Exporters' product vectors across markets

Lionel Fontagné[a,*], Angelo Secchi[b], Chiara Tomasi[c]

[a] *PSE, Université Paris 1 Panthéon-Sorbonne and CEPII*
[b] *PSE, Université Paris 1 Panthéon-Sorbonne and LEM*
[c] *Department of Economics and Management, University of Trento and LEM*

Abstract

The paper provides an original empirical approach to investigate multi-product firms' export patterns across destinations by considering the whole mix of products exported by a firm, formally defined as a product-vector. The proposed methodology allows to take into account a firm's choice of both exporting and non-exporting a product to a destination and to consider different forms of product complementarity that can generate product combinations. The empirical analysis uses a panel of transactions level data for the universe of Italian and French firms and complements the existing evidence along a few dimensions. First, we show that there is a high level of sparsity: selection of products at destination is indeed very severe. Second, we document that firms export several different combinations of product vectors across markets. Relatedly a high level of diversity is detected also when considering the intensive margin, pointing to a substantial departure from a stable global product hierarchy. Finally, we provide evidence that at the same time there exists a stable component in firms' product vectors across destinations composed by products which are not necessarily the most important in terms of sales, suggesting rich form of complementarities across goods. Products belonging to this stable component are less likely to be discarded as a consequence of an exogenous shock such as the dismantling of the MFA quotas after accession of China to the WTO.

Keywords: multi-product multi-country firms, product vectors, sparsity, fickleness, stability
JEL codes: F14, L11, L22

*Corresponding author. *Email:* lionel.fontagne@univ-paris1.fr *Address*: Maison des Sciences Economiques (Office #301),106-112 Boulevard de l'Hpital 75647 Paris Cedex 13 - France.

Email addresses: `lionel.fontagne@univ-paris1.fr` (Lionel Fontagné), `angelo.secchi@univ-paris1.fr` (Angelo Secchi), `chiara.tomasi@unitn.it` (Chiara Tomasi)

1. Introduction

The literature on international trade has emphasized the relevance of multi-product firms. These firms account for a considerable share of international trade[1] and contribute to a great extent to shape aggregate outcome. Multi-product firms update their product portfolio in response to trade shocks, exchange rate movements or tougher competition and such intra-firm adjustments are essential mechanisms that enhance the overall productivity and the welfare gains from trade (Bernard et al., 2011b; Mayer et al., 2014; Chatterjee et al., 2013). Despite their prevalence and their importance, little is known about the extent to which multi-product firms diversify across destinations and the patterns of such diversification. Expansion into foreign markets requires decisions about which products to export and which countries to approach with these products and such decisions are the result of a complex combination of factors related to firms, markets and products characteristics.

The contribution of this paper is to provide an original empirical investigation of the structure of export decisions of individual multi-product firms across destinations, improving in so doing our understanding of their behavior in international markets.

From a technical point of view, the paper proposes an approach that differentiates with respect to existing studies on the topic. In line with Mayer et al. (2014), we consider the whole list of products exported – a product-vector – and we define as a firm's global product vector the overall set of products exported worldwide and as a local product vector the set of products exported to a specific destination. Novel with respect to this approach, in defining the local product vectors we explicitly consider a firm' choice of not exporting one of its products to a given destination when it is already exported elsewhere. This is a strategic choice that certainly defines the core competences of a firm. We also introduce an element of novelty by prioritizing the identity of the goods exported by a firm across destinations. Looking at the identity of products allows us to provide evidence on the different set of combinations exported by a firm across destinations and, at the same time, to detect those combinations of products that are more likely to be co-exported, irrespective of their importance in terms of export value. In line with previous studies, we characterize the structure of the product-vectors by exploiting the information on product export shares but we refine the analysis by using novel metrics that overcome some of the limitations of the Spearman correlation index, which is the measure typically used within this literature.

We apply this empirical strategy to the universe of Italian and French multi-product exporters for the period 2000-2007, improving upon our knowledge on the structure of export decisions of multi-product exporters along three specific dimensions: sparsity, fickleness and stability. Throughout our empirical analysis, we highlight the importance to check if the observed regularities are statistical artifacts compatible with a null model of "random diversification" along the lines of Armenter and Koren (2014) or, on the contrary, they contain relevant information on the decision making of multi-product exporters in different destinations.

By using a suitable metric that assesses the distance between vectors, the Levenshtein dissimilarity index, we investigate how different is a firm's local product vector with respect to its global one, that is how common is for a firm not to export one of its products to a given destination even if it does export it elsewhere. Selection of products at destination emerges as very severe: the vector of products exported to each destination contains, on average, at least 80% of zeros with respect to the worldwide portfolio of products exported by a firm. The frequency and pervasiveness of the zeros within a firm across destinations is a novel result

[1] In the US, only 30% of exporters ship ten or more HS6 products and account for 97% of all exports (Bernard et al., 2009). We find similar figures in Europe: in Italy and France 42% and 40% of exporters ship more than five HS6 products and account for 96% and 95% of total export flows, respectively. This pattern is also observable in developing countries. In Brazil 25% of exporters ship ten or more HS6 products and account for 75% of total exports Arkolakis et al. (2016), while across 32 developing countries, the top five exporters make up 30% of total exports (Freund and Pierola, 2015).

within the existing evidence. Our result complements the observation of a strongly right-skewed distribution of the exporters' number of goods (the exporter scope), with most firms selling only one or two goods (Arkolakis et al., 2016).

The large number of zeros observed in the list of products potentially sold by a firm in a destination opens the possibility of having a high number of different combinations of products in the different destinations served by this exporter. Systematic differences in the arrangements of products exported by a firm across destinations confirm that product mixes are fickle across markets: 1/3 of French and Italian firms export a different product-mix in each destination, while on average two local product vectors belonging to the same firm and with the same number of active products differ in more than 50% of their composition.

Such high level of variability is difficult to be reconciled with the existing theoretical models on multi-product firms which assume a product hierarchy either based on production efficiency (Eckel and Neary, 2010; Mayer et al., 2014; Arkolakis et al., 2016) or on product quality (Manova and Yu, 2017; Manova and Zhang, 2012; Eckel et al., 2015). In both types of models, selection of products exported by a firm across markets can be predicted by its core competences. Assuming a product ladder where productivity or quality falls for each additional variety produced, firms enter export markets with their most efficient or higher quality product first, and then expand their scope moving down the ladder of efficiency/quality.[2] As a result the same product hierarchy should be observed across markets. Deviations from this perfect product ladder are not compatible with these models unless one explicitly assume unobserved shocks at the product-destination or firm-product-destination level as in Bernard et al. (2011b) and Arkolakis et al. (2016). Within this literature, the statistical footprint of a common pecking-order of a firm's products across destinations is provided by the rank correlation between local and global rankings of its product sales. By using the Spearman rank correlation, the existing studies suggest that, although far from being perfect, deviations from a product hierarchy seems to be only marginal.[3]

This paper shows that, once the high level of zeros is properly accounted for, the idea that firms adhere to the same global hierarchy in every destination served is much less supported by data. To overcome some of the limitations of the Spearman index, we use the Bray-Curtis measure to quantify the compositional similarity between vectors. We provide evidence of systematic differences in the relative importance in terms of revenues of a firm's products across destinations. Traditional determinants of the composition of the exported mix, such as size and toughness of competition in the destination market, firm efficiency and product quality matter, but our econometric analysis show that they do not quantitatively impact the large presence of zeros and the diversity of a firm's product vectors across destinations. The latter results confirm that destination-specific characteristics or firm-product attributes such as efficiency or quality alone cannot explain all the variation observed in the data and that demand factors idiosyncratic to the firm are important too (Eaton et al., 2011; Parenti et al., 2017; Osharin et al., 2014; Forlani et al., 2016; Comite et al., 2014). Indeed, models with consumer preferences that are asymmetric across varieties and heterogeneous across countries display enough versatility and are consistent with the variability observed in the data.

The last novel element we detect in the data concerns the existence, together with all the above diversities, of a common component in product mixes a firm's export that is stable

[2] Alternatively, Eckel and Neary (2010) consider a flexible manufacturing approach, where firms face declining efficiency in supplying less-successful products away from their core competency. In the same model, on the demand side, new varieties reduce the demand for firms' existing varieties. Other models assume that products are symmetric on both the demand and supply sides. As a result, the same amount of all products is sold. See, e.g. Nocke and Yeaple (2014); Feenstra and Ma (2008); Dhingra (2013).

[3] Mayer et al. (2014) find a correlation coefficient between French firms' local and global product rank of 0.68. Arkolakis et al. (2016) calculate the correlation in Brazilian firm-product sales' ranking by destination, using either the US or Argentina as the reference country, and find an average coefficient of 0.85. Manova and Yu (2017) report for Chinese firms a rank correlation for each firm-destination pair of about 0.69.

across destinations. Product complementarity or technological relatedness may imply that the production and thus export of one product leads to the production and export of its components or of other related goods. This dimension of stability shows up in the data by looking at the extent to which there are combinations of products that are more likely to be co-exported, irrespective of their importance in terms of export value.[4] We observe that on average 2/3 of French and Italian multi-product firms show subset of products typically co-exported that we define as the typical product vector. The probability of co-exporting these products tends to be systematically higher than that of exporting one of these products alone. Also, we observe significant departures from the theoretical benchmark of core competences following a rigid ordering in the product export sales across destinations: on average between 1/4 and 1/3 of the typical product vectors are composed by products which are not the most important in terms of sales. By considering the variation in export prices and quality across products, we provide evidence that the choices governing products belonging to a firm's typical product vectors may not be directly linked to product quality or efficiency.

To go beyond statistical regularities, we examine the response of firms' product vectors to an exogenous shock: the increase in competition from China and the removal of MFA quotas on textiles and apparel. We show that products belonging to the typical product vector are less likely to be discarded by a firm's product portfolio, even if they are not so relevant from a quantitative (sales) or qualitative (quality) point of view. The evidence points to richer forms of interdependence across products that can be due to technical, organizational, or strategic reasons.[5] By exploiting the information on the the good classification according their main end use, we suggest that technological factors are important factors that trigger the complementarities observed across products.

Our study is related to the relatively recent and rapidly evolving literature on multi-product firms in international markets. In Bernard et al. (2011b), a firm-product participation across markets is driven by firm-product specific attributes and by product-destination and firm-product-destination unobserved shocks that determine a firm's scale and scope of sales in different markets. While Mayer et al. (2014) investigate the consequences of demand shocks for the skewness of exports in the cross-section across destinations, Mayer et al. (2016) examine the response of skewness within a destination over time using as an exogenous demand shocks the change in the GDP at destination. Iacovone and Javorcik (2010) consider the North-American regional integration experience as a competitive shock leading Mexican firms to adjust their product scope. Product hierarchy can be also governed by quality: firms' core competences are in varieties with superior quality that command higher sales and prices (Eckel et al., 2015; Manova and Yu, 2017).[6] In Manova and Yu (2017) there is a rigid ordering of products within each firm based on the level of quality: as the firm extends her scope to marginal products, sales decrease and quality as well. Using Chinese firm-level data, the authors show that in response to the exogenous removal of quotas on textile and apparel quotas firms expand their product scope by adding lower quality varieties. Similar evidence is provided by Martin and Mejean (2014) that focus on between rather than within firms movements. They show that the quality of French firms' exports has increased in response to the growing market share of

[4]Bernard et al. (2010) pioneered this idea by focusing on co-production of goods within a firm.

[5]In this respect, the industrial organization literature has largely discussed the reasons why firms diversity and the existence of possible complementarities across goods (see the classic review in Bailey and Friedlaender (1982) and Montgomery (1994)). For example Stigler (1963) and Whinston (1990) explain product bundling/tying for price-discrimination and market-power leverage purposes, respectively. A typical example of demand complementarity, associated with compatibility issues, is in Yu and Wong (2015). Baldwin and Woodard (2009) provide an overview of production platforms, another potential reason of interdependence across products within a firm.

[6]Several other papers, such as Khandelwal (2010); Hallak and Schott (2011); Baldwin and Harrigan (2011); Crozet et al. (2011); Kugler and Verhoogen (2012); Sutton and Trefler (2016), have highlighted the role played by product 'quality', where quality is typically modeled as a demand shifter capturing therefore very different effects.

low-wage countries on destination markets.

With respect to previous empirical analyses we emphasize additional element of complexity characterizing multi-product firms' decisions. Indeed, while providing a useful and coherent framework to investigate decisions of multi-product exporters, the existing theoretical framework necessary leaves out of the picture some elements of complexity. Our empirical work points to the need of further addressing the role of product selection and the richer forms of product complementarities identified here.

The remainder of the paper is organized as follows. Section 2 describes the data used in the empirical analysis and it introduces the concept of product vector. Section 3 presents the some empirical regularities on firms' product vectors concerning sparsity, that is the frequency and the pervasiveness of zeros, and fickleness, i.e. the diversities in firms' product vector across destinations. Section 4 brings additional evidence on the stable component of a firm's product vector across markets. Section 5 concludes.

2. Definitions and data

This section operationalizes the concept of firms' product vectors through which we address the behavior of multi-product firms in international markets. We highlight how this complements the previous approaches. We also present our main data sources and describe the simple statistical benchmark used to compare our predictions with those of a simple random-assignment model.

Global, Local and Typical product vectors

We provide an empirical approach that allows to take into account the entire mix of products exported by a firm, including a firm's strategic decision of not exporting certain products to a destination while exporting the same products somewhere else, revealing in so doing the variety of capabilities of a firm. To this aim we formally define a product mix as a vector that we label *product vector*, and more specifically we introduce three concepts: the *Global Product Vector* (GPV_f), the *Local Product Vector* (LPV_{fd}) and the *Typical Product Vector* (TPV_f).

First, the global product vector is the overall set of products exported worldwide by a firm. Formally, a GPV_f is an ordered vector of 1s whose dimension is equal to the total number of products exported by firm f.

Second, the local product vector in a destination captures the combination of products a firm sells to that foreign market, explicitly including elements of the GPV_f actually exported to that destination and those that are not. This definition allows to control for a firm's choice not to export a given product to a given destination, when it does export it elsewhere. This an economic choice that likely contains relevant information. Formally, a LPV_{fd} is an ordered binary vector of the same dimension of the GPV_f reporting 1 when a product is actually exported by firm f to destination d and 0 otherwise. All the local product vectors of a given firm have therefore the same length as the corresponding global vector.

Figure 1 reports an example for a firm's GPV and its different LPVs.

$$\begin{array}{c}\text{HS 6 code}\\ 850152\\ 850110\\ 850151\\ 850300\\ 847990\\ 842430\end{array} \quad \text{GPV}_f = \begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix} \quad \text{LPV}_{fd_1} = \begin{pmatrix}1\\0\\1\\1\\0\\0\end{pmatrix} \quad \text{LPV}_{fd_2} = \begin{pmatrix}0\\1\\0\\0\\1\\1\end{pmatrix} \quad \text{LPV}_{fd_3} = \begin{pmatrix}1\\0\\1\\1\\0\\0\end{pmatrix} .$$

Figure 1: Global product vector (GPV) and local product vectors (LPV) of a firm exporting six different products (HS6) to three destinations. Each element of the vector takes the value one if the corresponding product is exported, and zero otherwise.

Third, we define the typical product vector (TPV_f) as the most frequent combination of products exported across destinations, conditional to be exported in at least two served markets. Figure 2 presents the TPV for the firm in the example above. A firm may not have a TPV at all.

$$\begin{array}{c}\text{HS 6 code}\\ 850152\\ 850110\\ 850151\\ 850300\\ 847990\\ 842430\end{array} \qquad \text{GPV}_f = \begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix} \quad \text{TPV}_f = \begin{pmatrix}1\\0\\1\\1\\0\\0\end{pmatrix} .$$

Figure 2: Typical product vector (TPV) of an exporting firm.

Working with product vectors as simple lists of HS6 codes has the principal advantage of allowing to identify the effects of different forms of product complementarity (demand- or supply-driven) that can generate product combinations with equal parts of high- and low-value items: an example is Apple Inc., which exports both its I-pad and the dedicated Apple pencil.[7] However, it is clear that not all the products a firm exports to a destination play the same strategic competitive role and that export sales may differ substantially across products, both worldwide and within a destination. We therefore associate product market shares to the product vectors replacing the 1s in the GPV_f by a firm's worldwide product market share and the 1s in the LPV_{fd} by a firm's product market share in the destination. The GPVs and LPVs continue to have the same length and their components sum up to 1, as shown in Figure 3.

$$\begin{array}{c}\text{HS 6 code}\\ 850152\\ 850110\\ 850151\\ 850300\\ 847990\\ 842430\end{array} \qquad \text{GPV}_f = \begin{pmatrix}0.4\\0.15\\0.3\\0.04\\0.1\\0.01\end{pmatrix} \quad \text{LPV}_{fd_1} = \begin{pmatrix}0.6\\0\\0.35\\0.05\\0\\0\end{pmatrix} \quad \text{LPV}_{fd_2} = \begin{pmatrix}0\\0.7\\0\\0\\0.2\\0.1\end{pmatrix} \quad \text{LPV}_{fd_3} = \begin{pmatrix}0.5\\0\\0.45\\0.05\\0\\0\end{pmatrix} .$$

Figure 3: Market shares associated with the global product vector (GPV) and the local product vectors (LPV).

Comparison with the sales share approach

Using worldwide or destination product export sales ratio is closer to the approach generally used in the literature and exemplified by Mayer et al. (2014) and in the related working paper version (Mayer et al., 2011). The main prediction of their model is that exporters skew their export sales towards their best performing product in presence of tougher competition, following their core competency ladder. They use a cross-section of (manufacturing) firm-level exports across destinations for one exporting country (France) and empirically check whether the "product ladder" of a given exporter, as revealed by the pecking order of product export sales, is stable across destinations. To proceed, they define two ladders – *global* (the ranking of products of an exporter according to its total exports) and *local* (the ranking in the considered destination). The prediction on the skewness of a firm's exported product mix across destinations is brought to the data using three strategies. First, the product rankings should be reflected by the rank correlation between local and global rankings. Mayer et al. (2014)

[7] In the example provided in Figure 1 the TPV is composed by two types of motors with different powers (HS6-850152 and HS6-520151) and a product that represent the parts and components of these final good (HS6-850300).

report a Spearman rank correlation of 0.68, which points to "substantial departures from a steady global product ladder".[8] Second, beyond the ranking, Mayer et al. (2014) consider the ratio of sales in the best performing product to the second (or third) best product, and this can be computed using either the global or local ranking of products in the ladder. Third, the skewness of export sales can be measured directly using the standard deviation of export sales, a Herfindhal index or a Theil index, considering the set of products exported to a destination.[9]

What one does not observe with this methodology is: i) the entire scope of the local product mix including the choice of not exporting a product to a certain destination; ii) the identity of products belonging to this mix; iii) the existence of a frequent combination of products within the exported mix – what we coined as a typical product vector. Accordingly, the method we propose complements nicely the former approach. For each firm, the identity of products exported (the list of products identified by their code/name) is the statistical footprint of her capabilities. This list is a (global) vector comprising only ones, and we observe a vector of zeros and ones at each destination for the same firm, and certain arrangements of the ones can be most frequent.

