Dynamics of relative product diversification in the course of economic development: import-export comparative analysis

Aleksandra Parteka[*] and Massimo Tamberi[**]

Paper to be presented at XIV ETSG Conference (Leuven, 13-15 September, 2012)

This version: 26.06.2012 [Please do not cite]

Abstract

This paper contributes to trade diversification literature by comparing changes in relative (i.e. assessed vis-à-vis world patterns) heterogeneity of import and export structures in the process of economic development. In particular, by focusing on the diversification of imports we add a missing piece to already analyzed export trends. We use highly disaggregated import and export statistics (4963 product lines) for 163 countries (1988-2010) and find that despite differences in levels (imports being typically more diversified than exports, especially at lower stage of economic development) they follow similar evolution in the development process. Progressing relative diversification of both import and exports structures accompanies economic growth. Comparing the results obtained with alternative estimation methods (nonparametric, semiparametric and parametric), we demonstrate the robustness of this finding. As income per capita grows, countries' import and export structures become less specialized with respect to the typical benchmark.

Keywords: diversification, specialization, trade, economic development

JEL: F14, O11

---

* Gdansk University of Technology, Faculty of Management and Economics, Narutowicza 11/12, 80-233 Gdansk, Poland. tel. (+48 58) 348-60-04 (direct), fax (+48 58) 348-60-07. Corresponding author (aparteka@zie.pg.gda.pl). The paper presents the results of the research project realized while A.Parteka was visiting Universitat Pompeu Fabra (Barcelona, Spain) - financial support from the Polish Ministry and Science and Higher Education is gratefully acknowledged (Mobilnosc Plus 652/MOB/2011/0).

** Universita' Politecnica delle Marche, Facolta' di Economia, Piazzale Martelli 8, 60121 Ancona, Italy.

1. Introduction

This paper deals with the empirical relationship between relative trade diversification (analyzed simultaneously from the point of view of exports and imports) and economic development process. In particular, by focusing on the relative aspect, we assess changes in the composition of countries' trade baskets vis-à-vis a typical world pattern of traded products' diversity.

Several arguments motivate the subject our research. Looking from the side of *exports*, their diversification (de-specialization) can be seen as a factor reducing risk and exposure to idiosyncratic shocks - particularly important in case of low income countries, often with very specialized (i.e. poorly diversified) economic structures and/or being dependant on natural resources. Diversification of *imports*, instead, is directly related to the ability of countries to experience welfare and productivity gains resulting from increasing variety in consumer and input goods, respectively. Finally, from 'academic' point of view our import-export diversification study allows us to test jointly the predictions of alternative theoretical models linking product variety with economic growth.

Product (sector) diversity can be analyzed from two different perspectives, depending if the focus is on *(i)* the degree of economic activity concentration assessed versus uniform distribution or *(ii)* relative specialization of countries' economic structure assessed with respect to overall (world) benchmark[1]. Empirical literature on both aspects of diversification and its link with economic development has been expanding rapidly in the recent years. Within the first stream of research, Imbs and Wacziarg (2003) were the first to find specific non-monotonic "stages of diversification" in the course of economic development. By linking measures of concentration (calculated with sector level employment and value added data) with countries' income per capita, they revealed U-shaped curve[2] with initial diversification and re-concentration at higher levels of economic development. This somehow surprising result was confirmed by Koren and Tenreyro (2007) using different production data. Similar effect was found for exports by Klinger and Lederman (2006), as well as Cadot et al. (2011a). Empirical studies remaining within the second stream of literature and focusing on relative patterns of diversification (thus being closer to our paper than the aforementioned ones) find slightly different outcome, with initial diversification but with no evident trend of re-specialization at higher stages of development: de Benedictis et al. (2009) obtain such a result with exports data while Parteka (2010) confirms it for exports and employment.

---

[1] Note that high degree of product diversification implies low degree of product 'concentration' (or 'specialization') so the terms are used as antonyms.

[2] In this particular case the curve illustrates the nonmonotonic relationship between GDP per capita on x-axis and the index of production concentration on the y-axis through nonparametric lowess approximation. U-shape results from the use of an inverse measure of diversification (based on inequality/concentration index): the decreasing part of the U-curve corresponds to decreasing concentration and increasing diversification of economic activity along the development process, while the upward rising part of the U-curve illustrates the re-concentration course.

Main observation based on the examination of current state-of-the-art is that while diversification studies have been appearing frequently in the recent years (see more detailed literature review in the next section), some aspects of the phenomenon still remain unexplored. In particular, compared to the considerable effort made to investigate production or export diversification patterns across countries, empirical evidence concerning the diversification of *imports* is much more scarce and evident research gap is present here. In particular, cross-country studies presenting product-wise import diversification in the context of countries' development process (complementing Cadot et al., 2011a; de Benedictis et al., 2009; Parteka, 2010 analysis concerning exports) are missing.[3] In other words, we still lack of detailed empirical evidence on how the set of imported goods changes along the path of economic growth of various countries. Moreover, it would be interesting to directly compare patterns of import and export product wise diversification within the same sample of countries, checking if diversification of exports goes in parallel with diversification of imports.

Consequently, the aim of this paper is to fill in the research gap concerning empirical evidence on the process of relative diversification of imported goods, comparing it with exports diversification. We draw on highly disaggregated import statistics (HS6 – subheadings), matching them with equally detailed data on exports[4] and various country-specific characteristics. Our sample (163 countries observed across 23 years) enables us to compare patterns of relative diversification visible in import and export structures of countries at very different levels of economic development.

