table 15: Markets and Heterogeneity

	Heterogeneity Across Suppliers in Market-Quarter
Average Lead Time (logs) in Market	-0.391*
	(0.20)
Average Order Turnover (logs) in Market	-0.226***
	(0.08)
Female Dummy	0.093**
	(0.04)
St.Dev. Price of Fabric (logs) in Market	0.124**
	(0.06)
Constant	2.924***
	(0.84)
Quarter FE	Yes
Observations	993
R^2	0.095

An observational unit in this regression is a market - quarter combination. Standard errors are bootstrapped and clustered at the HS6 level. The outcome is constructed as the standard deviation of the types of the sellers active in each market - quarter combination. Fixed effects for the quarter are included. All other covariates are constructed at the market level.

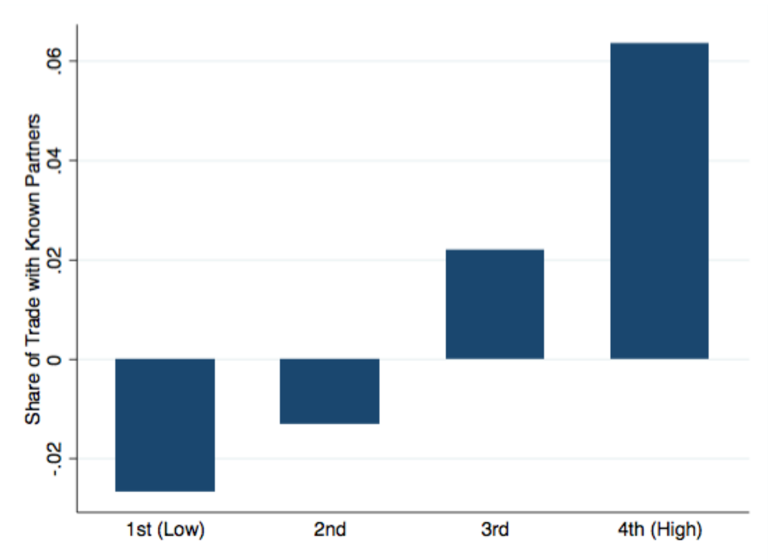


Figure 5: Persistency in trade and Heterogeneity within Markets.
The categories on the horizontal axis represent quartiles in the distribution of markets, characterised by the heterogeneity of available suppliers. Markets are defined as product (HS6) - season interactions. For each market, I generate the standard deviation over the types of all the suppliers available in that market. I rank the markets according to this characteristic and generate quartiles of this distribution such that the first quartile contains markets with low heterogeneity across suppliers. For each buyer I construct the share of its demand that is channeled to suppliers that are already known to the buyer. For this purpose, orders placed within the first 14 months of the panel or within the first 14 months of the buyer present in the market are not taken into account. I de-mean those shares for each buyer using a regression of such shares on buyer fixed effects. Each bar in the graph, then, represents the average, taken over all the markets in the corresponding quartile, of the (buyer de-meaned) share of exports channeled to known suppliers.

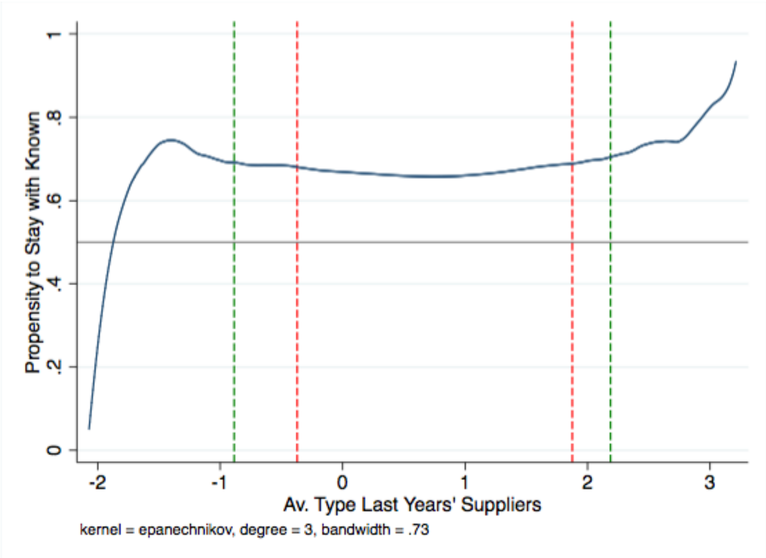
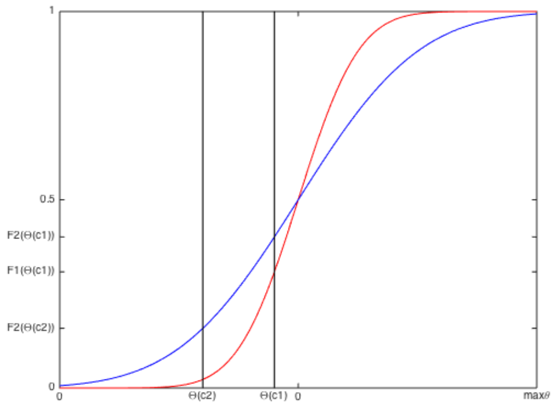