Micro-level trade data

The concepts of global, local and typical product vectors are applied here to the universe of French and Italian manufacturing exporters. The annual data on firm-product-destination level export flows are in values and quantities.[10] Product categories are classified according to the Harmonized System and are reported at the six-digit (HS6) level.[11] For both countries we exploit the panel and use the information available for the 2000-2007.[12] Since we focus on the variability of firms' exports across products and countries, we exclude those companies exporting only one product or serving only one destination.

Table 1 presents the total value of exports and the number of firms in the whole and restricted samples for Italy and France for two different years, 2000 and 2007. In the restricted sample the number of firms falls substantially but the value of exports changes only marginally. Indeed, manufacturing firms exporting several products to different destinations account for almost 98% of a country's total exports. To ensure that the results reported in the paper are not driven by particular features about Italian and French data, we replicate in Appendix A some of the stylized facts emerged so far in the literature on multi-product firms concerning firms' product extensive margins, their scope and scale (Arkolakis and Muendler, 2013; Arkolakis et al., 2016). The similarity in the results gives us confidence about the external validity of our results.

An illustrative example

To illustrate the new approach of firms' product vectors, we look at a real example taken from our data set. This is indeed just an example aiming at clarifying the way our indicators are constructed and one should refrain from drawing conclusions based on this example. We will systematize below the use of these indicators.

[8]Mayer et al. (2014) show that the correlation is not driven by single-product firms and income at destination does not profoundly affect the correlation either.

[9]This third approach is completed in Mayer et al. (2011) by reporting the distribution of within-firm sales to a destination: the average share of sales of a product plotted against its local rank fits well a Pareto distribution although departures are observed at the two extremes of it.

[10]The datasets were accessed at ISTAT and Banque de France facilities and were made available for analysis after careful screening to avoid disclosure of individual information.

[11]This is the finest level of disaggregation available for Italy, while the French customs data are available at the eight-digit (CN8) level. However, we consider the six-digit classification to make the analysis comparable across the two countries. This approach also gives us a better chance of finding the stable product lists the literature proposes.

[12]Because some product categories are assigned different HS6 product codes at different points in time, we use concordance tables provided by Eurostat to harmonize the classifications to the 1996 version.

Table 1: DESCRIPTIVE STATISTICS ON EXPORT VALUE AND NUMBER OF EXPORTERS, 2000 and 2007

	Year	Whole sample	Restricted sample
ITALY			
Total exports (billions Euro)	2000	194.1	190.0
	2007	287.0	281.5
# Exporters	2000	77,477	46,325
	2007	69,363	42,622
FRANCE			
Total exports (billions Euro)	2000	204.6	201.6
	2007	283.1	279.5
# Exporters	2000	35,796	21,967
	2007	31,798	18,632

Notes: In the restricted sample we keep only those firms exporting more than one product and serving more than one destination.

We consider an Italian firm producing electrical motors, shipping 16 different products to 20 different destinations. We do not identify this firm and omit, on purpose, some information (e.g. on certain destinations) for confidentiality reasons. The example is illustrated in Figure 4, where each color represents a different HS6 product and products are ranked according to the lexicographic order from the HS classification. The destinations on the X-axis are ranked according to their share of total export value. The left top panel shows the LPV_{fd} of this firm in each of the 20 destinations served, i.e. the combination of products a firm sells to each foreign market. The right top panel includes also the GPV_f, i.e. the overall set of products exported by this firm worldwide. The bottom left panel reports the TPV_f for this firm. Finally, the bottom right panel shows the product vectors for this firm with product market shares. Figure 4 illustrates the richness of the information contained in the global and local product vectors for the selected firm.

First, in each destination the firm exports only a small subset of its GPV: LPVs indeed contain a large number of zeros. In other words, the non-export of a product to a destination is frequent, underlining the importance of controlling for product selection when looking at firm diversification in international markets. This illustrates the sparsity of the product vector. Moreover, we note that the combinations of products exported by this firm appear to a large extent fickle across markets: it exports 14 different LPVs to the 20 active destinations. By considering product export shares we add a further dimension of variability. The firm's GPV hierarchy in terms of market shares does not appear in any of the 20 destinations. In this example, the explanation is simple: the US is a large market, and the first product exported to the US accounts for a large share of a firm's total exports. This product is specific to US demand patterns in terms of electrical motors.

Second, while some variability is detected across destinations driven by the large number of zeros and by then different combinations of product exported, we also observe that there are some combinations of products that are more likely to be exported than others. Indeed, the firm possesses a typical product vector suggesting that there is a structural component in the product vectors of a firm that is stable across destinations. Looking in more detail, it exports the typical product vector to four out of its top 10 destinations, which cover 90% of the firm's total exports (DEU,FRA,GRC,NLD). The typical product vector is composed by three products: "AC Motors" of two different powers (orange and red) plus an additional generic good labeled "Parts of these motors" (yellow). This illustrates stability. Importantly, when looking at the sales share of these products, we note that goods with very high sales value are exported together with products with low sales value. Two of these are specific types of electric motors, and account for 52% and 11% of the firm's total exports. The third

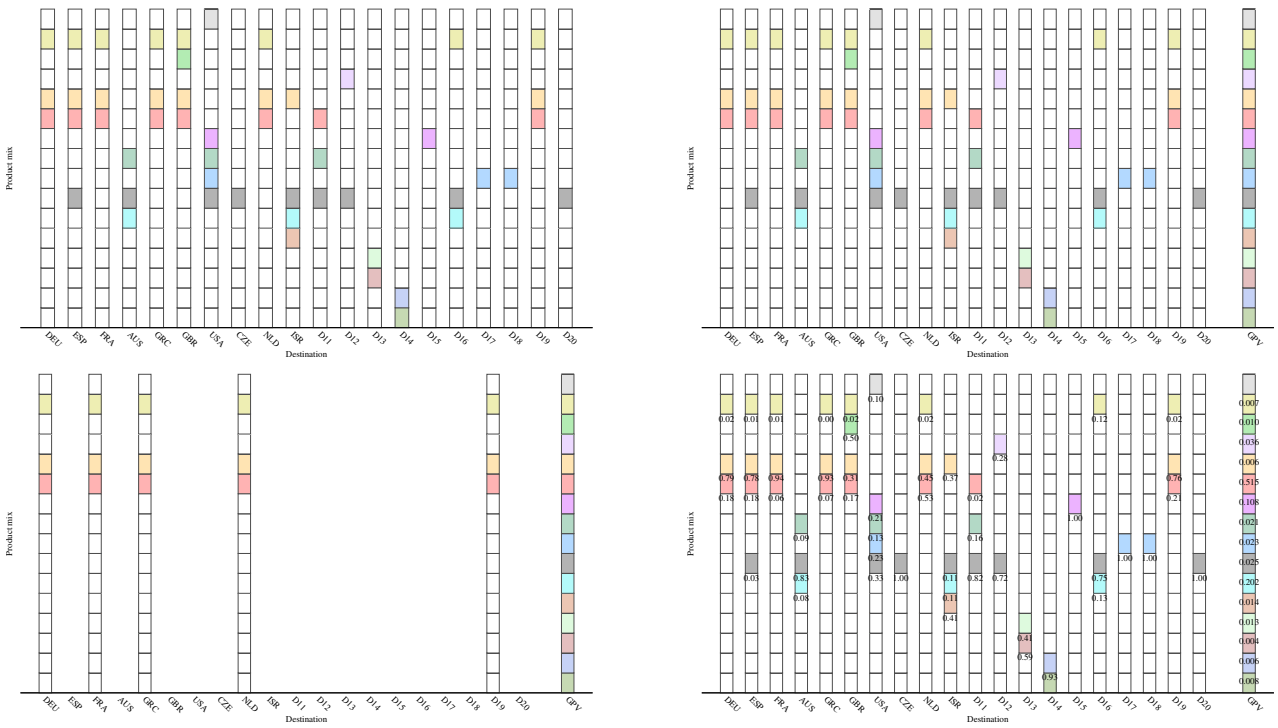

Figure 4: The figure reports for a firm producing electrical motors the local product vectors (left top panel), the global product vector (right top panel), the typical product vector (bottom left panel), and the product vectors with market share (bottom right panel). Each color represents an HS6 code and products are ranked according to the lexicographic order from the HS classification.

is labeled parts of motors, with a market share of only 1%.[13] In that peculiar example, a group of products is typically sold to a destination, even it deviates from the rank of sales and presumably profitability. Accordingly, stability reflects a richer form of interdependence across products due to technological complementarity between products. This type of configuration is not the one suggested by the international trade theoretical literature.

In what follows, we will systematize this approach and characterize the export patterns of multi-country and multi-product firms. This will provide a set of empirical regularities that improve upon our understanding on multi-product firms' behavior. However, before proceeding it is important to ascertain that observed regularities are not statistical artifacts driven precisely by the sparsity of our large data. To do so, we need a statistical benchmark that we now present.

Statistical benchmark

A major caveat in interpreting stylized facts on GPVs, LPVs and TPVs comes from the possibility that the observed regularities may represent statistical artifacts simply induced by few key features of a firm's export structure such as the product-, destination- and product-destination extensive margins. Armenter and Koren (2014), for example, show that in presence of highly sparse data a simple balls-and-bins model is able to quantitatively reproduce the pattern of zero product- and firm-level trade flows across destinations and the frequency of multi-product, multi-destination exporters. In order to overcome this potential limitation we provide the reader with a "null model" against which comparing our empirical results. This model represents a statistical benchmark in which we fix some of these key properties of a firm's export structure but we randomize the exact composition of a firm's LPVs generating a sort of

[13] We observe some variability also within this stable component as the relative importance of the three products exported to the top-EU destinations, while stable in a first approximation, changes across markets. Exports to Germany are mainly of high-power AC motors (of between 750W and 75KW), while in the Netherlands the greatest export share comes from low-power AC motors (with power below 750W).

random diversification model.

To build this statistical benchmark we implement for each firm the following procedure:

1. We set its number of products, destinations and product per destination equal to those observed in the data. This implies that the GPV remains equal to the one observed;
2. Within each destination we randomly select which products are exported, that is we randomly reshuffle the 1s in each LPV;
3. We compute the TPV (if any);
4. We compute descriptive statistics on LPVs and TPV;
5. We repeat the procedure 50 times and we average descriptive statistics among them.[14]

Three remarks are in order. First, this procedure implies that a firm's product-, destination- and product-destination extensive margins are unaffected by the randomization. Second, computationally finding a binary random matrix satisfying the constraints implied by the procedure above, especially for firms with several products and few destinations, might be very time consuming. For this practical reason sometimes we allow for minor violations of the constraints, in terms of the number of destinations where a product is exported.[15] Third, when we perform the random reshuffling of the products exported by a firm to a destination we might generate triplets firm-destination-product that do not exist in our dataset and for which we do not observe the associated export value. For this reason, in the following, we cannot exploit the statistical benchmark in those investigations involving export shares.

In the rest of the paper we then report, when possible, results computed on the randomly generated data for France and Italy and we compare them with those obtained with the original data. The results confirm that the observed regularities are not a statistical artifacts induced by few key features of multi-product, multi-destination exporters.

3. Diversities in product vectors

Building on the illustrative example above, this section explores the differences among firms' product vectors across destinations providing evidence on the behavior of multi-product firms. Using a suitable metric we start by showing that the vast majority of the elements in a firm's LPVs are zeros, a fact in line with the existence of a strong selection of products across destination markets. Next, we use the same metric to show that a firm's LPVs tend to be diverse also in terms of their composition even after controlling for selection. We then turn to the role of the intensive margin and we show that properly assessing the extent to which firms adhere to their global product hierarchy requires to account for the sparse structure of the LPV data and to fully exploit the information contained in market shares. Finally, by mean of regression analysis, we document that, while traditional factors such as size and toughness of competition in the destination market and firm efficiency and product quality matter, they do not quantitatively impact the large presence of zeros and the diversity of LPVs across destinations.

Product selection across destinations

Traditionally the trade literature has emphasized the importance of the extensive margin at the firm level: selection through entry and exit contributes to a great extent to shape observed international trade patterns.[16] A more recent literature on multi-product firms has turned the

[14]The exact way in which we average across the 50 replications depends on the specific empirical exercise we perform and details on that are provided case by case along the paper.

[15]This happens for 4 out of 45,276 firms in Italy and for 6 out of 19,506 firms in France.

[16]Building on Melitz (2003)'s model with heterogeneous firms, Helpman et al. (2008) and Chaney (2008) among others, developed trade models that explicitly consider the extensive margin of firm entry and exit. Empirical studies have analyzed how the firm' extensive margin is affected by destination-specific factors such as destination income, market size, economic and political integration, as well as variable and fixed trade costs.

Table 2: STRUCTURE OF PRODUCT VECTORS

	Year	Mean	Std. Dev.	Min.	1Q	Median	3Q	Max.	Obs.
					ITALY				
# Products	2000	11.23	16.00	2.00	3.00	6.00	13.00	413.00	46,228
	2007	12.34	19.39	2.00	3.00	6.00	13.00	520.00	42,539
# Products in LPV	2000	2.37	2.16	1.00	1.35	1.80	2.56	86.05	46,228
	2007	2.45	2.59	1.00	1.33	1.75	2.58	70.89	42,539
# Destinations	2000	12.47	12.87	2.00	3.00	8.00	17.00	118.00	46,228
	2007	14.95	15.51	2.00	4.00	9.00	21.00	134.00	42,539
# LPV	2000	7.63	8.26	1.00	3.00	5.00	9.00	107.00	46,228
	random	[10.38]	[11.49]	[1.00]	[3.00]	[6.00]	[13.94]	[118.00]	[46,228]
	2007	8.35	9.58	1.00	3.00	5.00	10.00	116.00	42,539
	random	[11.92]	[13.64]	[1.00]	[3.00]	[6.00]	[16.00]	[124.00]	[42,539]
					FRANCE				
# Products	2000	12.64	20.45	2.00	3.00	6.00	13.00	615.00	21,726
	2007	13.18	22.72	2.00	3.00	6.00	14.00	608.00	19,246
# Products in LPV	2000	2.52	2.49	1.00	1.33	1.80	2.74	48.48	21,726
	2007	2.62	3.19	1.00	1.33	1.77	2.72	76.00	19,246
# Destinations	2000	12.62	14.85	2.00	3.00	7.00	16.00	159.00	21,726
	2007	13.75	15.79	2.00	3.00	8.00	18.00	162.00	19,246
# LPV	2000	7.84	9.72	1.00	2.00	4.00	9.00	139.00	21,726
	random	[10.68]	[13.60]	[1.00]	[3.00]	[5.00]	[13.00]	[149.00]	[21,726]
	2007	8.13	10.17	1.00	2.00	4.00	9.00	144.00	19,246
	random	[11.33]	[14.34]	[1.00]	[3.00]	[6.00]	[14.00]	[158.00]	[19,246]

Notes: The statistics are calculated on the restricted sample where we keep only those firms exporting more than one product and serving more than one destination. Statistics for the statistical benchmark (random) are computed for 2000 and 2007 averaging over 50 replications.

attention to the role of selection among products within a firm and to the factors driving it. Decisions of which good to offer in each market have been shown to depend on production costs, product quality, destination country-specific characteristics such as market structure, competition, consumer's preferences and income. Two issues have been largely neglected however. First, little is known on the pervasiveness of the zeros within a firm across destinations. Second, there is no evidence concerning the identity of products exported by a firm across destinations. Investigating these two aspects throughout the lens of the GPV and of the LPVs will turn to be useful to better understand the behavior of multi-product firms in international markets.

We begin by asking how common is that a firm does not export a product in the GPV to one of its already active destination, that is how severe is the within firm product selection. Table 2 presents informative descriptive statistics for LPVs of Italian (top panel) and French (bottom panel) exporters for 2000 and 2007.[17] The first set of rows of each panel reports the distribution of the number of products exported by each firm. Italian and French average exporters are remarkably similar in terms of their product diversification: the average number of products exported is in both cases around 12.[18] The second set of rows of each panel shows the descriptive statistics for the number of active products in the LPV (i.e. the number of ones in the LPV). The average number of products in the LPV is around 2.5 in Italy and France. This simple figure, similar for the two countries, suggests that on average an exporter sells only a small subset of its products to each destination: LPVs are very sparse.

[17]Figures for other years in the data set look very similar to the ones shown. They are available upon request.

[18]If we consider the whole population of exporters, the average number of products exported by both Italian and French firms is 8.5 which is in line with what observed in other micro-level dataset.

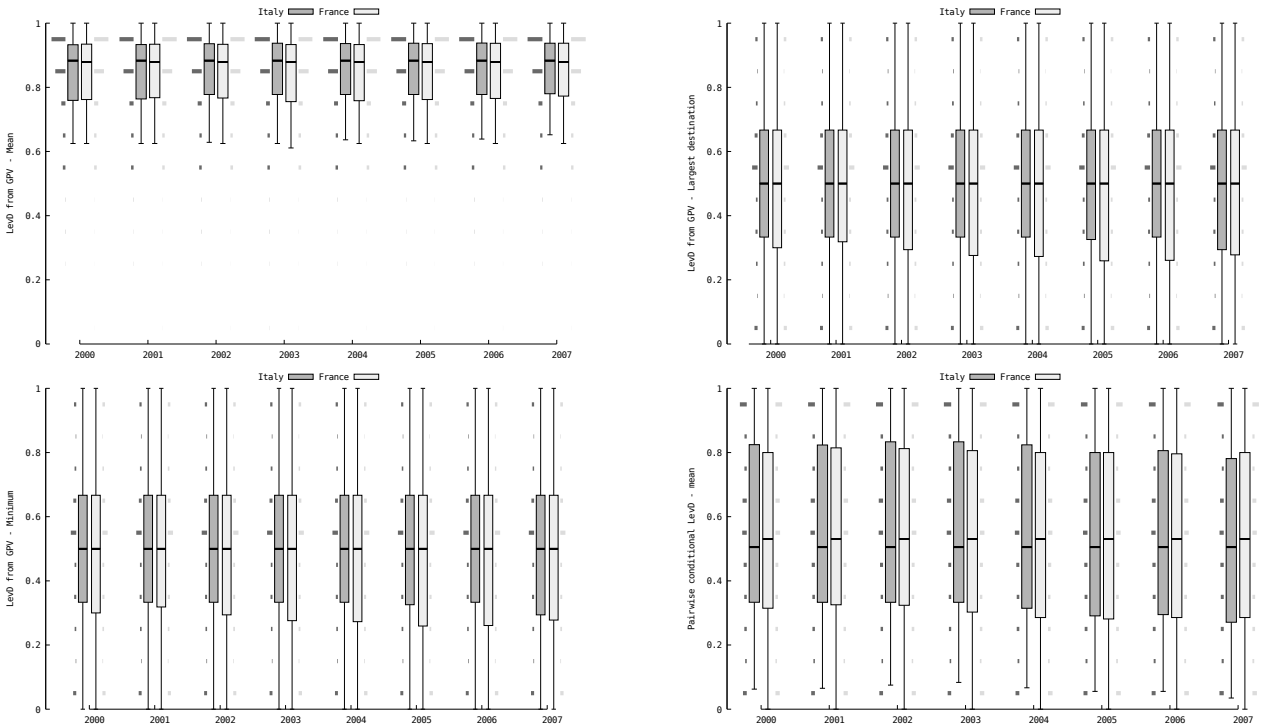

Figure 5: Firm-level box-plots of the average Levenshtein dissimilarity index (LevD) between a firm's global product vector (GPV) and local product vectors (LPV) using the entire sample (top-left panel), using only the destination with the largest revenues (top-right panel) and using the destination with the lowest product scope (bottom-left panel). For each firm, the LevD is computed between each LPV and the corresponding GPV and then averaged over its destinations. Bottom-right panel report the average pairwise Levenshtein distance among a firm's destination with the same product scope. Box-plot represents inter-quartile range with the corresponding median. Whiskers bars cover from the 10th to the 90th percentile of the distribution. To guide the reading of the figure the corresponding histogram is also reported.

While the average number of products exported on a given destination gives us a prima facie idea of the level of sparsity, to dig deeper in this direction we compute the distance of LPV_{fd} from the corresponding GPV_f by using a normalized version of the Levenshtein dissimilarity index (LevD from now on). Also known as the Edit distance, LevD is a string metric developed to measure the difference between two sequences and in our case it captures the share of zeros in a LPV.[19] For each firm we compute the LevD between each LPV_{fd} and the corresponding GPV_f.

The top-left panel in Figure 5 reports the box-plot together with a 10%-90% whisker bar of the distribution of the LevD (averaged within firms across their destinations) for Italy and France in different years. The mean and the median of this distribution are higher than 0.80 in both countries and they are very stable across years. The implication is that on average a LPV contains 80% of zeros with respect to the corresponding GPV. While hiding a certain degree of heterogeneity this average does not misrepresent the phenomenon: indeed 50% of the mass of firms lies in terms of LevD in 0.75-0.95 and 80% in 0.65-1. This makes apparent that exporters do sell in each destination only a small subset of the GPV which can be seen as representing the full potential of the firm in terms of range of products.

One may suspect that the large presence of 0 in LPVs is driven by marginal destinations

[19]Formally, it represents the minimum number of single edits (insertion, deletion, substitution) required to change one sequence into the other divided by the number of elements in the longest sequence minus 1. See Appendix B for more details regarding this measure. Note that in our case we need to consider only substitutions since all the product vectors have the same dimension. In this case this metric is also known as Hamming distance.

or marginal products which account for a small fraction of a firm's total export value.[20] To explore the robustness of our results along both dimensions we perform two stress-tests of our exercise. In the top-right panel of Figure 5 we compute the LevD between the GPV and the LPV only for the largest destination of each firm. While the inter-quartile range increases, suggesting a higher overall heterogeneity in LevD across firms, the mean remains close to 0.70. Even in the largest destination firms tend to export a subset of their products as small as 30% of the GPV. Similarly, bottom-left panel of Figure 5 replicate the computation of the LevD focusing only on a firm's destination with the highest product scope, that is with the lowest LevD. Although this is a very demanding exercise, results appear reassuring. While the the inter-quartile range further shifts downward, the mean of the LevD remains as high as about 0.48. Comforted by these checks we conclude that sparsity is a robust feature of LPVs: within a firm product selection at destination is very tough.[21]

Moving to the second issue neglected in the literature, the large presence of zeros in LPVs opens the possibility of observing diverse combinations of products in the different destinations. To explore this possibility we check to what extent the combinations of products exported by a firm in different destinations are diverse. Descriptive statistics reported in Table 2 show that Italian and French exporters display similar geographical diversification patterns: they are active, on average, in between 12 to 15 destinations depending on the year and they possess, on average, approximately up to 8 different LPVs.[22] The same data reveals that about one third of the firms export a different LPV in each destination and that these latter firms account for between 5 and 10% of a country's total exports. These evidences suggest that the average exporter tend to sell diverse product combinations in different destinations. In order to go beyond these indirect evidences and to quantitatively assess the variability across destinations in the composition of LPVs, we revert to the LevD metric. For each firm we compute the average pairwise LevD but only among those LPVs with the same number of active products, that is we compare product mixes among destinations where a firm features the same product scope.[23] The bottom-right panel of Figure 5 reports the box-plot of the conditional pairwise LevD. In 2007 the inter-quartile range for Italy is large, spanning 0.27-0.78, with a mean approximately equal to 0.52. As it should be apparent form the Figure, similar numbers would emerge for France and for other years in our data. These evidence implies that two LPVs, belonging to the same firm and with the same number of active products, differ on average in more than 50% of their composition.[24] Summing up the evidence documented so far, LPVs appear very sparse and they tend to contain combinations of products with a significant degree of heterogeneity across destination.

A regression analysis will now correlate the observed dissimilarity with determinants identified in the literature on multi-product firms: firm-product efficiency and quality, as well as country characteristics (market size and competition) drive product selection within firm (see Bernard et al., 2011b; Mayer et al., 2014; Manova and Yu, 2017, among others). Here we apply a similar logic to the LevD computed with respect to the GPV and in line with previous

[20]This concern is particularly important here since it is well known that the empirical distributions of these variables are highly skewed.

[21]In Appendix C we report a number of further robustness checks. We first ask whether firm size matters. We construct quartiles of the firm-size distribution at the sectoral level (ISIC 3digit when possible, 2digit otherwise) in terms of number of employees and look at the LevD for the first and fourth quartiles. Second, we consider highly-diversified firms (those exporting over 15 products) versus little-diversified firms (those exporting fewer than 6).

[22]Note that a comparison with the statistical benchmark, reported in Table 2 for 2000 and 2007, suggests that the observed degree of diversity is significantly higher when we compute it using randomly generated LPV.

[23]Note that the logic behind computing the LevD with respect to the GPV or pairwise among LPVs is different. In the second case in fact LevD embeds factors belonging to different destination markets. This is also why in our regression analysis we will focus only on the LevD from the GPV.

[24]Figure D1 in Appendix D reports the same plot computed using random data generated by the statistical benchmark. It confirms that our results are not a mere statistical artifact.

empirical studies we estimate

$$\text{LevD}_{fd,t} = \alpha + \beta_1 \text{Market size}_{fd,t} + \beta_2 \text{Market concentration}_{fd,t} + \beta_3 \text{X}_{fd,t} + \delta_{ft} + \delta_d + \epsilon_{fd,t} \quad ,$$

where $\text{LevD}_{fd,t}$ is the dissimilarity index between the GPV and the LPV in destination d. Our interest lies in the variability of the dependent variables detected within a firm across destinations hence we include in our regression firm-year fixed effect δ_{ft} to account for any systematic difference in ability between exporters that might affect trade outcomes across markets. We also add destination fixed effects, δ_d, which implicitly account for cross-country differences in total income, market toughness, trade costs and other institutional frictions common to all exporters. With this structure of fixed effects coefficients are identified only through the variability within a firm-year across countries and they capture a conditional correlation between the dissimilarity index and a set of firm-destination level variables.

We include in our specification two controls capturing firm-specific characteristics of the destination markets: size and concentration. We measure the market size for firm f in destination d (Market size_{fd}) as the simple average, over the products in GPV_f, of the total (log) import to d from any country in the world, except Italy and France respectively.[25] Higher values of market size in destination d reflect a greater demand potential for the whole set of products exported by firm f. Similarly, market concentration (Market $\text{concentration}_{fd}$) is measured as the simple average, over the products in GPV_f, of the (log) Herfindal-Hirschman index for import to d from any country in the world, again except Italy and France.[26] A higher value of market concentration implies that in destination d, for the set of all products exported by firm f, on average imports are highly concentrated among few importing countries.

Finally, our specification accounts for possible efficiency or quality sorting effects. Indeed, again following a consolidated literature (Manova and Yu, 2017; Manova and Zhang, 2012), we add a further control $\text{X}_{fd,t}$ that is either the (log) total bilateral export sales of a firm f in destination d at time t ($\text{Sales}_{fd,t}$) or the (log) average "inferred quality" of the products exported in that destination ($\text{AvgQuality}_{fd,t}$).[27] In all regressions standard errors are clustered at the firm-year and destination level but our results are robust to alternative treatments of the error term.

Columns (1)-(2) in Table 3 show the estimation results for both Italy and France. Signs of

[25] This market size proxy is built following a two-step procedure. First we compute, using data from BACI (see Gaulier and Zignago, 2010), the total imports of product p in destination d, $\text{IMP}^*_{pd} = \sum_{o \in O_{pd}} \text{IMP}^*_{pod}$, where O_{pd} is the set of countries exporting product p to destination d, excluding in turn Italy or France. Measures built from BACI data are identified by a $*$ superscript. We then calculate the simple average of (log) IMP^*_{pd} for each firm-destination pair as

$$\text{Market size}_{fd} = \frac{1}{|\Pi_f|} \sum_{p \in \Pi_f} \log \text{IMP}^*_{pd} \quad ,$$

where $|\Pi_f|$ is the cardinality of the set of products exported by firm f.

[26] To compute our proxy of market concentration we first compute the Herfindal index of imports of product p in destination d, $\text{HHI}^*_{pd} = \sum_{o \in O_{pd}} (\text{IMP}^*_{pdo} / \sum_{o \in O_{pd}} \text{IMP}^*_{pdo})^2$. We then take the average over products for each firm-destination pair:

$$\text{Market concentration}_{fd} = \frac{1}{|\Pi_f|} \sum_{p \in \Pi_f} \log \text{HHI}^*_{pd} \quad .$$

[27] We measure export quality at the product-destination-firm level by applying the methodology implemented by Manova and Yu (2017) (originally developed by Khandelwal (2010)) where quality plays the role of a demand shifter. Accordingly, quality is obtained as the residual of a regression of $\ln q_{fpd,t} + \sigma \ln p_{fpd,t}$ on $\alpha_p + \alpha_{d,t} + \epsilon_{fpd,t}$, where elasticities of substitution σ are sector (3-digit) specific and taken from Imbs and Mejean (2017). $\text{AvgQuality}_{fd,t}$ is then computed as a simple average for each firm-destination pairs. One important caveat to keep in mind when using this measure is that it can accurately approximate quality assuming a theoretical framework with CES preferences and constant mark-ups.

Table 3: SELECTION OF PRODUCTS ACROSS DESTINATIONS

Dependent variable:	$\text{LevD}_{fd,t}$ (1)	(2)	(3)	(4)
		ITALY		
Market size$_{fd,t}$	-0.008^a	-0.024^a	-0.007^a	-0.009^a
	(0.001)	(0.002)	(0.000)	(0.001)
Market concentration$_{fd,t}$	0.008	0.019^a	0.008^a	0.010^a
	(0.006)	(0.006)	(0.001)	(0.001)
Sales$_{fd,t}$	-0.036^a		-0.029^a	
	(0.002)		(0.000)	
AvgQuality$_{fd,t}$		0.001^a		0.002^a
		(0.000)		(0.000)
N	5,059,814	4,993,475	4,531,125	4,474,050
adj. R^2	0.359	0.241	0.593	0.565
		FRANCE		
Market size$_{fd,t}$	-0.014^a	-0.031^a	-0.004^a	-0.006^a
	(0.002)	(0.003)	(0.000)	(0.000)
Market concentration$_{fd,t}$	0.010	0.019^b	0.001	0.002^c
	(0.008)	(0.008)	(0.001)	(0.001)
Sales$_{fd,t}$	-0.035^a		-0.027^a	
	(0.002)		(0.000)	
AvgQuality$_{fd,t}$		0.001^a		0.002^a
		(0.000)		(0.000)
N	2,144,583	2,056,169	1,935,821	1,856,872
adj. R^2	0.365	0.266	0.621	0.606
Firm-Year FE	Yes	Yes	No	No
Destination FE	Yes	Yes	No	No
Firm-Destination FE	No	No	Yes	Yes
Year FE	No	No	Yes	Yes

Notes: This table shows the regressions for the Levenshtein dissimilarity measure ($\text{LevD}_{fd,t}$) and the index computed conditional on the same number of products exported ($\text{CLevD}_{fd,t}$) between a firm's global product vector (GPV) and local product vectors (LPV). Columns showing the same specification with different fixed effects do not report the same number of observations due to the removal of singletons. Results show the standard errors (in parenthesis) clustered at the firm-year and destination level in columns 1-2, 5-6 and at the firm-destination and year level in columns 3-4, 7-8: $^c < 0.1$, $^b < 0.05$, $^a < 0.01$.

all coefficients that are statistically significant appear consistent with other results reported in the existing literature on multi-product firms. First, the dissimilarity between LPVs and the corresponding GPV tends to be lower in larger markets: a firm exports more products of its portfolio to destinations where there is a greater demand for these products, implying a lower number of zeros and less sparsity. Second, LevD increases in markets characterized by higher import concentration: a firm tends to export a small subset of its worldwide product portfolio in destination where imports are highly concentrated among few countries. Third, in markets where a firm earns systematically higher revenues there is a lower dissimilarity with respect to the GPV, while there is a positive association between the LevD and the average quality of the products exported to that destination. These latter results are qualitatively in line with Manova and Yu (2017).

While statistically significant the strength of these correlations is not high enough to affect the degree of sparsity in a firm's LPVs and the corresponding $\text{LevD}_{fd,t}$. The typical Italian or French exporter generates a LPV with (on average) only half of a product more in a destination where it observes an increase of one standard deviation (around 120%) of bilateral revenues and with (on average) only one third of a product more in a destination where the average

quality of its products is twice as big.[28] This confirms that penetrating a destination market with a new product is costly (Arkolakis et al., 2016) and corroborates the idea that the large presence of zeros is a robust feature of LPVs.

A potential caveat in interpreting our regression results concerns the fact that firm-destination determinants of product-vectors variability might be poorly controlled for in our specification and an omitted variable bias might be in place. To check to what extent this might affect our estimates in columns (3)-(4) we propose an alternative specification that further exploits the time-dimension in our data by including firm-destination and year fixed effects.[29] While other firm-destination factors seem to play a relevant role increasing the explanatory power of the model, all in all our results appear rather robust to the inclusion of more demanding fixed effects.

In Appendix E we provide a set of exercises that further test the robustness of our results with respect to a number of confounding factors. First, as suggested by a recent empirical literature on multi-product exporters, a firm's export product mix in different markets may be affected by tariffs. We thus add an additional regressors, $\text{Tariff}_{fd,t}$ to our specification. Second, we estimate our regression only on those products that are exported to EU countries, where there are no tariffs for French and Italian exporters. Third, to check if our empirical results do not simply reflect a product-mix heterogeneity caused by destination-country income, we re-calculate our dependent variables and re-run the regressions focusing only on a set of developed countries. Fourth, we add a sensitivity check, regarding "carry-along trade", in which manufacturing firms export products that they do not produce (Bernard et al., 2012). Lacking data on domestic production at the firm-product level, we control for this potential confounding effect by excluding products that are contemporaneously exported and imported by the same firm. Last, as emphasized by Baldwin and Ottaviano (2001) multi-product multinational (MNC) firms have an incentive to produce and export different products in different countries. Global MNC companies are indeed complex organizations that sprawl across industries and countries and their activity is an amalgamation of several very distinct types that must be explained by several very distinct models. Given their complexity and peculiarity trade patterns of these firms may be different from those of other companies, requiring therefore an in-depth investigation. Due to the lack of detailed information on FDI vs trade activities, as a robustness check we simple remove MNC from the regressions. The latter exercise is done only for Italy.[30]

Overall, the results of these robustness checks are in line with those reported in Table 3, excluding the possibility that our results reflect particular issues not accounted for in the baseline specification.[31]

Hierarchy of products across destinations

The investigations conducted so far have neglected the role of the intensive margin and, in particular, they can say nothing about the possibility, suggested in the theoretical literature (Mayer et al., 2014; Eckel et al., 2015; Manova and Yu, 2017), that firms adhere to the same

[28]These figures are computed using averages across years. In Italy the standard deviation of firm level bilateral export sales is about 1.2 while the average number of products exported is 12.5. The half product change is the obtained by (12.5-1)*(-0.036*1.2) where the -1 in the first term comes from the normalization of the LevD (cfr. Appendix B).

[29]Standard errors are now clustered at firm-destination and year but again the results are robust to alternative treatments of the error term.

[30]Our dataset for France did not include the LIFI file reporting exporters' ownership.

[31]In additional robustness checks, available upon request, we ask whether our results change if we modify the way the regressors market size and market concentration are defined. We calculate the two measures using only the top product in each destination for each firm. Second, we apply weights to a firm's products given by the total exports of all Italian firms of product p to destination d over the total exports of all Italian firms to destination d, excluding the exports of the firm under consideration. We also check that our findings are not a statistical artifact produced by the structure of the product classification by considering those firms that export over more than one HS2-digit product.