Figure 6: Propensity to Engage with New Suppliers and Past Experience.
The figure shows the fitted curve of a kernel (Epanechnikov) weighted local polynomial regression. The outcome is the proportion of new orders the buyer allocates to already known suppliers in a given quarter. The independent variable is the average of the *types* of the suppliers the buyer has interacted with in the last four quarters. Only buyers with more than 6 quarters of experience are considered. Buyer-quarters pairs in which no new orders are allocated are discarded. Top and bottom 5% of the distribution of average types are trimmed off. The vertical lines represent the bounds of the 75-th and 25-th (red) and 87.5-th and 12.5-th (green) percentiles of the overall distribution of types of suppliers in the market. The figure is based on 6,366 buyer-quarter combinations.



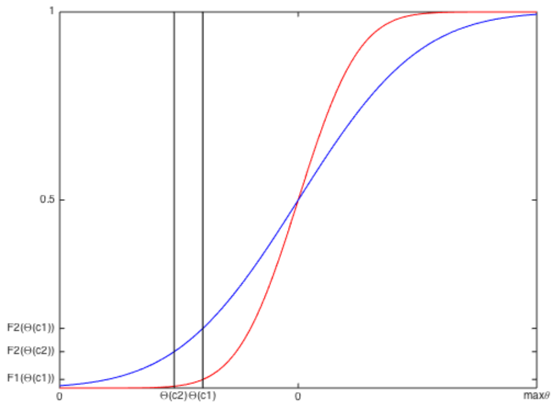


Figure 7: Mean Preserving Spread of the Distribution of Types of Suppliers.
F_1 is the original distribution over seller's types, θ. F_2 is a mean preserving of F_1. Θ_L and Θ_H correspond to the threshold suppliers buyers with high and low cost respectively are willing to accept (note that the subscripts L and H refer to the level of the threshold and *not* to the search costs). The figure is for illustrative purposes only.

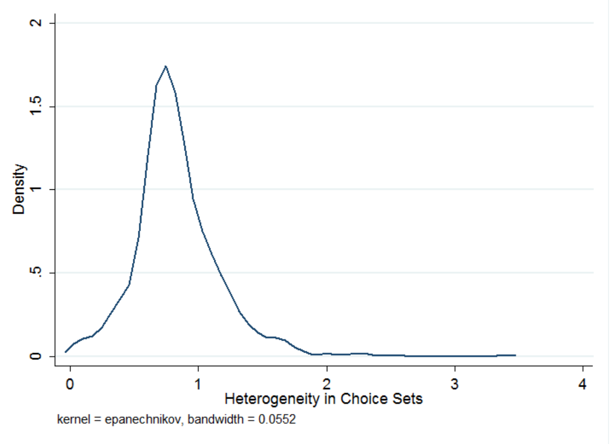


Figure 8: Distribution of Heterogeneity Measure across Choice Sets.
For each choice set, constructed as explained in the main text, the heterogeneity measure is obtained by computing the standard deviation of types of the suppliers available in such choice set. The figure depicts the distribution of the dispersion measure in all search spells considered in columns (2) to (7) of Table 16.