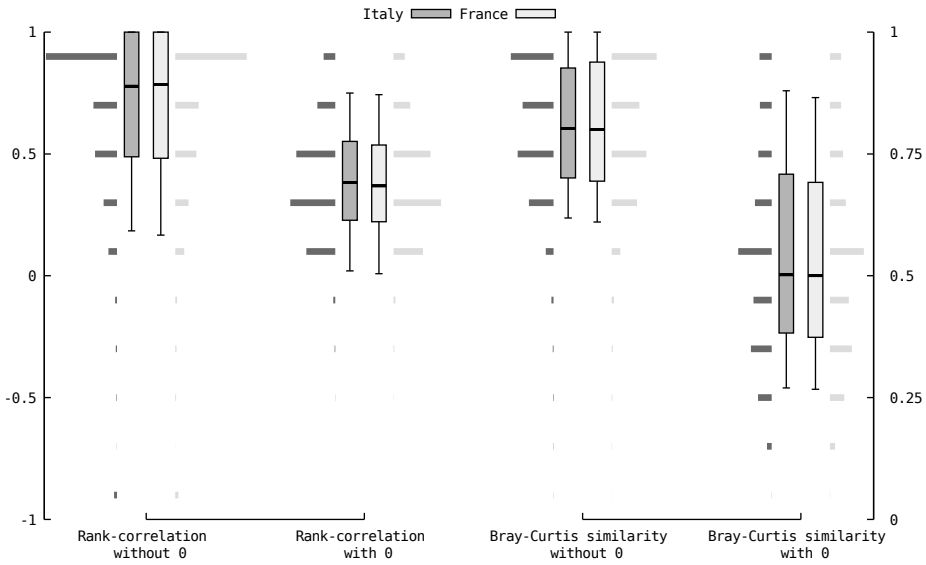

Figure 6: Firm-level box-plot of rank correlation index and of the Bray-Curtis similarity index between a firm's global product vector (GPV) and local product vectors (LPV) with and without the zeros. Products in the GPV and LPV are associated with their export shares. For each firm, each index is computed between its GPV and each LPV and then averaged across its destinations. Box-plot represents inter-quartile range with the corresponding median. Whiskers bars cover from the 10th to the 90th percentile of the distribution. To guide the reading of the figure the corresponding histogram is also reported. Note the for the Bray-Curtis index the y-scale is on the right. Year is 2007.

global product hierarchy in all destinations they are active in. We now move in this direction and explicitly take into account the role of product export revenues in studying the relationship between LPVs and the GPV.

Traditionally to compare product mixes across destinations, in terms of market shares, rank correlations have been used providing evidence of a rather high similarity among product vectors. This approach suffers of two limitations. First, by disregarding a firm's choice of not exporting a product in a given destination (i.e. by removing the zeros from LPVs), it mechanically inflates (in absolute terms) the correlation index. This happens, on the one hand, because we do not assign a rank to zeros and, on the other, because all the LPVs with a single product are by construction discarded from the analysis. Second, while rank-based statistics are designed to be less sensitive in presence of significant heterogeneity among observations[32], it is certainly also true that they imply a loss of information. Neglecting these two aspects affects the actual measurement of the diversity among LPVs.

To assess to what extent this is true. Figure 6 (left-panel) reports box-plots of the Spearman rank correlation between LPVs and GPV both neglecting and accounting for the role of zeros in the computation.[33] Two comments are in order. First, in line with previous results, for both Italy and France the median and mean correlation computed without the zeros are quite high, 0.76 and 0.65 respectively, suggesting an almost stable hierarchy across destinations. However, if we include the zeros in the computation, both mean and median appear significantly lower being around 0.4 in France and Italy as well. Second, looking at the entire distribution we observe that for a non-negligible group of firms product hierarchies in different destinations are significantly diverse. This is the case irrespective of considering or not the zeros.[34]

[32]As discussed above this is certainly relevant for international trade studies where the variables under investigations are characterized by highly skewed distributions.

[33]Figure 6 is built using data for 2007. Other years returns similar results (available upon request).

[34]When we do not consider the zeros the inter-quartile ranges are $(0.49, 1)$ and $(0.48, 1)$ in Italy and France respectively, while the 80% of the firms have rank-correlation in $(0.18, 1)$ and $(0.17, -1)$. On the contrary with the zeros inter-quartile ranges read $(0.23, 0.55)$ and $(0.22, 0.54)$ for Italy and France, and 80% of the firms have

Table 4: HIERARCHY OF PRODUCTS ACROSS DESTINATIONS

Dependent variable:	$BC_{fd,t}$				$RC_{fd,t}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ITALY							
Market size$_{fd,t}$	0.008^a	0.032^a	0.013^a	0.016^a	0.005^a	0.028^a	0.016^a	0.019^a
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Market concentration$_{fd,t}$	-0.021^a	-0.037^a	-0.012^a	-0.015^a	-0.018^a	-0.034^a	-0.015^a	-0.018^a
	(0.004)	(0.005)	(0.002)	(0.002)	(0.004)	(0.005)	(0.002)	(0.002)
Sales$_{fd,t}$	0.055^a		0.054^a		0.053^a		0.052^a	
	(0.002)		(0.000)		(0.003)		(0.001)	
AvgQuality$_{fd,t}$		0.003^a		0.002^a		0.004^a		0.003^a
		(0.000)		(0.000)		(0.000)		(0.000)
N	5,059,814	4,993,475	4,531,125	4,474,050	5,059,570	4,993,239	4,531,004	4,473,929
adj. R^2	0.558	0.477	0.668	0.634	0.364	0.295	0.499	0.467
	FRANCE							
Market size$_{fd,t}$	0.006^a	0.030^a	0.008^a	0.009^a	0.005^a	0.029^a	0.010^a	0.012^a
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Market concentration$_{fd,t}$	-0.030^a	-0.043^a	-0.008^a	-0.009^a	-0.026^a	-0.039^a	-0.010^a	-0.012^a
	(0.005)	(0.006)	(0.002)	(0.002)	(0.004)	(0.005)	(0.002)	(0.002)
Sales$_{fd,t}$	0.053^a		0.047^a		0.052^a		0.049^a	
	(0.002)		(0.000)		(0.003)		(0.000)	
AvgQuality$_{fd,t}$		0.005^a		0.003^a		0.006^a		0.005^a
		(0.000)		(0.000)		(0.001)		(0.000)
N	2,144,583	2,056,169	1,935,821	1,856,872	2,144,480	2,056,070	1,935,759	1,856,812
adj. R^2	0.556	0.481	0.698	0.674	0.365	0.308	0.524	0.502
Firm-Year FE	Yes	Yes	No	No	Yes	Yes	No	No
Destination FE	Yes	Yes	No	No	Yes	Yes	No	No
Firm-Destination FE	No	No	Yes	Yes	No	No	Yes	Yes
Year FE	No	No	Yes	Yes	No	No	Yes	Yes

Notes: This table shows the regressions for the Bray-Curtis distance (BC_{fd}) and the Rank Correlation (RC_{fd}) between a firm's global product vector (GPV) and local product vectors (LPV). Columns showing the same specification with different fixed effects do not report the same number of observations due to the removal of singleton. Results show the standard errors (in parenthesis) clustered at the firm-year and destination level in columns 1-2, 5-6 and at the firm-destination and year level in columns 3-4, 7-8: $^c < 0.1$, $^b < 0.05$, $^a < 0.01$.

Overall this evidence suggests that the idea that firms adhere to the same global hierarchy in every destination served seems much less supported by data. A firm's LPVs display indeed a high degree of diversity.

To further corroborate this conclusion we move forward and we explicitly associate to each product both in the GPV and in LPVs a firm's worldwide and bilateral export share respectively.[35] Irrespective of including or not the zeros, comparing product vectors with export shares using rank-based statistics implies a loss of information and might influence our assessment of

rank-correlations in $(0, 0.75)$.

[35]In practice this consists in replacing the 1s in the GPV and LPV with the corresponding product market share. GPV and LPVs are not binary vectors anymore, they become vectors whose components sum up to 1.

the overall diversity of a firm's LPVs.[36] To evaluate the what extent this is true we complement the information obtained using a standard rank-correlation index with that associated with a similarity measure that properly accounts for product market shares. This measure, known as Bray-Curtis similarity (BC from now on), is defined as

$$BC_{fd} = 1 - \frac{\sum_p |s_{fp} - s_{fdp}|}{\sum_p |s_{fp} + s_{fdp}|} \quad ,$$

where s_{fp} and s_{fdp} represent the export shares of the product p in the GPV and LPV respectively.[37] $\sum_p |s_{fp} + s_{fdp}|$ is a normalizing factor, which always equals two in our case. The BC similarity index is bound between 0 and 1, where 1 means the two vectors have the same composition both in terms of products and in terms of their market shares ($s_{fp} = s_{fdp}$ for all p) and 0 means the two vectors are completely disjoint (no products in common).[38] Figure 6 reports (last plot on the right) the box-plot of the BC similarity index computed including the zeros in LPVs. Three remarks are in order. First, the mean is about 0.55 both in France and in Italy and the corresponding median is only marginally lower and not distant from the median point of the theoretical support $[0, 1]$. Second the dispersion of the distribution is rather large with the 10%-90% range covering almost two thirds of the entire support. Third, not considering the sparsity of LPV significantly upward biases also this similarity measure.[39] This analysis documents that, once properly accounted for product selection, we tend to observe a high degree of heterogeneity among a firm's LPV also in terms of the relative importance of each product with respect to the GPV.

As before, we conclude our investigations by asking how much of the detected variability can be ascribed to factors that are destination-specific or firm-destination specific by estimating

$$\{\mathrm{BC}_{fd,t}, \mathrm{RC}_{fd,t}\} = \alpha + \beta_1 \text{Market size}_{fd,t} + \beta_2 \text{Market concentration}_{fd,t} + \beta_3 \mathrm{X}_{fd,t} + \delta_{ft} + \delta_d + \epsilon_{fd,t} \quad ,$$

where the dependent variable is either the Bray-Curtis similarity or the rank-correlation (RC) both computed including the zeros. Results are reported in Table 4. In commenting these results we focus on France and on the BC index only (columns (1)-(2)) since it turns out that estimates are qualitatively similar between Italy and France and between the two measures. The average similarity between LPVs and the GPV tends to be higher where firms face bigger demand and lower import concentration. The same is true in those destination markets where exporters earn higher revenues and sell higher quality goods. However, as before, the impact of these determinants is not enough economically meaningful to contradict our main stylized

[36] Consider the following hypothetical firm:

$$\mathrm{GPV}_f = \begin{pmatrix} 0.97 \\ 0.02 \\ 0.01 \end{pmatrix} \quad \mathrm{LPV}_{fd_1} = \begin{pmatrix} 0.51 \\ 0.49 \\ 0 \end{pmatrix} \quad \mathrm{LPV}_{fd_2} = \begin{pmatrix} 0.85 \\ 0.15 \\ 0 \end{pmatrix} \quad .$$

In this example, although LPV_{fd_1} and LPV_{fd_2} are very different, the rank correlation between each LPV and the GPV is one. Similarly if the hypothetical firms is

$$\mathrm{GPV}_f = \begin{pmatrix} 0.51 \\ 0.48 \\ 0.01 \end{pmatrix} \quad \mathrm{LPV}_{fd_1} = \begin{pmatrix} 0.51 \\ 0.49 \\ 0 \end{pmatrix} \quad \mathrm{LPV}_{fd_2} = \begin{pmatrix} 0.49 \\ 0.51 \\ 0 \end{pmatrix} \quad ,$$

rank-correlations would signal diversities where there are almost none.

[37] See Appendix B for more details regarding the definition of this measure.

[38] In the example discussed in footnote 36 the BC similarity for LPV_{fd_1} and LPV_{fd_2} would be 0.53 and 0.87 respectively for the first hypothetical firm and 0.99 and 0.97 for the second one.

[39] In Appendix C we report a number of robustness checks. We construct quartiles of the firm-size distribution at the sectoral (ISIC 3digit when possible, 2digit otherwise) in terms of number of employees and look at the LevD for the first and fourth quartiles. Second, we consider highly-diversified firms (those exporting over 15 products) versus little-diversified firms (those exporting fewer than 6). They are all consistent with our story.

Table 5: STRUCTURE OF TYPICAL PRODUCT VECTORS.

Year	Number of multi-product firms with TPV (%)	Share of destinations with TPV	Share of exports covered by products in TPV	Share of firms with rank of the product of the TPV >1
		ITALY		
2000	65%	44%	68%	26%
random	[42%]	[27%]	[25%]	[77%]
2007	69%	46%	68%	26%
random	[46%]	[26%]	[25%]	[78%]
		FRANCE		
2000	66%	46%	67%	26%
random	[40%]	[30%]	[27%]	[75%]
2007	67%	47%	68%	26%
random	[43%]	[29%]	[27%]	[75%]

Notes: The statistics are calculated averaging over firms in the restricted sample where we keep only those firms exporting more than one product and serving more than one destination. Statistics for the statistical benchmark (random LPV) are computed averaging over 50 different replications. Statistics for the statistical benchmark (random) are computed for 2000 and 2007 averaging over 50 replications.

facts. A one standard deviation increases in bilateral revenues or average quality is associated with only a negligible variation in the similarity index. As before we test the robustness of these results against a more demanding specification with firm-destination plus year fixed-effects (columns (3)-(4) and (7)-(8)). As above we also run a battery of test to investigate if tariffs, focusing on EU or Developed country destinations, controlling for "carry-along trade" and removing multi-product multinational companies drive our results. Results are reported in Appendix E: no substantial deviations from our baseline are observed.

The picture emerged so far suggests that, albeit important, quality and efficiency firm-product attributes can not entirely explain multi-product firms' behavior in international markets. Expansion into foreign markets requires decisions about which countries to approach and which products to export and such decisions are the result of a complex combination of factors. This complexity can be detected also when looking at firms' core products: beyond the simple pecking order reflecting quality or efficiency the strategic core of a firm could hide product complementarities driven by technical, organizational, or management reasons. Defining "core products" is indeed more complex than simply rank them in terms of export sales. In the next section, we investigate deeply the issue by complementing and improving upon the existing evidence on firms' core products and competencies.

4. Similarities in product vectors

Inspired again by our illustrative example in this section we introduce the notion of typical product vector (TPV), as the most frequent combination of products observed across destinations. We characterize its composition and we produce evidence that the TPV composition is not governed only by the level of products' quality or efficiency. By exploiting as a natural experiment the exogenous shock driven by the removal of quotas on Chinese exports in textile and clothing industries after the entry of China in the WTO, we show that firms' switching decisions for products belonging to the TPV are less likely to be driven by quality or efficiency reasons. We conclude the section by providing some tentative evidence on the possible technological complementarities guiding the TPV.

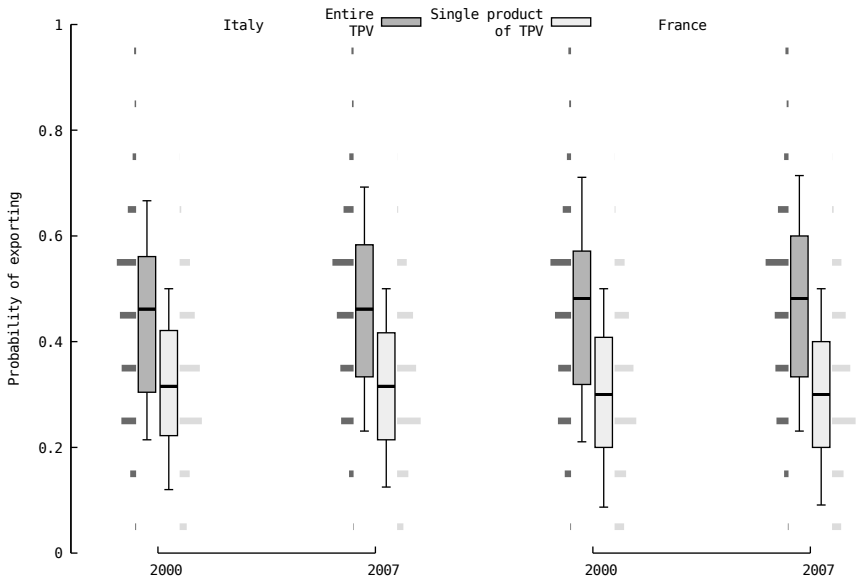

Figure 7: Top panels report for Italy (left) and France (right) empirical distribution of the probability of observing a destination with a typical product vector (TPV) and of observing a destination with single product of the TPV, in both cases possibly together with other products not included in the TPV.

Typical product vector

An exporter's TPV has been defined above as the most frequent (across destinations) combination of products exported, independently of their market shares.[40]

Table 5 presents the descriptive statistics on the relevance of the TPV in Italy and France. First, the TPV is pervasive: two thirds of exporters do possess a TPV. Second, the TPV is economically relevant: on average it is exported in 45% of a firm's destinations and the products it contains account for 67% of a firm's total exports. Finally, the TPV is not a statistical artifact: when we use randomly generated LPVs the percentage of firms with a TPV is significantly lower and, while the TPV tends to cover a similar share of destinations, the products in the TPV are less important in terms of sales than in the case of real data.

We next compare the distribution of the probability of observing a destination with the TPV, possibly with other products, together with the distribution of the probability of observing a destination where one single product of the TPV is exported, possibly together with other products not in the TPV. Visual inspection of Figure 7 clearly suggests that the probability of exporting the TPV tends to be higher than the probability of exporting a single product of the TPV alone, hence suggesting that these products tend to be co-exported, which is confirmed by a robust non-parametric test of stochastic dominance with a significance level lower than 10^{-4}, again both for France and Italy.[41] This result is in line with the idea that the stable component of a firm's export portfolio, identified by the TPV, may capture some demand or technological complementarities across goods that are missed when one uses a single-product perspective.[42]

[40]In case of ties we select the combination associated with the more important destination in term of sales. We require the frequency to be higher than 1 so that it is possible that a firm does not possess the TPV. This happens when it exports a different product mix in each destination.

[41]The null hypothesis in this test is that the probability that a random draw from the first density displays a higher value as compared with a random draw from the second one is equal 1/2. A negative (positive) value of the test statistics implies that this probability is higher (lower) than 1/2. test statistics are -59.6 and -46.0 for Italy and France respectively. See Fligner and Policello (1981) for details.

[42]Again we check if this result is compatible with the statistical benchmark. Figure D2 in Appendix D reports a similar box-plot built using randomly generated LPV data for 2007. As opposed to Figure 7 in this case the probability of exporting a single product of the TPV alone is higher than the probability of exporting the TPV. This result, which holds for both Italy and France, suggests that also this regularity does not represent a simple statistical artifact.