Table 16: Search Spells and the Intensive Margin

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Search Intensity: SI_{bms}			
$St.Dev.(\hat{\theta})_{ms}$	-0.356***	-0.293**	-0.274**	-0.284**	-0.315	-0.301**	-0.324**
	(0.112)	(0.110)	(0.131)	(0.126)	(0.217)	(0.127)	(0.124)
$\bar{\hat{\theta}}_{ms}$	-0.038	-0.043	-0.017	-0.042	-0.043	-0.031	-0.031
	(0.033)	(0.035)	(0.041)	(0.049)	(0.078)	(0.050)	(0.049)
$q_{bmt(s)}$	0.160***	0.155***	0.156***	0.184***	0.151***	0.198***	0.200***
	(0.016)	(0.015)	(0.012)	(0.019)	(0.021)	(0.014)	(0.014)
$N^s_{mt(s)}$	0.018	0.023	0.038*	0.054**	0.051	0.063***	0.061***
	(0.015)	(0.017)	(0.021)	(0.023)	(0.047)	(0.019)	(0.019)
θ^x_{bms}						0.023	
						(0.029)	
sh^x_{bmT-1}							0.061
							(0.099)
Product FEs	yes	yes	yes	.	yes	.	.
Quarter FEs	yes	yes	yes	yes	.	yes	yes
Buyer FEs	yes	yes	.	.	.	.	.
Buyer-Quarter FEs	.	.	.	.	yes	.	.
Buyer-Season FEs	.	.	yes	.	.	.	.
Buyer-Product FEs	.	.	.	yes	.	yes	yes
R^2	0.48	0.49	0.65	0.66	0.84	0.66	0.66
Observations	21,421	16,434	16,434	16,434	16,434	14,002	16,433

Notes: The dependent variable, SI_{bms}, is the search intensity of buyer b when undergoing search spell s in product m, measured as the (log) of the count of short lived relations during s. Fixed effects for the product category (HS4/5), the calendar quarter of the spell (this is the quarter in which the majority of the spell fell in or the quarter of the end of the spell), the buyer and the interaction between the buyer and the product, the quarter or the season are included in different specifications. $\bar{\hat{\theta}}_{ms}$ and $St.Dev.(\hat{\theta})_{ms}$ stand for the mean and standard deviation of qualities of available suppliers in spell s in product m. $q_{bmt(s)}$ id the (log) size of the demand in the quarter of the spell, $N^s_{mt(s)}$ the number of available suppliers, θ^x_{bms}, the quality of the supplier in the broken relation and sh^x_{bmT-1}, the share of the buyer's demand this supplier explained in the last year of activity in the relation. Column (1) includes all search spells. The rest of the columns include search spells that don't satisfy the criteria for re-allocation to existing suppliers or anticipated search. Standard errors clustered at the HS4 product code in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: Search Spells and Size of Buyer

	(1)	(2)	(3)
	Search Intensity SI_{bms}		
Buyer Decile 2	0.166**		
	(0.066)		
Buyer Decile 3	0.184**		
	(0.075)		
Buyer Decile 4	0.336***		
	(0.108)		
Buyer Decile 5	0.492***		
	(0.113)		
Buyer Decile 6	0.605***		
	(0.164)		
Buyer Decile 7	0.815***		
	(0.208)		
Buyer Decile 8	0.973***		
	(0.267)		
Buyer Decile 9	1.136***		
	(0.271)		
Buyer Decile 10	1.439***		
	(0.206)		
$StDev(\hat{\theta}_{ms})$	-0.298***	-0.252**	-0.250*
	(0.095)	(0.118)	(0.136)
$StDev(\hat{\theta}_{ms})$ Buyer Decile 2	0.491***	0.704***	0.842***
	(0.118)	(0.169)	(0.136)
$StDev(\hat{\theta}_{ms})$ Buyer Decile 3	0.515***	0.568***	0.482*
	(0.111)	(0.113)	(0.253)
$StDev(\hat{\theta}_{ms})$ Buyer Decile 4	0.411***	0.532***	0.557***
	(0.083)	(0.120)	(0.169)
$StDev(\hat{\theta}_{ms})$ Buyer Decile 5	0.272***	0.284***	0.261
	(0.081)	(0.094)	(0.184)
$StDev(\hat{\theta}_{ms})$ Buyer Decile 6	0.232***	0.218**	0.263
	(0.077)	(0.098)	(0.168)
$StDev(\hat{\theta}_{ms})$ Buyer Decile 7	0.070	0.042	0.182*
	(0.064)	(0.099)	(0.091)
$StDev(\hat{\theta}_{ms})$ Buyer Decile 8	-0.018	0.016	0.001
	(0.096)	(0.127)	(0.145)
$StDev(\hat{\theta}_{ms})$ Buyer Decile 9	-0.042	-0.081	-0.121
	(0.098)	(0.115)	(0.148)
$StDev(\hat{\theta}_{ms})$ Buyer Decile 10	-0.006	-0.001	-0.042
	(0.054)	(0.060)	(0.073)
Additional Controls	yes	yes	yes
Product FEs	yes	yes	.
Quarter FEs	yes	yes	yes
Buyer FEs	.	yes	.
Buyer-Product FEs	.	.	yes
R^2	0.30	0.49	0.67
Observations	16,207	16,207	16,207