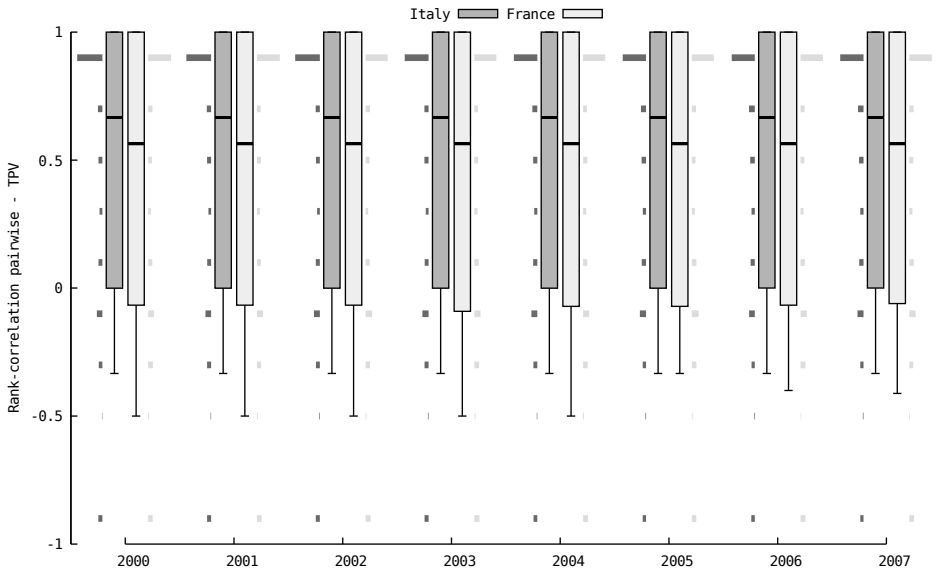

Figure 8: Firm-level frequency distributions of the average pairwise rank correlation between export shares in terms of sales for products of a firm's typical product vector in the different destinations where it is exported. For each firm, the pairwise rank correlation is computed between any two destinations where the TPV is exported, averaged within each destination first and then across destinations. Box-plot represents inter-quartile range with the corresponding median. Whiskers bars cover from the 10th to the 90th percentile of the distribution. To guide the reading of the figure the corresponding histogram is also reported.

In what follows, we further characterize a firm's TPV by looking at its composition in terms of export sales. According to the standard definition a firm's core product should be the most important one in terms of export sales, either due to efficiency or quality reasons. In order to verify whether the efficiency or quality sorting models apply to the TPV, we exploit the information concerning the ranking of the TPV products in terms of export sales. We observe in the last column of Table 5 that, for more than one quarter of firms with a TPV, the TPV is not simply composed by the set of the first k products ranked in terms of export value.[43]

The lack of a perfect ordering among products belonging to the TPV is detected also when looking at the ranking across destinations. We rank the products belonging to a firm's TPV in terms export sales and for each destination where the TPV is exported we compute the pairwise rank correlation. We then take the average within each firm's destination first and the across its destinations. The firm-level frequency distribution is reported in Figure 8 for Italian and French firms, respectively. The average rank correlation is around 0.4 and almost 1/3 of the firms presents a negative rank correlation among TPV: even within the stable part of a firm, there is a great deal of destination-specific heterogeneity in terms of export shares of the products in the TPV. As shown in Figure C3 in Appendix C this result is robust when we split our sample by firm size or by the overall number of products exported.[44]

As a final exercise, to further corroborate the hypothesis that the TPV is not only governed by quality and efficiency mechanisms, we consider the variation in export prices and quality across products by means of regression analysis. By drawing on Manova and Yu (2017), we

[43]The relative rank of the TPV is computed as

$$\frac{\sum_{p\in \mathrm{TPV}} \mathrm{rank}_p}{\sum_{i=1}^{|\mathrm{TPV}|} \mathrm{i}},$$

where |TPV| represent the cardinality of the TPV.

[44]In this case using only the top destination is in this case unreasonable as this would imply compute the pairwise rank correlation between the TPV and itself, in case the top destination has the TPV, or between TPV and a vector without the TPV, in case the top destination does not have the TPV.

Table 6: TYPICAL PRODUCT VECTORS - THE ROLE OF QUALITY AND EFFICIENCY

Dependent variable:	$\text{Price}_{fp,t}$	$\text{Quality}_{fp,t}$	$\text{Price}_{fp,t}$	$\text{Quality}_{fp,t}$
	(1)	(2)	(3)	(4)
	ITALY		FRANCE	
$\text{Sales}_{fp,t}$	0.005^a	1.075^a	0.037^a	1.181^a
	(0.001)	(0.007)	(0.002)	(0.009)
$\times$ $\text{DTPV}_{fp,t}$	-0.013^a	-0.087^a	-0.025^a	-0.110^a
	(0.001)	(0.008)	(0.002)	(0.011)
$\text{DTPV}_{fp,t}$	-0.025	0.127	-0.005	0.148
	(0.016)	(0.093)	(0.028)	(0.132)
N	3,383,322	3,325,388	1,513,875	1,468,725
adj. R^2	0.707	0.437	0.685	0.426
Firm-Year FE	Yes	Yes	Yes	Yes
Product-Year FE	Yes	Yes	Yes	Yes

Notes: Regression using data on 2000-2007. All the regressions include a constant term. Robust standard errors in parenthesis, clustered at firm level: $^c < 0.1$, $^b < 0.05$, $^a < 0.01$.

estimate the following model:

$$y_{fp,t} = \alpha + \beta_1 \text{Sales}_{fp,t} + \beta_2 \text{DTPV}_{fp,t} + \beta_3 \text{DTPV}_{fp,t} \times \text{Sales}_{fp,t} + \delta_{f,t} + \delta_{p,t} + \epsilon_{fp,t} \quad (1)$$

where y_{fpt} is either the (log) price charged by a firm f in product p at time t, $(\text{Price}_{fp,t})$[45], or the (log) inferred export quality $(\text{Quality}_{fp,t})$ using the Manova and Yu (2017) approach.[46] $\text{DTPV}_{fp,t}$ is a dummy variable for products belonging to a firm's TPV. We regress either price or quality on a firm's sales in product p and on its interaction with the DTPV. We include firm-time fixed effect (δ_{ft}) to account of all time variant and time-invariant exporters' characteristics that might affect trade outcomes symmetrically across products and product-year fixed effect δ_{pt} which allow us to make prices and quality comparable across goods by demeaning them by their product-year specific average across firms.

In line with previous findings in Table 6 we observe that the conditional correlation between export price/quality and sales across goods within a firm is positive (Manova and Yu, 2017). However, we observe that the quality-sales relationship is weakened for products that belong to the TPV: export revenues are positively correlated with quality across products but less so for the TPV products.

All together these evidences, the fact that the typical product vector does not necessary include a firm's most important products in term of sales and the lack of a perfect correlation between TPV across destinations, suggest a substantial departure from the concept of core products traditionally employed in the literature and the idea of a stable ranking of product exports across destinations. Overall our results reinforce the idea of product completementarity or technological relatedness which imply co-exportation of goods. We will further investigate these potential rich forms of inter-dependencies between products in the following.

Typical product vector and trade shocks

According to a stylized conceptual framework where multi-product firms compete on production efficiency or product quality, firms' product mix adjustments in general and in response

[45]The price is approximated by the trade unit value and constructed by taking the ratio $\frac{\sum_d \text{sales}_{fpdt}}{\sum_d \text{quantity}_{fpdt}}$.

[46]Export quality is now estimated at product-firm as the residual of a regression of $\ln q_{fp,t} + \sigma p_{p,t}$ on $\alpha_{p,t} + \epsilon_{fp,t}$, where elasticities of substitution σ are sector (3-digit) specific and taken from Imbs and Mejean (2017).

Table 7: PRODUCT DYNAMICS and TYPICAL PRODUCT VECTOR

Dependent variable:	$\text{Drop}_{fp,t=1}$					
	(1)	(2)	(3)	(4)	(5)	(6)
	ITALY					
$\text{DTPV}_{fp,t=0}$	-0.033^a	-0.084^a	-0.045^a	-0.098^a	-0.067^a	-0.035^a
	(0.002)	(0.010)	(0.002)	(0.027)	(0.014)	(0.002)
$\text{Sales}_{fp,t=0}$		-0.027^a		-0.028^a	-0.027^a	
		(0.000)		(0.001)	(0.001)	
$\times$ $\text{DTPV}_{fp,t=0}$		0.006^a		0.006^a	0.005^a	
		(0.001)		(0.002)	(0.001)	
Export $\text{Share}_{fp,t=0}$			-0.272^a			
			(0.007)			
$\times$ $\text{DTPV}_{fp,t=0}$			0.139^a			
			(0.006)			
$\text{Quality}_{fp,t=0}$						-0.003^a
						(0.000)
$\times$ $\text{DTPV}_{fp,t=0}$						0.002^a
						(0.000)
$\text{NPE}_{f,t=0}$	-0.074^a	-0.062^a	-0.078^a	-0.052^a	-0.062^a	-0.073^a
	(0.002)	(0.002)	(0.002)	(0.007)	(0.003)	(0.002)
N	2,845,140	2,845,140	2,845,140	601,383	2,243,757	2,791,136
adj. R^2	0.468	0.473	0.469	0.520	0.478	0.469
	FRANCE					
$\text{DTPV}_{fp,t=0}$	-0.042^a	-0.094^a	-0.055^a	-0.141^a	-0.092^a	-0.042^a
	(0.003)	(0.015)	(0.004)	(0.037)	(0.020)	(0.003)
$\text{Sales}_{fp,t=0}$		-0.030^a		-0.034^a	-0.030^a	
		(0.001)		(0.002)	(0.001)	
$\times$ $\text{DTPV}_{fp,t=0}$		0.006^a		0.010^a	0.006^a	
		(0.001)		(0.003)	(0.001)	
Export $\text{Share}_{fp,t=0}$			-0.320^a			
			(0.010)			
$\times$ $\text{DTPV}_{fp,t=0}$			0.165^a			
			(0.010)			
$\text{Quality}_{fp,t=0}$						-0.004^a
						(0.000)
$\times$ $\text{DTPV}_{fp,t=0}$						0.001^a
						(0.000)
$\text{NPE}_{f,t=0}$	-0.070^a	-0.055^a	-0.075^a	-0.056^a	-0.053^a	-0.068^a
	(0.004)	(0.004)	(0.004)	(0.011)	(0.005)	(0.004)
N	1,308,970	1,308,970	1,308,970	248,779	1,060,191	1,238,615
adj. R^2	0.488	0.494	0.489	0.552	0.499	0.489
Firm-Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Regression using data on 2000-2007, for firms exporting at least one product in both $t-1$ and t (Surviving firms). Column 3 the variable ln $\text{Sales}_{fp,t=0}$ (and its interaction with the TPV) is replaced with a firm-product export share. Column 4 includes only those firms with a TPV with more than one product while column 5 include firms with a TPV made by one product. All the regressions include a constant term. Robust standard errors in parenthesis, clustered at firm level: $^c < 0.1$, $^b < 0.05$, $^a < 0.01$.

to an exogenous shocks should follow a precise ordering that depends on the product ranking within a firm (Mayer et al., 2014; Manova and Yu, 2017). However, under the hypothesis that core competences reflect richer forms of interdependence that can be captured by those goods that are more likely to be co-exported, products belonging to the TPV should be more resilient to shock, irrespective of their quality or efficiency level.

To test this hypothesis we draw on Manova and Yu (2017) relating a firm's probability of dropping a variety from its export portfolio with product attributes, such as revenues and quality, and we study whether these attributes are relevant for products belonging to the TPV. With this aim we estimate the following model

$$\begin{aligned}\text{Drop}_{fp,t=1} &= \alpha + \beta_1 \text{DTPV}_{fp,t=0} + \beta_2 \text{Product Attribute}_{fp,t=0} + \\ &+ \beta_3 \text{DTPV}_{fp,t=0} \times \text{Product Attribute}_{fp,t=0} + \beta_4 \ln \text{NPE}_{f,t=0} + \delta_t + \delta_{fp} + \epsilon_{fp,t} \quad .\end{aligned} \tag{2}$$

The dependent variable, $\text{Drop}_{fp,t=1}$, is a binary indicator set equal to 1 if a firm exporting the product p at time $t=0$ drops it at time $t=1$. $\text{DTPV}_{fp,t=0}$ is a dummy variable for products belonging to a firm's TPV while Product $\text{attribute}_{fp,t=0}$ refers either to the (log) export revenues or the (log) quality of a firm's products. Our baseline specification also includes a control for a

firm's product scope, namely the (log) number of products exported by the firm in year $t = 0$ ($\text{NPE}_{f,t=0}$). Our interest lies in the value of the coefficients β_1 and β_3 which capture the extent to which products belonging to the TPV differ with respect to other goods in terms of their probability of being dropped. Since we include firm-product fixed effects, to capture all the time invariant unobserved determinants of export dynamics which may be correlated with the product attribute, these coefficients are identified from the variation in the product dynamics within a firm-product over time. To further clean our identification we also include year fixed effects, δ_t, to control for possible time-trends. Standard errors are clustered by firm to allow for correlations within firms over time, but our results are robust to alternative treatments of the error term. Note that these regressions only considers firms which do not drop all their products, that is only surviving firms. This avoids to confound factors related to the likely different motivation behind the choice to completely exit from export markets. Results are therefore informative on dropping probabilities conditional on survival.[47]

Table 7 presents the results which are very similar across countries allowing to comment them only for Italy. We start with Column (1) where we include only the the dummy capturing if a product belongs to the TPV. In this baseline specification the estimated coefficient is -0.033. Given an average drop rate of about 45% among products that are not in the TPV, this means that goods in the TPV have a 7% lower probability of being dropped than other products.[48] This result corroborates the idea that the TVP represents a stable ensemble of products even across time. More interestingly, in column (2) we observe that exporters on average tend to drop more their relatively marginal products in terms of sales but this negative relationship is weaker among products that belong to the TPV. In this specification a 100% increase in a firm-product export revenues decreases the probability of dropping by 6.5% and 4.6% for non-TPV and TPV products, respectively. That is products in the TPV are 30% less sensitive to changes in revenues, pointing to some forms of inter-dependence across products in the TPV that go beyond single product revenue-based considerations. As a robustness check in column (3) we replicate the analysis replacing revenues with corresponding export shares, a measure of the relative importance of a product within the firm, while in columns (4) and (5) we split the sample to focus on firms having a multiple-products TPV or a single-product TPV, respectively. Results remain robust across different specifications and samples. Finally, in column (6) we move our attention to quality estimated as above at firm-product level. Result suggests a deviation from a strict product ordering also in terms of quality for goods belonging to the TPV. Indeed, the higher is the quality the lower is the probability of dropping a product but less so for goods belonging to the TPV.

While in the previous exercises we have been agnostic about the reasons why a firm should drop its products and eventually reallocate its resources among the surviving ones, we now consider a specific episode that has induced substantial adjustments along firms extensive margin: the dismantling of the Multi-fibre Arrangement quotas on Chinese textile products in conjunction with China's accession to the World Trade Organization (WTO). Following Martin and Mejean (2014), Utar (2014), and Bloom et al. (2016), we use the observed change in the market share of Chinese worldwide exports by product as a proxy for the exogenous increase in competitive pressures faced by Italian and French exporters. As well documented by Brambilla et al. (2010), after the quotas were removed China's worldwide exports in the textile and clothing sector experienced substantial growth.[49]

In order to identify the effect of such an increase in competition, we add to equation (2) the change in China's export share of product p between t and $t+1$ (Δ China Share_{pt}) and

[47]Because it is not feasible to estimate an adding regression at the firm-product level as the set of possible added products includes all the products not currently exported by the firm, we keep the focus on firm-product dropping behavior. See also Bernard et al. (2011a) for a similar approach.

[48]On average the fraction of exported firm-products that is dropped every year is 41% among all firm-product.

[49]This growth led some European Countries to reintroduce some limited quotas after 2005.

its interaction with the TPV dummy.[50] Following Bloom et al. (2016), we restrict the sample to those firms belonging to sectors with toughest quotas in 2000 which were likely to be more exposed to the increase in China competition after its accession to the WTO and the dismantling of the MFA quotas.[51] The estimates of this regression are reported in Panel A of Table 8. Because some temporary quotas have been reintroduced by the European Union after 2005, we restrict the analysis on the period 2000-2004. In columns (1)-(2) and (4)-(5) we run the equation on all firms while in columns 4 and 6 we remove those sectors without binding quotas in 2000. In line with the economic intuition and with existing evidence, firms are more likely to drop products that are strongly exposed to competition from Chinese firms. However, for those products that are in TPV the effect is reversed suggesting that the tougher is the Chinese competition the more likely is that firms consolidate around the ensemble of its core products. This result is robust even when controlling for the relative importance of the products in terms of sales (columns (2) and (3), and (5) and(6)).

As an alternative identification strategy, we draw on Martin and Mejean (2014) and we propose a difference-in-difference framework that exploits the removal of quotas in 2002 as an exogenous trade shock from low-wage countries. We use the information on the quotas in 2000 made available by Bloom et al. (2016).[52] We estimate the following equation

$$\begin{aligned}\text{Drop}_{fp,t=1} = \alpha + \beta_1\text{Post2002} + \beta_2\text{Post2002} \times \text{Quota}_p + \beta_3\text{Quota}_p \times \text{DTPV}_{fp,t} + \\ + \beta_4\text{DTPV}_{fp,t} + \beta_5\text{Post2002} \times \text{Quota}_p \times \text{DTPV}_{fp,t} + \delta_t + \delta_{fp} + \epsilon_{fp,t}\ ,\end{aligned} \tag{3}$$

where Post2002 is a dummy equal to one after 2001 and Quota_p is a dummy equal to one for the treated group, i.e. those products belonging to sectors with a large share of goods subjected to a quota on Chinese exports before 2002. With this specification we compare a firm's probability of dropping a product before and after the entry of China in the WTO and use as a treated group those products belonging to sectors with a large proportion of good categories covered by a binding quota before 2002. Hence, we expect the interaction term Post2002 $\times$ Quota_p to be positive: the treated group experienced a surge in the intensity of competition from China that should drive firms' export dynamic. However, to corroborate our argument this effect should be less pronounced among products belonging to the TPV. In this case the coefficient associated with the triple interaction Post2002 $\times$ Quota_p $\times$ $\text{DTPV}_{fp,t}$ should be negative.