Notes: The dependent variable, SI_{bms}, is the search intensity of buyer b when undergoing search spell s in product m, measured as the (log) of the count of short lived relations during s. Fixed effects for the product category (HS4/5), the calendar quarter of the spell (this is the quarter in which the majority of the spell fell in or the quarter of the end of the spell) and the buyer are included in different specifications. Deciles over buyers' sizes are constructed over the total volume the buyer purchases in Bangladesh (across all products), with the first decile being the smallest buyers. $St.Dev.(\hat{\theta}_{ms})$ stands for the standard deviation of qualities of available suppliers in spell s in product m. $q_{bmt(s)}$, $\bar{\hat{\theta}}_{ms}$ and $N^s_{mt(s)}$, as defined in the main text, are included as controls in all specifications. All regressions include search spells that don't satisfy the criteria for re-allocation to existing suppliers or anticipated search. Standard errors clustered at the HS4 product code in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 18: Dispersion and Sorting

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Average Accepted Supplier Type $\bar{\bar{\theta}}_{bmt}\|\theta > \Theta_{bmt}$						
q_{bmt}	0.041***	0.041***	0.059***	0.058***	0.058***	0.048***	
	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.006)	
q_{bt}							0.033***
							(0.004)
$St.Dev.(\hat{\theta})_{mt}$	-0.171**	-0.165**	-0.177***	-0.155***	-0.124**		
	(0.072)	(0.069)	(0.056)	(0.057)	(0.059)		
$StDev(\hat{\theta}_{mt})$ Buyer Quartile 1						0.072	0.085
						(0.076)	(0.078)
$StDev(\hat{\theta}_{mt})$ Buyer Quartile 2						-0.121	-0.097
						(0.080)	(0.079)
$StDev(\hat{\theta}_{mt})$ Buyer Quartile 3						-0.155**	-0.137*
						(0.075)	(0.076)
$StDev(\hat{\theta}_{mt})$ Buyer Quartile 4						-0.182**	-0.193***
						(0.072)	(0.071)
Product FEs (HS6)	yes	yes	.	.	yes	yes	yes
Category FEs (HS4)	.	.	yes	yes	.	.	.
Year FEs	yes	.	yes	.	yes	yes	yes
Category-Year FEs	.	yes	.	.	.	.	.
Buyer FEs	.	.	yes	.	yes	.	.
Buyer-Category FEs	.	.	.	.	.	.	.
Buyer-Year FEs	.	.	.	yes	.	.	.
R^2	0.04	0.05	0.28	0.45	0.29	0.05	0.04
Observations	38,423	38,423	38,423	38,423	38,423	38,423	38,423

Notes: The dependent variable, $\bar{\bar{\theta}}_{bmt}|\theta > \Theta_{bmt}$, is the average type of supplier the buyer has accepted (is trading with beyond a one-off interaction) in the product-year combination. The size of the buyer is generated as the log volumes the buyer imports in the market-year or in the market, q_{bmt} and q_{bt}. The main regressor in all specifications, $StDev(\hat{\theta}_{mt})$ is the standard deviation of the qualities or types of suppliers available in the $-mt$ combination. This is interacted with dummies collecting the quartile in the size distribution in which the buyer falls, based on q_{bt}. Fixed effects at different levels are introduced in turn in the different specifications. Standard errors clustered at the HS4 product code in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 19: Probability of allocating an order to a known seller

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	=1 if allocated to known supplier (Reporting Mg.Eff. from Probit)						
$\bar{\hat{\theta}}^k_{oimt}$	0.050***	0.007***	0.032***	0.007***	0.007***	0.008***	0.007***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
$Med(\hat{\theta}^u_{oimt})$	0.008		0.016	0.014	0.003	0.008	0.014
	(0.020)		(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
$StDev(\hat{\theta}^u_{oimt})$	0.041**	0.032**	0.037**	0.030*	0.032**	0.034**	0.030*
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
$\bar{\hat{\theta}}^u_{oimt}$		0.013					
		(0.020)					
Volume order, logs				0.033***	0.033***	0.033***	0.031***
				(0.001)	(0.001)	(0.001)	(0.001)
Volume demand (prod-quart), logs				0.003	0.025***	0.014**	0.003
				(0.005)	(0.007)	(0.007)	(0.005)
Count buyers, logs					-0.055***		
					(0.014)		
Count suppliers, logs						-0.029**	
						(0.013)	
Duration order (log) days							0.003***
							(0.001)
Product FEs	yes	yes	yes	yes	yes	yes	yes
Quarter FEs	.	.	.	yes	yes	yes	yes
Buyer FEs	.	yes	yes	yes	yes	yes	yes
Observations	79,892	78,809	79,892	78,809	78,809	78,809	78,809

Notes: An observation in these estimations is an order. The outcome variable a^K_{oimt} takes value one if order o, in product category m at time t, is allocated to a supplier that is known by buyer i. $\bar{\hat{\theta}}^k_{oimt}$ is the average type of the known or existing suppliers to buyer i, relevant to the current order. Similarly, $\bar{\hat{\theta}}^u_{oimt}$ denotes the average type of all available suppliers that are unknown to the buyer. St.Dev. and Median correspond to Standard Deviation and Median over those objects. Product fixed effects are introduced at the level of HS6 codes. Quarter and buyer effects are included in some specifications. Standard errors clustered at the HS4 product code in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 20: Uncertainty and Size of Initial Order

	(1)	(2)	(3)	(4)	(5)	(6)
	Av. (Log) Size of All First Orders: $\overline{q^i}_{mt}$					
Av. (Log) Size NonFirst: $\overline{q^{ni}}_{mt}$	0.246***	0.208***	0.187***	0.178***	0.092**	0.067
	(0.042)	(0.042)	(0.042)	(0.042)	(0.045)	(0.046)
$St.Dev.(\hat{\theta})_{mt}$	-0.310**	-0.314*	-0.283*	-0.349**	-0.356**	-0.354**
	(0.145)	(0.164)	(0.163)	(0.168)	(0.158)	(0.157)
$\bar{\hat{\theta}}_{mt}$		0.576***	0.454**	0.434**	0.202	0.208
		(0.181)	(0.179)	(0.177)	(0.148)	(0.145)
Count Suppl. N^s_{mt}					0.150***	0.173***
					(0.027)	(0.029)
Av. (Log) Price						-0.823***
						(0.231)
Product FEs	.	.	yes	yes	yes	yes
Quarter FEs	.	.	.	yes	yes	yes
R^2	0.09	0.13	0.16	0.21	0.24	0.26
Observations	814	814	814	814	814	814

Notes: The dependent variable is the average log size of all the starting orders (first order between a buyer and a seller) placed in the HS6-quarter combination. Fixed effects for the quarter and HS4 code are introduced in some of the specifications. The regressors include: the average log volume of all the orders placed in the HS6-quarter combination, that correspond to non first time interactions and that are part of sustained relations (relations that last for more than a year); the mean and standard deviation in the types of available suppliers in the HS6-quarter combination; the count of available suppliers in the HS6-quarter combination; the (weighted) average price of the product (HS6) in the quarter. Standard errors clustered at the HS6 product code in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 21: Price Cost Margins