Panel B of Table 8 reports the results. As before the analysis is restricted over the period 2000-2004. Columns (1) and (3), for Italy and France respectively, show the result of a baseline specification with the effect of the removal of quotas for the treated group without distinguishing between products that are or not in the TPV. The positive sign on the coefficient β_2 means that a firm's probability of dropping a product increases more strongly in the sectors where quotas were relaxed, i.e. sectors that have faced a more intense competition from China. In columns (2) and (4) we add the triple interactions. The coefficient β_5 is negative indicating that the effect of the competitive pressure coming from Chinese exporters is less pronounced for the treated products that belong to the TPV. Indeed, column (2) for Italy suggests that after the entry of China in the WTO firms show higher probability of dropping a product than before 2002 (18.7 percentage points) and this probability increases of other 1.6 percentage

[50] China's export share is computed using data from BACI as the worldwide exports of China in product p at time t divided by the total exports of all the other countries excluding Italy or France to avoid the introduction of a source of endogeneity in the analysis.

[51] In a similar vein, Manova and Yu (2017) replicate the analysis of product switching on the restricted sample of Chinese firms belonging to the textile and apparel industries that enjoyed a large increase in foreign demand after the removal of MFA quotas.

[52] This data come from the SIGL (System for the Management of Licenses for Textile Imports) database, available at *http://trade.ec.europa.eu/sigl/*, is classified according to 172 grouped quota categories defined by the EU. Bloom et al. (2016) map these categories into the HS6 classification and, for each four-digit industry, they calculate the proportion of product categories that are covered by a quota in 2000 weighting each HS6 in an industry by its 2000 import value.

Table 8: PRODUCT DYNAMICS and TYPICAL PRODUCT VECTORS - TRADE SHOCK

PANEL A - Dependent variable:			Drop$_{fp,t=1}$			
	(1)	(2)	(3)	(4)	(5)	(6)
		ITALY			FRANCE	
DTPV$_{fp,t=0}$	-0.020^a	-0.066^a	-0.125^a	-0.030^a	-0.068^a	-0.097
	(0.002)	(0.013)	(0.036)	(0.003)	(0.019)	(0.060)
Δ China Share$_{fp,t=0}$	0.069^a	0.065^a	0.056^c	0.058	0.057	-0.012
	(0.024)	(0.024)	(0.031)	(0.039)	(0.039)	(0.057)
$\times$ DTPV$_{fp,t=0}$	-0.362^a	-0.357^a	-0.238^a	-0.406^a	-0.395^a	-0.323^a
	(0.041)	(0.040)	(0.065)	(0.066)	(0.065)	(0.122)
Sales$_{fp,t=0}$		-0.025^a	-0.023^a		-0.028^a	-0.029^a
		(0.001)	(0.001)		(0.001)	(0.002)
$\times$ DTPV$_{fp,t=0}$		0.005^a	0.010^a		0.005^a	0.007
		(0.001)	(0.003)		(0.001)	(0.005)
NPE$_{f,t=0}$	-0.070^a	-0.059^a	-0.076^a	-0.063^a	-0.049^a	-0.078^a
	(0.003)	(0.003)	(0.009)	(0.005)	(0.005)	(0.016)
N	2,008,955	2,008,955	441,648	944,569	944,569	164,587
adj. R^2	0.498	0.502	0.481	0.513	0.518	0.484
PANEL B						
		ITALY			FRANCE	
Post2002$_t$	0.200^a	0.187^a		0.200^a	0.192^a	
	(0.003)	(0.003)		(0.004)	(0.005)	
Post2002$_t$ $\times$ Quota$_p$	0.011^a	0.016^a		0.026^a	0.038^a	
	(0.004)	(0.006)		(0.006)	(0.009)	
DTPV$_{fp,t}$		-0.021^a			-0.029^a	
		(0.002)			(0.003)	
Post2002$_t$$\times$Quota$_p$$\times$DTPV$_{fp,t}$		-0.077^a			-0.093^a	
		(0.007)			(0.012)	
Quota$_p$ $\times$ DTPV$_{fp,t}$		0.039^a			0.040^a	
		(0.008)			(0.013)	
N	2,503,618	1,972,075		121,0745	936,123	
adj. R^2	0.470	0.496		0.492	0.512	
Firm-Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Regression using data on 2000-2004, for firms exporting at least one product in both $t-1$ and t (Surviving firms). In Panel A we add as independent variable the change in China's export share of product p between t and $t+1$ (Δ China Share$_{pt}$) and its interaction with $\times$ DTPV$_{fp,t=0}$. In columns 3 and 6 of panel A the sample includes only those firms belonging to sectors with toughest quotas in 2000. In Panel B we propose a difference-in-difference approach that exploits the removal of quotas in 2002 as an exogenous trade shock. All the regressions include a constant term. Robust standard errors in parenthesis, clustered at firm level: $^c < 0.1$, $^b < 0.05$, $^a < 0.01$.

points for the treated products, i.e those goods on which a binding quota was imposed before 2002. However, the likelihood of being dropped after 2002 is reduced by 7.7 percentage points for the treated products belonging to the TPV. Similar findings are observed for France.

Although in the analysis so far we have agnostically studied the TPV composition, without taking a stance on the possible type of interdependence across products, in the next and final exercise we try to disentangle possible technological relatedness. We do that by exploiting the Broad Economic Categories (BEC) classification to identify in our dataset those products that belongs to the "intermediate" class.[53] These category includes raw materials that are more likely to be used in a successive productive step, as well as parts and components. These products, although marginal from a quantitative point of view, may be part of a firm's export bundle as they are indeed linked to the final good by technical reasons.[54]

[53]The BEC classification has been widely used in the literature of international trade to identify intermediate inputs (Amiti et al., 2014; Brandt et al., 2017).

[54]This is indeed what we observe in our illustrative example where the TPV is composed by two products, defined as final according to the BEC classification, and one good labeled parts and components that belong to

Table 9: TECHNOLOGICAL COMPLEMENTARITIES in the TYPICAL PRODUCT VECTOR

Dependent variable:	$\text{Drop}_{fp,t=1}$			
	(1)	(2)	(3)	(4)
	ITALY		FRANCE	
$\text{DTPV}_{fp,t=0}$	-0.430^a	-0.415^a	-0.435^a	-0.412^a
	(0.001)	(0.002)	(0.002)	(0.003)
Intermediate_p	0.013^a	0.014^a	0.001	0.005^b
	(0.002)	(0.002)	(0.002)	(0.002)
$\times\ \text{DTPV}_{fp,t=0}$	-0.042^a	-0.045^a	-0.034^a	-0.040^a
	(0.002)	(0.002)	(0.003)	(0.003)
$\text{Quality}_{fp,t=0}$		-0.006^a		-0.009^a
		(0.000)		(0.000)
$\times\ \text{DTPV}_{fp,t=0}$		0.002^a		0.003^a
		(0.000)		(0.000)
$\text{NPE}_{f,t=0}$	0.126^a	0.124^a	0.124^a	0.119^a
	(0.002)	(0.002)	(0.003)	(0.003)
N	2,845,140	2,791,136	130,8970	123,8615
adj. R^2	0.124	0.128	0.133	0.145
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: Regression using data on 2000-2007, for firms exporting at least one product in both $t-1$ and t (Surviving firms). The variable Intermediate_p is a dummy equals 1 if the product is classified as an intermediate inputs in the BEC category and zero otherwise. All the regressions include a constant term. Robust standard errors in parenthesis, clustered at firm level: $^c < 0.1$, $^b < 0.05$, $^a < 0.01$.

As before, we estimate the model on the probability of dropping a product adding the information on a product characteristic as follows

$$\begin{aligned}\text{Drop}_{fp,t=1} &= \alpha + \beta_1 \text{DTPV}_{fp,t=0} + \beta_2 \text{Intermediate}_p + \\ &+ \beta_3 \text{DTPV}_{fp,t=0} \times \text{Intermediate}_p + \beta_4 \ln \text{NPE}_{f,t=0} + \delta_t + \delta_f + \epsilon_{fpt}\end{aligned} \tag{4}$$

where the variable Intermediate_p is a dummy equals 1 if the product is classified as an intermediate inputs in the BEC category and zero otherwise. Because the product characteristic is now time invariant we include firm rather than firm-product fixed effect (δ_f). If technological complementarities are at work, as for instance suggested by our illustrative example, a TPV should include intermediate products that might be relatively marginal for the firm in terms of quality or efficiency but yet representing its core competences and less likely to be dropped when adjustments over time occur. Results, presented in of Table 9, indeed confirm that firms are more likely to drop intermediate products but less so when they belong to the TPV. This is consistent with firms adjusting their product portfolio taking into account other firms of goods' interdependence that deviate from the rank of quality or profitability.

5. Conclusion

In this paper we proposed an original empirical strategy to investigate the structure of export decisions of individual multi-product firms across destinations. Several departures from the usual empirical approach employed to study multi-product firms helped refining our understanding of how the product mix is adjusted. First, considering the whole list of products exported, we explicitly introduced a firm' choice of not exporting some of its products in a

the intermediate class.

given destination by including the zeros in her local product vector. Second, instead of only considering the contribution of each product to a firm's product in terms of export sales, we prioritized the identity of the goods exported by a firm across destinations. So doing we shed light on the different set of combinations exported by a firm across destinations. Finally, we detected combinations of products more likely to be co-exported irrespectively of their importance in terms of sales. On the measurement front we also used novel metrics overcoming some of the limitations of the Spearman correlation index typically used in this literature.

Considering two comprehensive panels of the universe of Italian and French multi-product exporters we confirmed the importance of the three dimensions of exporters choices, meaning sparsity, fickleness and stability. The observed regularities were shown not to be statistical artifacts.

Beyond the descriptive evidence on the structure of firms' local product vectors and global ones, we showed that selection of products at destination is severe: the product mix at destination contains, on average, at least 80% of zeros with respect to the worldwide portfolio of products exported by a firm, which opens the possibility of having a high number of different combinations of products in the different destinations served by this exporter. We found that 1/3 of French and Italian firms export a different product-mix in each destination. Also, two local product vectors belonging to the same firm and with the same number of active products will differ in more than 50% of their composition. Finally, our approach pointed to rich and understudied forms of complementarities or technological relatedness between exported products.

Beyond statistical evidence, or identified correlation between these patterns and the usual determinants of the composition of the product mix, we used an exogenous competitive shock (the removal of MFA quotas on textiles and apparel) faced by French and Italian firms to show how our findings amend the results based on a cost- or quality-driven pecking order of products. Products belonging to the typical product vector are less likely to be dropped, as well as products complementing each other in technological terms. Also considering the quality of exported products, we show that the richer forms of complementarity between products here identified cushion the expected adjustment of exporters, whereby tougher competition leads exporters to concentrate on core products of higher quality.

Accordingly, our findings do not simply refine our understanding of the strategy of multi-product exporters. They also suggest extending multi-product export models to capture richer form of product complementarities, as well as product selection as key determinants of trade patterns. Another and related important avenue for future research is to understand how sparsity, fickleness and stability in the exported product mixes affect the welfare and distributional consequences of international trade.

Acknowledgments

The present work was made possible thanks to a research agreement between the Italian Statistical Office (ISTAT) and the Scuola Superiore Sant'Anna. Lionel Fontagné acknowledges support by the Bank of France at earlier stages of this paper. Angelo Secchi gratefully acknowledges the Paris School of Economics for granting him a *Résidence de Recherche*. We also acknowledge financial support from the Institute for New Economic Thinking, INET inaugural grant #220. We thank Lorenzo Caliendo, Lionel Nesta, Thierry Mayer and Stephanie Haller for their useful comments. We thank seminar participants at the RIDGE December forum (Montevideo), University of Nice-Sophia Antipolis (Nice), S.Anna School of Advance Studies (Pisa) and the FREIT conference (Ljubljana) for insightful comments.

References

Amiti, M., O. Itskhoki, and J. Konings (2014): "Importers, Exporters, and Exchange Rate Disconnect," *American Economic Review*, 104, 1942–78.

Arkolakis, C., S. Ganapati, and M.-A. Muendler (2016): "The Extensive Margin of Exporting Products: A Firm-level Analysis," Cowles Foundation Discussion Papers 2028, Cowles Foundation for Research in Economics, Yale University.

Arkolakis, C. and M.-A. Muendler (2013): "Exporters and their products: a collection of empirical regularities," *CESifo Economic Studies*, 59, 223–248.

Armenter, R. and M. Koren (2014): "A Balls-and-Bins Model of Trade," *American Economic Review*, 104, 2127–2151.

Bailey, E. E. and A. F. Friedlaender (1982): "Market structure and multiproduct industries," *Journal of economic literature*, 1024–1048.

Baldwin, C. Y. and C. J. Woodard (2009): "The Architecture of Platforms: A Unified View," in *Platforms, Markets and Innovation*, ed. by A. Gawer, Edward Elgar.

Baldwin, R. and J. Harrigan (2011): "Zeros, Quality, and Space: Trade Theory and Trade Evidence," *American Economic Journal: Microeconomics*, 3, 60–88.

Baldwin, R. E. and G. I. Ottaviano (2001): "Multiproduct multinationals and reciprocal FDI dumping," *Journal of International Economics*, 54, 429–448.

Bernard, A. B., E. J. Blanchard, I. V. Beveren, and H. Y. Vandenbussche (2012): "Carry-Along Trade," NBER Working Papers 18246, National Bureau of Economic Research, Inc.

Bernard, A. B., M. Grazzi, and C. Tomasi (2011a): "Intermediaries in International Trade: Direct versus indirect modes of export," NBER Working Papers 17711, National Bureau of Economic Research, Inc.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2009): "The Margins of US Trade," *American Economic Review*, 99, 487–93.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010): "Multiple-Product Firms and Product Switching," *American Economic Review*, 100, 70–97.

——— (2011b): "Multi-Product Firms and Trade Liberalization," *Quarterly Journal of Economics*, forthcoming.

Bloom, N., M. Draca, and J. V. Reenen (2016): "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity," *Review of Economic Studies*, 83, 87–117.

Brambilla, I., A. K. Khandelwal, and P. K. Schott (2010): "China's Experience under the Multi-Fiber Arrangement (MFA) and the Agreement on Textiles and Clothing (ATC)," in *China's Growing Role in World Trade*, National Bureau of Economic Research, Inc, NBER Chapters, 345–387.

Brandt, L., J. V. Biesebroeck, L. Wang, and Y. Zhang (2017): "WTO Accession and Performance of Chinese Manufacturing Firms," *American Economic Review*, 107, 2784–2820.

Bray, J. R. and J. T. Curtis (1957): "An ordination of upland forest communities of southern Wisconsin," *Ecological Monographs*, 27, 325–349.

Chaney, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98, 1707–21.

Chatterjee, A., R. Dix-Carneiro, and J. Vichyanond (2013): "Multi-product Firms and Exchange Rate Fluctuations," *American Economic Journal: Economic Policy*, 5, 77–110.

Comite, F. D., J.-F. Thisse, and H. Vandenbussche (2014): "Verti-zontal differentiation in export markets," *Journal of International Economics*, 93, 50 – 66.

Crino', R. and P. Epifani (2012): "Productivity, Quality and Export Behaviour," *Economic Journal*, 122, 1206–1243.

Crozet, M., K. Head, and T. Mayer (2011): "Quality sorting and trade: Firm-level evidence for French wine," *Review of Economic Studies*.

Dhingra, S. (2013): "Trading Away Wide Brands for Cheap Brands," *American Economic Review*, 103, 2554–84.

Eaton, J., S. Kortum, and F. Kramarz (2011): "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, 79, 1453–1498.

Eckel, C., L. Iacovone, B. Javorcik, and J. P. Neary (2015): "Multi-product firms at home and away: Cost- versus quality-based competence," *Journal of International Economics*, 95, 216–232.

Eckel, C. and J. P. Neary (2010): "Multi-Product Firms and Flexible Manufacturing in the Global Economy," *Review of Economic Studies*, 77, 188–217.

Feenstra, R. and H. Ma (2008): "Optimal Choice of Product Scope for Multiproduct Firms under Monopolistic Competition," in *The Organization of Firms in a Global Economy*, ed. by E. Helpman, D. Marin, and T. Verdier, Harvard University Press.

Flach, L. and E. Janeba (2017): "Income inequality and export prices across countries," *Canadian Journal of Economics/Revue canadienne d'conomique*, 50, 162–200.

Fligner, M. A. and G. E. Policello (1981): "Robust rank procedures for the Behrens-Fisher problem," *Journal of the American Statistical Association*, 76, 141–206.

Forlani, E., R. Martin, G. Mion, and M. Muls (2016): "Unraveling firms: Demand, productivity and markups heterogeneity," Working Paper Research 293, National Bank of Belgium.

Freund, C. and M. D. Pierola (2015): "Export superstars," *Review of Economics and Statistics*, 97, 1023–1032.

Gaulier, G. and S. Zignago (2010): "BACI: International Trade Database at the Product-Level. The 1994-2007 Version," Working Papers 2010-23, CEPII research center.

Hallak, J. C. and P. K. Schott (2011): "Estimating Cross-Country Differences in Product Quality," *The Quarterly Journal of Economics*, 126, 417–474.

Helpman, E., M. Melitz, and Y. Rubinstein (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 123, 441–487.

Iacovone, L. and B. Javorcik (2010): "Multi-Product Exporters: Product Churning, Uncertainty and Export Discoveries," *Economic Journal*, 120, 481–499.

Imbs, J. and I. Mejean (2017): "Trade Elasticities," *Review of International Economics*, 25, 383–402.

Khandelwal, A. (2010): "The Long and Short (of) Quality Ladders," *The Review of Economic Studies*, 77, 1450–1476.

Kugler, M. and E. Verhoogen (2012): "Prices, Plant Size and Product Quality," *Review of Economic Studies*, 79, 307–339.

Manova, K. and Z. Yu (2017): "Multi-product firms and product quality," *Journal of International Economics*, 109, 116–137.

Manova, K. and Z. Zhang (2012): "Export Prices across Firms and Destinations," *Quarterly Journal of Economics*, 127, 379–436.

Martin, J. and I. Mejean (2014): "Low-wage country competition and the quality content of high-wage country exports," *Journal of International Economics*, 93, 140–152.

Mayer, T., M. J. Melitz, and G. I. Ottaviano (2011): "Market Size, Competition, and the Product Mix of Exporters," NBER Working Papers 16959, National Bureau of Economic Research, Inc.

——— (2014): "Market Size, Competition, and the Product Mix of Exporters," *American Economic Review*, 104, 495–536.

——— (2016): "Product Mix and Firm Productivity Responses to Trade Competition," NBER Working Papers 22433, National Bureau of Economic Research, Inc.

Melitz, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

Montgomery, C. A. (1994): "Corporate diverification," *Journal of Economic Perspectives*, 8, 163–178.