	(1)	(2)	(3)	(4)	(5)	(6)
Linear Trend	0.001**	0.001*	0.001**	0.001*	0.001*	0.001*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Av. Price Fabric, logs	-1.085***	-1.086***	-1.085***	-1.095***	-1.096***	-1.097***
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Number buyers, logs		0.084*				0.087*
		(0.05)				(0.05)
Volume demand (prod-quart), logs			0.039			
			(0.03)			
θ supplier				0.238***	0.235***	0.235***
				(0.03)	(0.03)	(0.03)
Av. θ alternative suppliers					0.083	0.100
					(0.08)	(0.09)
St.Dev. θ alternative suppliers					0.110*	0.102*
					(0.06)	(0.06)
Constant	4.722***	4.348***	4.201***	4.656***	4.506***	4.113***
	(0.15)	(0.29)	(0.43)	(0.17)	(0.22)	(0.38)
Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Buyer-Seller FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	53848	53848	53848	53848	53762	53762
R^2	0.611	0.611	0.611	0.612	0.613	0.613

An observation in these estimations is an order. The outcome variable is measured as (revenue - cost)/cost, which in the data, across all product categories and time periods, has a median of 0.87 and a Standard Deviation of 1.3. Standard errors are bootstrapped in all cases and clustered by HS6 categories. Time fixed effects are taken according to the quarter in which the order starts, irrespective of its span over time. Product fixed effects correspond to HS6 codes.

Table 22: Time Induced Endogenous Mobility

	(1)	(2)
	$\Delta_b \overline{q}_{st} = \overline{q}_{b_2;st} - \overline{q}_{b_1;st-1}$	
Δq_{st-1}	0.561	
	(0.502)	
Δp_{st-1}	-0.914	
	(0.854)	
q_{st-1}		-0.323
		(0.977)
q_{st-2}		-0.225
		(0.579)
p_{st-1}		1.655
		(1.685)
p_{st-2}		-1.585
		(1.536)
Product-Quarter FEs	yes	yes
Seller FEs	yes	yes
R^2	0.95	0.97
Observations	506	506

Notes: The dependent variable is the change in the buyer's size when the supplier switches buyers, $\Delta_b \overline{q}_{st} = \overline{q}_{b_2;st} - \overline{q}_{b_1;st-1}$. The main regressor is the change in past volumes traded with first incumbent buyer, $\Delta q_{st-1} = q_{b_1;st-1} - q_{b_1;st-2}$. Regressions include product-quarter fixed effects and a seller fixed effects. I include only seller-quarter (st) combinations where the following conditions are satisfied: (i) s is active in the panel for at least a year before t; (ii) s trades with b_1 for at least two consecutive quarters before switching to b_2; (iii) after switching, s and b_2 trade for at least two quarters before s returning to b_1. Standard errors clustered at the HS4 product code in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 23: Endogenous Mobility and Match Specific Errors

	(1)	(2) Δq_{st}	(3)
$\Delta_b \overline{q}_{st}$	0.529***	0.515***	0.515***
	(0.069)	(0.110)	(0.110)
$\Delta_b \overline{q}_{st}$ ×Move Up	0.101	0.109	0.109
	(0.108)	(0.177)	(0.177)
Time trend	-0.024		-0.042
	(0.024)		(0.177)
Seller FEs	yes	yes	yes
Product FEs	yes	.	.
Quarter FEs	yes	.	.
Product-Quarter FEs	.	yes	yes
R^2	0.40	0.76	0.76
Observations	1,548	1,548	1,548

Notes: The dependent variable is the change (gain or loss) in the demand from main buyer for seller s when switching buyers, $\Delta q_{st} = q_{b_1;st} - q_{b_1;st-1}$. The main regressor is the difference in the buyers' size $\Delta_b \overline{q}_{st} = \overline{q}_{b_2;st} - \overline{q}_{b_1;st-1}$. This is also interacted with a dummy that indicates whether the supplier has moved onto a "larger" buyer. Regressions include seller, product, quarter or product-quarter fixed effects. Standard errors clustered at the HS4 product code in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.