Nocke, V. and S. Yeaple (2014): "Globalization And Multiproduct Firms," *International Economic Review*, 55, 993–1018.

Osharin, A., J.-F. Thisse, P. Ushchev, and V. Verbus (2014): "Monopolistic competition and income dispersion," *Economics Letters*, 122, 348–352.

Parenti, M., P. Ushchev, and J.-F. Thisse (2017): "Toward a theory of monopolistic competition," *Journal of Economic Theory*, 167, 86–115.

Qiu, L. D. and W. Zhou (2013): "Multiproduct firms and scope adjustment in globalization," *Journal of International Economics*, 91, 142–153.

Stigler, G. J. (1963): "United States v. Loew's Inc.: A Note on Block-Booking," *The Supreme Court Review*, 1963, 152–157.

Sutton, J. and D. Trefler (2016): "Capabilities, Wealth, and Trade," *Journal of Political Economy*, 124, 826–878.

Utar, H. (2014): "When the Floodgates Open: Northern; Firms' Response to Removal of Trade Quotas on Chinese Goods," *American Economic Journal: Applied Economics*, 6, 226–250.

Verhoogen, E. (2008): "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector," *Quarterly Journal of Economics*, 123, 489–530.

Whinston, M. D. (1990): "Tying, Foreclosure, and Exclusion," *American Economic Review*, 80, 837–59.

Yeaple, S. R. (2013): "Scale, Scope, and the International Expansion Strategies of Multi-product Firms," Working Paper 19166, National Bureau of Economic Research.

Yu, C. and T. N. Wong (2015): "A product bundle determination model for multi-product supplier selection," *Journal of Intelligent Manufacturing*, 26, 369–385.

Appendix A

The above empirical analysis has produced new evidence regarding firm product vector across destinations. Concerns may be raised about the external validity of our results: can we extend to other countries these stylized facts we observe for France and Italy? Here below we report evidence in support of a positive answer to this question by referring at the existing empirical literature on multi-product and multi-destination exporters.

Indeed, Arkolakis and Muendler (2013); Arkolakis et al. (2016) report three robust stylized facts concerning the number of products exporters ship (export scope) and the corresponding average sales per product (export scale) to each destination, which characterize firms in Brazil, Denmark, Chile and Norway. First, a few large wide-scope exporters and many narrow-scope firms coexist in each destination. Second, the sales of wide-scope exporters are concentrated on a few products, and the same firms are able to cope with lower sales for their lowest-selling products. Third, there is a systematic positive relationship between average exporter scale (i.e the average sales per product within each country) and exporter scope. To further check that there is nothing particular about our data, we replicate their analysis and find the same stylized facts for both Italian and French exporters.

Figure A1 shows the relationship, for Italian firms (row 1) and French firms (row 2), between exporter scope and the corresponding percentile in the exporter-scope distribution for firms shipping to Germany and to the USA (similar patterns are found for other destinations). The distribution is very skewed: more than 70% of exporters export only three or fewer products, while only the top 10% ship more than 10 products. In Figure A1 we plot, for firms exporting 4, 8, 16, and 32 products, the average (across firms) export product value for products sharing the same rank against their rank within firm. These figures confirm first that in Italy (row 3) and France (row 4) wide-scope exporters are indeed much larger, in terms of sales, than narrow-scope exporters. But more importantly, wide-scope exporters are able to cope with lower sales of their lowest selling products. In this respect, regressing the exporter's lowest-ranked product's (log) sales against (log) exporter scope in a market conditioning on fixed effects for firm and destination produces an elasticity of -1.45 (0.007) and -1.56 (0.011) for Italy and France, respectively (Arkolakis et al. (2016) estimate an elasticity for Brazilian firms of 2.1). Finally, Figure A1 confirms for Italian (row 5) and French (row 6) firms that within-destination mean exporter scope and the corresponding mean exporter scale are positively related.

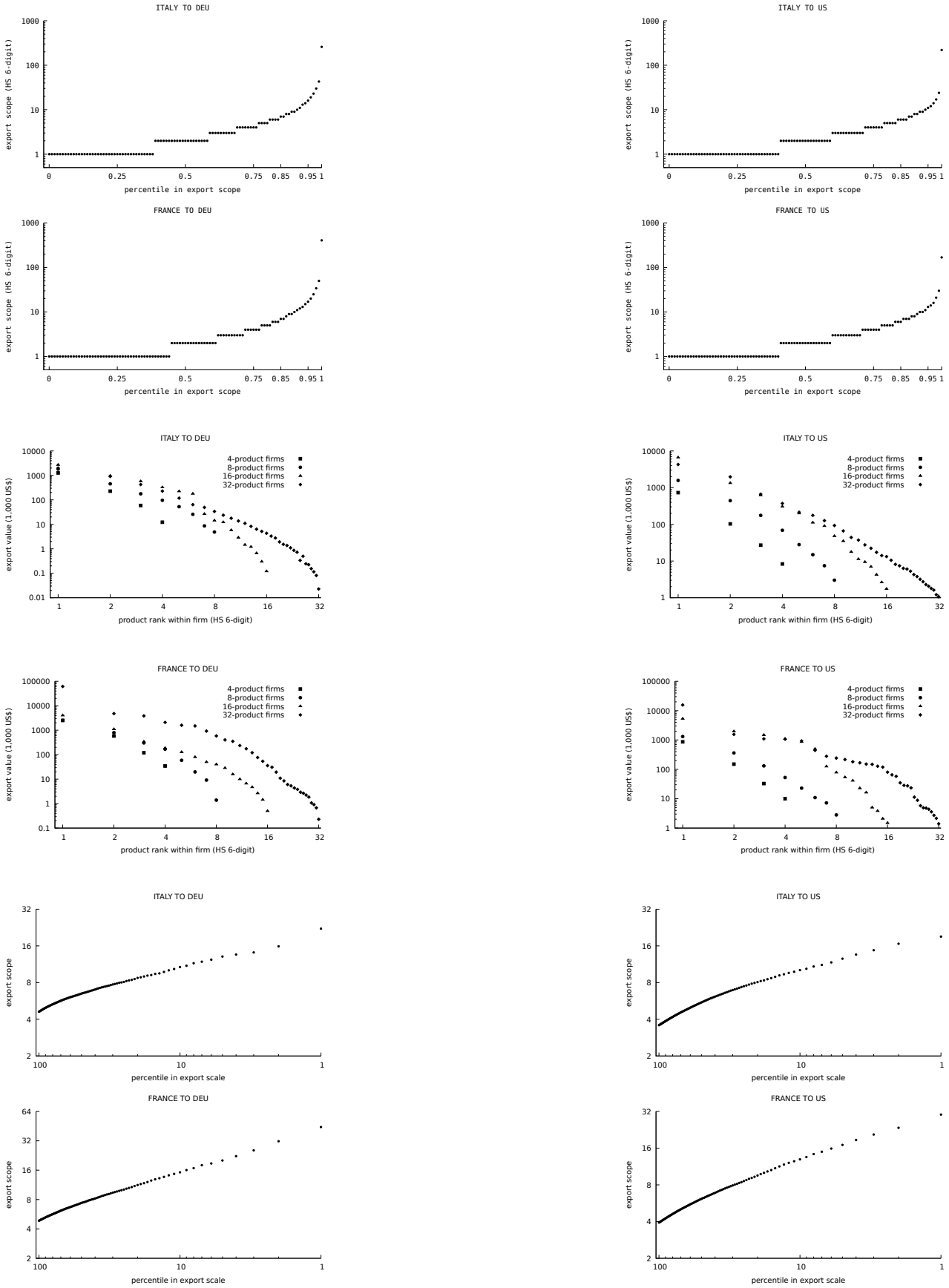

Figure A1: Empirical distributions of the number of product exported to Germany (DEU) and to the US by Italian (left and right panels, row 1) and French (left and right panels, row 2) exporters. Empirical distributions of product export sales to Germany (DEU) and to the US of Italian (left and right panels, row 3) and French (left and right panels, row 4) exporters. These latter figures shows the distributions only for firms exporting exactly 4, 8, 16 or 32 products. Relation between exporter scale and exporter scope for Italian (left and right panels, row 5) and French (left and right panels, row 6) firms exporting to Germany (DEU) and the US. Customs data 2007, products are defined at the HS 6-digit level. The scales of both the vertical and horizontal axes are logarithmic. Other years are available upon request.

Appendix B

Levenshtein distance

The Levenshtein distance between two sequences is the minimum number of single edits (insertion, deletion, substitution) required to change one sequence into the other, divided by the number of elements of the longest sequence. Mathematically the Levenshtein distance between two sequences $s = \{s_1, \ldots, s_{N_s}\}$ and $q = \{q_1, \ldots, s_{N_q}\}$ is

$$Lev(s,q) = \frac{\text{Edit}(N_s, N_q)}{\text{Max}(N_s, N_q) - 1} \quad ,$$

where $\text{Edit}(N_s, N_q)$ is given by the following recursion

$$\text{Edit}(i,j) = \begin{cases} \text{Max}(i,j) & \text{if Min}(i,j) = 0 \\ \text{Min} \begin{cases} \text{Edit}(i-1,j) + 1 \\ \text{Edit}(i,j-1) + 1 \\ \text{Edit}(i-1,j-1) + \mathbf{1}(s_i \neq q_j) \end{cases} & \text{otherwise} \end{cases}$$

in which $\mathbf{1}(\cdot)$ is the indicator function.

Consider the following simple example for a firm exporting four different HS6 products to two destinations. Each element of the vector takes the value one if the corresponding product is exported, and zero otherwise. We thus consider a firm's binary choice to export a product or not.

$$\text{GPV}_f = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{LPV}_{fd_1} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \quad \text{LPV}_{fd_2} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad .$$

In this example LevD(GPV_f,LPV_{fd_1}) is 1/3: one change is needed to transform LPV_{fd_1} into the GPV_f and the number of elements in the longest sequence is four. Instead, the LevD(GPV_f,LPV_{fd_2}) is equal to 2/3: two changes are required to transform LPV_{fd_2} into the GPV_f and the length of the longest sequence is still four. More generally, in our example introducing an additional difference between the two sequences (by changing a one in a LPV to a zero) affects the Levenshtein distance by 1/3.

Bray-Curtis similarity index

The Bray-Curtis similarity index, also known as the Sørensen index, is a well-known way of quantifying the similarity between samples.[55] Formally, the Bray-Curtis similarity index between vector i and vector j is defined as

$$BC_{i,j} = 1 - \frac{\sum_{k=1}^{K} |i_k - j_k|}{\sum_{k=1}^{K} |i_k + j_k|} \quad , \tag{5}$$

where i_k and j_k represent the number of elements observed in vector i and j along the kth dimension. The Bray-Curtis index is symmetric and ranges from 1, when the two vectors are

[55] Originally this measure was developed to study species abundance in different locations in ecological analysis.(Bray and Curtis, 1957)

identical, to 0 where the two vectors are disjoint.[56] The BC measure is not an Euclidean proximity index since it does not satisfy the triangular inequality axiom.

Usually, the BC index is computed on count data.[57] Consider the two following vectors i and j with dimension $K = 3$

$$i = \begin{pmatrix} 11 \\ 0 \\ 7 \end{pmatrix} \quad j = \begin{pmatrix} 24 \\ 37 \\ 5 \end{pmatrix} \quad ;$$

the $\mathrm{BC}_{i,j}$ between the two is 0.381, and is obtained as follows:

$$BC_{i,j} = 1 - \frac{|11-24| + |0-37| + |7-5|}{18+66} = 1 - 0.619 = 0.381 \quad . \tag{6}$$

When calculated with raw count data, the BC similarity index captures similarity associated with both the *size* and the *shape* of the two vectors, where the former refers to similarity in the total number of elements, $\sum_k i_k$ and $\sum_k j_k$, in vector i and j respectively while *shape* concerns the distribution of elements along the different dimensions of the two vectors.

We here calculate the BC similarity index between a firm's GPV and its LPVs where the elements of the vectors are product export shares. In this case, the denominator in equation (5) is always 2, and the index captures similarity only in *shape*. The BC index is particularly useful when we observe firms exporting products to different destinations with the same ranking but with very different export shares.

[56] Two vectors i and j are disjoint if whenever there is a non-zero entry in i, there is a zero entry in j and vice versa.

[57] For example, the number of firms in a location or the number of products exported to a destination.

Appendix C

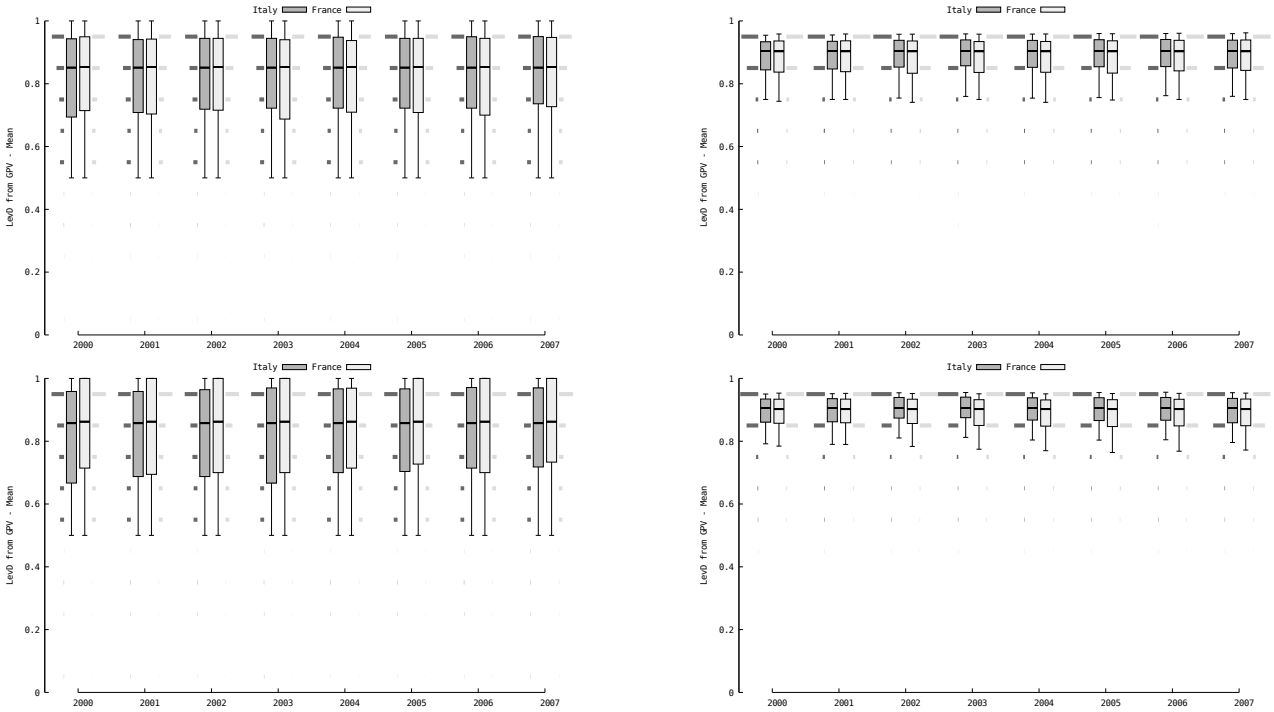

Figure C1: Firm-level frequency distributions of the average Levenshtein dissimilarity index (LevD) between a firm's global product vector (GPV) and local product vectors (LPV). For each firm, the LevD is computed between each LPV and the corresponding GPV and then averaged over its destinations. Box-plot represents inter-quartile range with the corresponding median. Whiskers bars cover from the 10th to the 90th percentile of the distribution. To guide the reading of the figure the corresponding histogram is also reported. Computations are performed with only firms in the first (**top-left panel**) and fourth (**top-right panel**) quartile of the sectoral (3-digit) firm size distribution and with only firms firms exporting fewer than 6 (**bottom-right panel**) and more than 14 products (**bottom-left panel**) products.

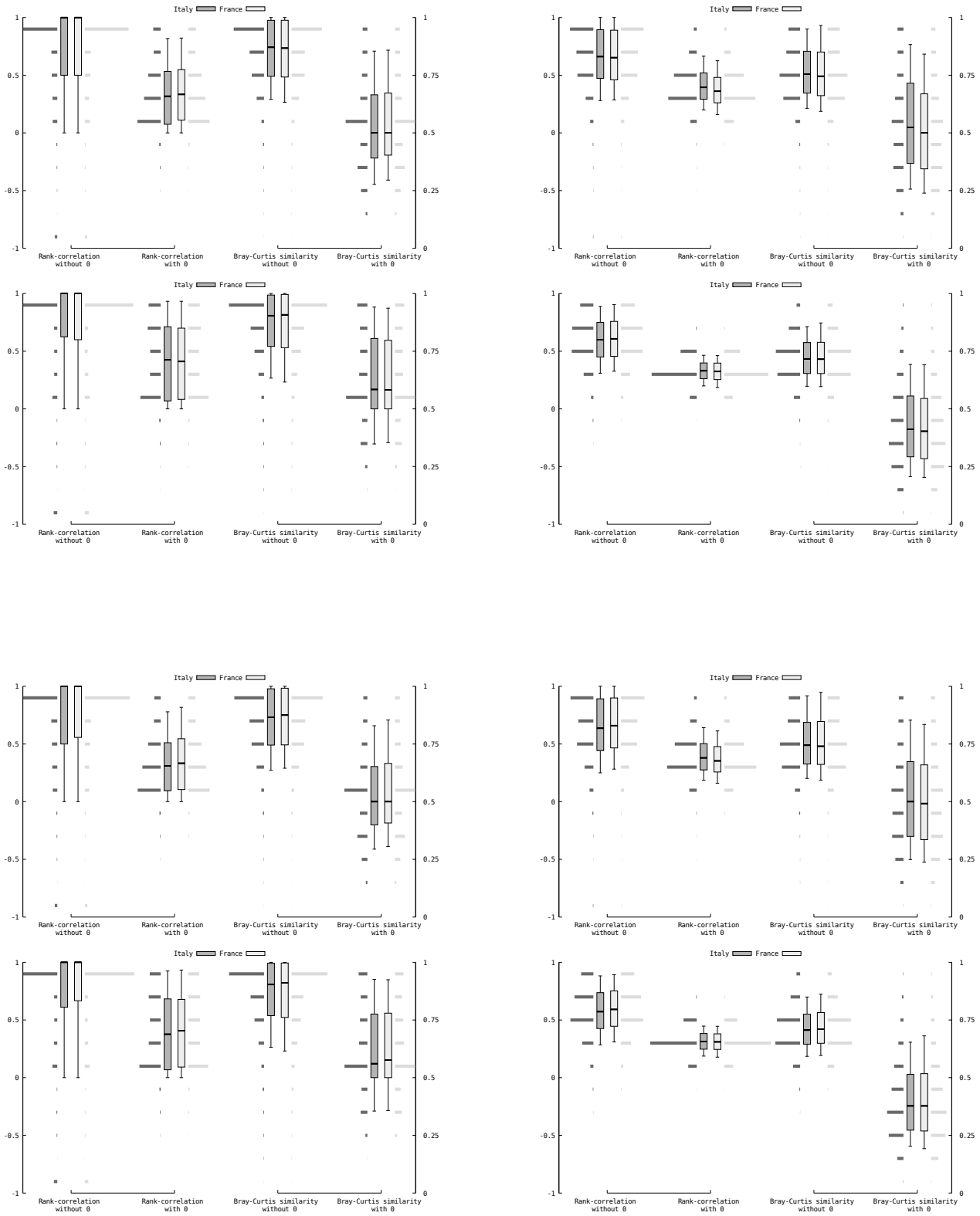

Figure C2: Firm-level box-plot of rank correlation index and of the Bray-Curtis similarity index between a firm's global product vector (GPV) and local product vectors (LPV) with and without the zeros. Products in the GPV and LPV are associated with their export shares. For each firm, each index is computed between its GPV and each LPV and then averaged across its destinations. Box-plot represents inter-quartile range with the corresponding median. Whiskers bars cover from the 10th to the 90th percentile of the distribution. To guide the reading of the figure the corresponding histogram is also reported. The first four panels refer to 2007 while the second four to 2000. Within each year computations are performed with only firms in the first (**top-left panel**) and fourth (**top-right panel**) quartile of the sectoral (3-digit) firm size distribution and with only firms firms exporting fewer than 6 (**bottom-right panel**) and more than 14 products (**bottom-left panel**) products. Note the for the Bray-Curtis index the y-scale is on the right.

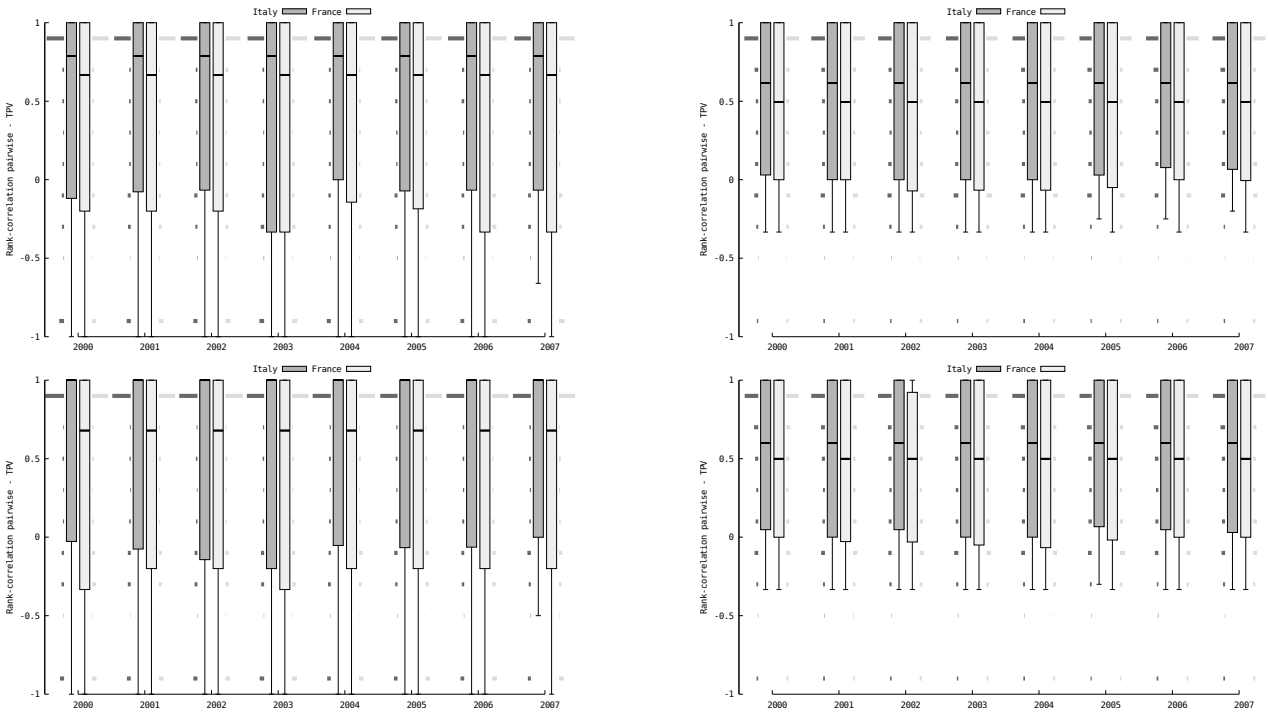

Figure C3: Firm-level frequency distributions of the average pairwise rank correlation between export shares in terms of sales for products of a firm's typical product vector in the different destinations where it is exported. For each firm, the pairwise rank correlation is computed between any two destinations where the TPV is exported, averaged within each destination first and then across destinations. Box-plot represents inter-quartile range with the corresponding median. Whiskers bars cover from the 10th to the 90th percentile of the distribution. To guide the reading of the figure the corresponding histogram is also reported. Computations are performed with only firms in the first (**top-left panel**) and fourth (**top-right panel**) quartile of the sectoral (3-digit) firm size distribution and with only firms firms exporting fewer than 6 (**bottom-right panel**) and more than 14 products (**bottom-left panel**) products.

Appendix D

Figure D1: Firm-level box-plots of the average pairwise Levenshtein distance among a firm's destination with the same product scope computed with random data generated by the statistical benchmark and averaged over 50 replications. Box-plot represents inter-quartile range with the corresponding median. Whiskers bars cover from the 10th to the 90th percentile of the distribution. To guide the reading of the figure the corresponding histogram is also reported.

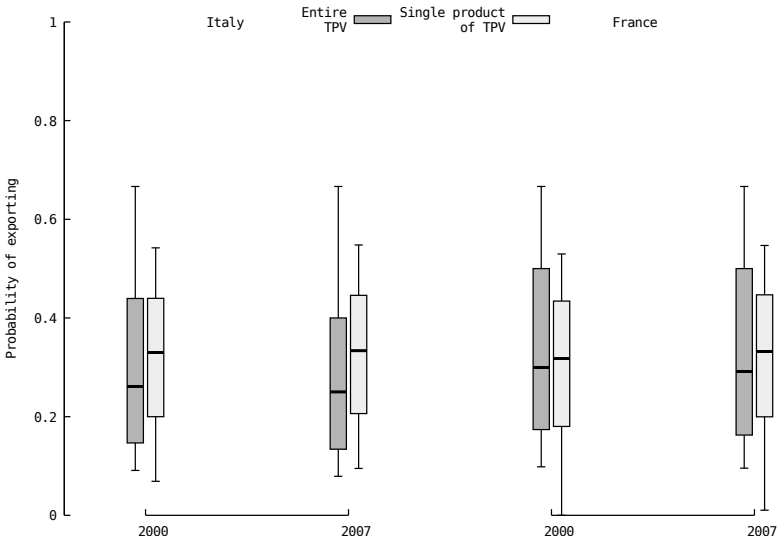

Figure D2: Firm-level box-plots of the distribution of the probability of observing a destination with a typical product vector (TPV) and of observing a destination with single product of the TPV, in both cases possibly together with other products not included in the TPV. Probabilities are computed with random data generated by the statistical benchmark and averaged over 50 replications.

Appendix E

This section presents a set of exercises that test the robustness of our results on the determinants of the composition of firms' product vector with respect to a number of confounding factors.

First, as suggested by a recent empirical literature on multi-product exporters, a firm's export product mix in different markets may be affected by tariffs. Bernard et al. (2011b) show, for example, that falling trade costs cause firms to drop their least-attractive products, while Dhingra (2013) finds that export-oriented firms in Thailand reduce their product lines in response to a unilateral tariff cut.[58] We thus add an additional regressor, $\text{Tariff}_{fd,t}$[59] to specifications (3), defined as

$$\text{Tariff}_{fd,t} = \frac{1}{|\Pi_f|} \sum_{p \in \Pi_f} \text{, Tariff rate}_{pd,t}$$

where $\text{Tariff rate}_{pd,t}$ is the tariff paid to export product p to destination d, and $|\Pi_f|$ is the cardinality of the set products exported by firm f. The estimated beta coefficients on $\text{Tariff}_{fd,t}$ appear in columns 1-2 of Tables E1,E2, and E3. Tariffs do not explain our results. The $\text{Tariff}_{fd,t}$ variable is not statistically significant in almost all regressions. This result may raise doubts about the validity of this robustness check, so we also estimate our regression only on those products that are exported to EU countries, where there are no tariffs for French and Italian exporters: the results appear in columns 3-4 of the same tables. The overall story remains the same.

Second, to check that our empirical results do not reflect product-mix heterogeneity caused by destination-country income, we re-calculate our dependent variables and re-run the regressions focusing only on a set of developed countries.[60] A number of empirical contributions have shown that firms are more likely to export high-quality and technologically-advanced products to high-income countries (Verhoogen, 2008; Crino' and Epifani, 2012; Flach and Janeba, 2017). More generally, firms segment markets and adapt products according to destination-country income. The results, in columns 5-6 of the tables, confirm that our findings also hold in exports to developed countries only. The two firm-country level variables continue to be important determinants of firms' product vectors across destinations.

Third, we add a sensitivity check, reported in columns 7-8 of the tables, regarding "carry-along trade", in which manufacturing firms export products that they do not produce (Bernard et al., 2012). In principle, we would need information on both production and exports at the product level to identify carry-along firms. As these data are not available, we make an approximation by excluding products that are contemporaneously exported and imported by the same firm. The findings are robust to this change in product vector composition, excluding the possibility that our results reflect the import of products which are then re-exported by the firm and (possibly) not produced by it.

Fourth, we remove multinational companies (MNCs), which are complex organizations that sprawl across industries and countries. As data on multinationals are available only for Italy, this robustness check only appears in column 9-10.[61] Our baseline results regarding the role of all two firm-country specific variables continue to hold.

[58] Qiu and Zhou (2013) propose a theoretical model where efficient firms expand their export product scope in response to foreign tariff cuts, whereas those of inefficient firms fall.

[59] The variable Tariff is taken from the WITS (World Bank) database. We use the Most Favored Nation.

[60] We define developed countries as those with *per capita* income levels above the 50th percentile according to the World Bank.

[61] Among the work on the behavior of multi-product multinationals see the recent contribution of Yeaple (2013).

Table E1: LEVENSHTEIN DISTANCE: ROBUSTNESS CHECKS

Dependent variable:	LevD$_{fd,t}$									
	Tariff		EU		Developed		No Carry Along		No MNC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	ITALY									
Market size$_{fd,t}$	-0.007^a	-0.009^a	-0.033^a	-0.034^a	-0.015^a	-0.015^a	-0.008^a	-0.010^a	-0.007^a	-0.009^a
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Market concentration$_{fd,t}$	0.008^a	0.010^a	0.023^a	0.024^a	0.014^a	0.016^a	0.009^a	0.011^a	0.008^a	0.010^a
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
ln Sales$_{fd,t}$	-0.029^a		-0.039^a		-0.033^a		-0.028^a		-0.030^a	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Tariff$_{fd,t}$	-0.025^a	-0.031^a								
	(0.005)	(0.007)								
ln AvgQuality$_{fd,t}$		0.002^a		0.004^a		0.002^a		0.001^a		0.002^a
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
N	4,448,652	4,393,183	2,099,722	2,080,221	3,265,531	3,227,052	4,436,880	4,380,579	4,178,310	4,126,640
adj. R^2	0.594	0.566	0.494	0.478	0.553	0.532	0.563	0.538	0.590	0.562
	FRANCE									
Market size$_{fd,t}$	-0.004^a	-0.005^a	-0.037^a	-0.038^a	-0.013^a	-0.014^a	-0.006^a	-0.008^a		
	(0.000)	(0.000)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)		
Market concentration$_{fd,t}$	0.001	0.002	0.035^a	0.036^a	0.001	0.003	0.001	0.002		
	(0.001)	(0.001)	(0.004)	(0.004)	(0.003)	(0.003)	(0.001)	(0.001)		
ln Sales$_{fd,t}$	-0.027^a		-0.036^a		-0.031^a		-0.026^a			
	(0.000)		(0.000)		(0.000)		(0.000)			
ln AvgQuality$_{fd,t}$		0.002^a		0.003^a		0.003^a		0.002^a		
		(0.000)		(0.000)		(0.000)		(0.000)		
Tariff$_{fd,t}$	0.000	-0.000								
	(0.008)	(0.009)								
N	1,879,043	1,802,780	915,048	880,522	1,385,894	1,329,813	1,853,618	1,774,830		
adj. R^2	0.623	0.608	0.533	0.522	0.570	0.557	0.568	0.555		
Firm-Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the regressions for the Levenshtein dissimilarity measure (LevD$_{fd,t}$) between a firm's global product vector (GPV) and local product vectors (LPV). Results show the standard errors (in parenthesis) clustered at the firm-year and destination level: $^c < 0.1$, $^b < 0.05$, $^a < 0.01$.

Table E2: BRAY-CURTIS: ROBUSTNESS CHECK

Dependent variable:	BC$_{fd,t}$									
	Tariff		EU		Developed		No Carry Along		No MNC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	ITALY									
Market size$_{fd,t}$	0.013^a	0.016^a	0.010^a	0.011^a	0.013^a	0.014^a	0.013^a	0.016^a	0.013^a	0.016^a
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Market concentration$_{fd,t}$	-0.012^a	-0.015^a	-0.006^b	-0.007^a	-0.013^a	-0.016^a	-0.011^a	-0.014^a	-0.012^a	-0.015^a
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
ln Sales$_{fd,t}$	0.054^a		0.050^a		0.053^a		0.049^a		0.054^a	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
ln AvgQuality$_{fd,t}$		0.002^a		0.004^a		0.002^a		0.003^a		0.002^a
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
Tariff$_{fd,t}$	0.007	0.004								
	(0.010)	(0.012)								
N	4,448,652	4,393,183	2,099,722	2,080,221	3,265,531	3,227,052	4,436,880	4,380,579	4,178,310	4,126,640
adj. R^2	0.668	0.634	0.679	0.641	0.678	0.643	0.629	0.602	0.668	0.633
	FRANCE									
Market size$_{fd,t}$	0.008^a	0.009^a	0.008^a	0.008^a	0.008^a	0.008^a	0.007^a	0.008^a		
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)		
Market concentration$_{fd,t}$	-0.008^b	-0.009^b	-0.008^a	-0.008^b	-0.008^a	-0.010^a	-0.007^a	-0.008^a		
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)		
ln Sales$_{fd,t}$	0.047^a		0.045^a		0.047^a		0.041^a			
	(0.000)		(0.000)		(0.000)		(0.000)			
ln AvgQuality$_{fd,t}$		0.003^a		0.004^a		0.003^a		0.004^a		
		(0.000)		(0.000)		(0.000)		(0.000)		
Tariff$_{fd,t}$	-0.004	-0.009								
	(0.010)	(0.010)								
N	1,879,043	1,802,780	915,048	880,522	1,385,894	1,329,813	1,853,618	1,774,830		
adj. R^2	0.698	0.674	0.698	0.671	0.701	0.677	0.656	0.640		
Firm-Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the regressions for the Bray-Curtis between a firm's global product vector (GPV) and local product vectors (LPV). Results show the standard errors (in parenthesis) clustered at the firm-year and destination level: $^c < 0.1$, $^b < 0.05$, $^a < 0.01$.

Table E3: RANK CORRELATION: ROBUSTNESS CHECK

Dependent variable:	RC$_{fd,t}$									
	Tariff		EU		Developed		No Carry Along		No MNC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	ITALY									
Market size$_{fd,t}$	0.017^a	0.020^a	0.019^a	0.021^a	0.017^a	0.019^a	0.015^a	0.018^a	0.016^a	0.019^a
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
Market concentration$_{fd,t}$	-0.014^a	-0.018^a	-0.008^c	-0.009^b	-0.018^a	-0.022^a	-0.013^a	-0.016^a	-0.015^a	-0.018^a
	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
ln Sales$_{fd,t}$	0.052^a		0.064^a		0.055^a		0.050^a		0.054^a	
	(0.001)		(0.001)		(0.001)		(0.001)		(0.001)	
ln AvgQuality$_{fd,t}$		0.003^a		0.006^a		0.004^a		0.004^a		0.004^a
		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)
Tariff$_{fd,t}$	0.027^b	0.031^b								
	(0.010)	(0.012)								
N	4,448,535	4,393,066	1,909,348	1,892,824	3,218,127	3,180,944	4,431,909	4,375,852	4,178,189	4,126,519
adj. R^2	0.499	0.467	0.448	0.414	0.483	0.452	0.479	0.453	0.496	0.463
	FRANCE									
Market size$_{fd,t}$	0.011^a	0.013^a	0.010^a	0.011^a	0.009^a	0.010^a	0.008^a	0.009^a		
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)		
Market concentration$_{fd,t}$	-0.010^a	-0.011^a	-0.005	-0.005	-0.013^a	-0.016^a	-0.010^a	-0.012^a		
	(0.002)	(0.003)	(0.006)	(0.006)	(0.003)	(0.003)	(0.002)	(0.003)		
ln Sales$_{fd,t}$	0.049^a		0.059^a		0.052^a		0.045^a			
	(0.001)		(0.000)		(0.001)		(0.001)			
ln AvgQuality$_{fd,t}$		0.005^a		0.007^a		0.006^a		0.006^a		
		(0.000)		(0.000)		(0.000)		(0.000)		
Tariff$_{fd,t}$	0.031^b	0.024^b								
	(0.010)	(0.010)								
N	1,878,983	1,802,722	808,914	781,505	1,362,653	1,308,020	1,848,781	1,770,242		
adj. R^2	0.524	0.501	0.460	0.435	0.504	0.482	0.499	0.484		
Firm-Destination FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the regressions for the Rank Correlation between a firm's global product vector (GPV) and local product vectors (LPV). Results show the standard errors (in parenthesis) clustered at the firm-year and destination level: $^c < 0.1$, $^b < 0.05$, $^a < 0.01$.