Figure 1 demonstrates that developed countries have typically more diversified import and export structures than developing economies, and in general major heterogeneity of exported goods is positively correlated with highly diversified basket of imported products. We explore this issue. Employing a wide range of measures (number of active and new product lines, synthetic indices of relative diversification) and alternative estimation methods (nonparametric, semiparametric and parametric) we are also able to examine the robustness of revealed curve of relative diversification along the course of economic development.

[3] We leave aside another strand of literature which focuses on diversification of imported goods (mainly inputs) and their influence on employment and labor markets (especially in terms of labor substitution across countries and resulting wage effects) as not directly linked to our research.

[4] We use mirrored data – see Data Appendix.

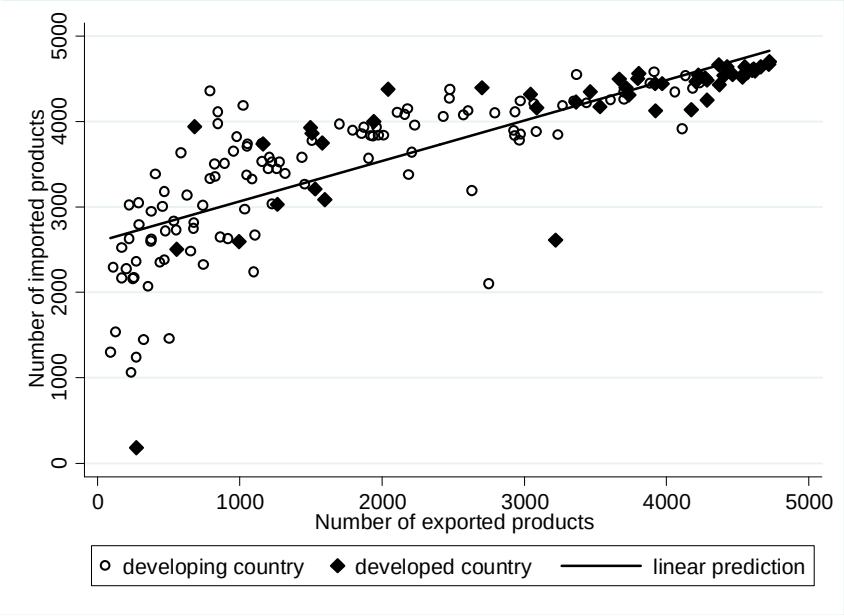

Figure 1. Diversification of imports versus diversification of exports – countries divided according to the development level

Note: Plot shows average number of imported products versus average number of exported products (theoretical max: 4963 product lines) - average values are calculated across time for 163 countries present in our sample (listed in Appendix 1A). Countries division into developed and developing ones based on World Bank's (2011) classification using GNI per capita 2010 (developing countries = low and middle income countries)

Source: own elaboration based on trade data from UNComtrade, 2011 (HS0, 4963 product lines).

Our results can be summarized as follows. We find that despite differences in levels (imports being typically more diversified than exports, especially at lower stage of economic development), they follow similar evolution in the development process. In line with love-of-variety models, progressing diversification of both import and exports structures accompanies the process of economic development and this finding is robust to changes in the methodological setting. Our relative perspective allows us to conclude that as income per capita grows, countries' import and export structures become less specialized with respect to the world pattern of trade. Through the comparison of alternative estimation methods we demonstrate the robustness of monotonically decreasing relative diversification curve (both for exports and imports).

Our paper is structured in the following way. In Section 2 we first give an overview of theoretical contributions which provide us a conceptual guideline for the subsequent empirical analysis. Then, we describe existing empirical evidence directly linked to the theme of our research. In Section 3 we present the data and the methodology used to analyze trade diversification patters, along with some descriptive statistics. Section 4 is devoted to the exploration of the changes in the degree of import and export diversification along the path of economic growth. We present our conclusions in Section 5.

2. Theoretical and empirical background for the study on trade diversification

2.1 Theoretical considerations

Most literature related to our research is purely empirical and not explicitly based on the predictions of theoretical models. However, alternative explanations of product differentiation can be searched for in trade theory and growth/development literature, offering two ways of thinking – focusing on demand or supply side of the phenomenon. Conventionally, 'old' trade theory (H-0 and Ricardian frameworks) draws the attention to productivity gains resulting from specialization; 'new' trade theory built on Dixit-Stiglitz (1977) emphasizes consumer gains from variety, while growth and development literature naturally views the rise in the differentiation of produced goods as a fundamental component of the development process.

As a starting point we could consider Armington's (1969) model of nationally differentiated goods (i.e. goods are differentiated by national origin and countries are specialized in different goods) in which each country is assumed to produce a fixed variety of good and thus there is no extensive margin. Large and more productive economies produce more, exploiting exclusively the intensive margin of trade which lowers prices of each variety; as a result larger economies intensively export higher quantities at lower prices.[5] The assumption that products are differentiated by their country of origin, but the number of varieties supplied by each country is fixed, is known as the 'Armington assumption'.

Instead, in monopolistic competition models of trade (assuming a market with a large number of firms, each producing a unique variety of a differentiated product) the number of varieties produced in each country *varies* due to free entry. 'New' trade theory contributions (Krugman 1979a, 1979b, 1980, 1981; Helpman and Krugman 1985) focused on benefits from international trade stemming from the expansion of available product varieties. Using Dixit-Stiglitz (1977) approach[6], Krugman introduced an explicit reference to the linkage between welfare gains and the variety of goods, formalized through CES utility function, capturing the notion of the so called "love of variety": consumer's utility increases with the number of product varieties available. The number of goods produced in the market depends

[5] Hummels and Klenow (2002) point out that Armington's model predictions, ignoring extensive margin of trade, are at odds with the data (e.g. missing two-thirds of how larger economies export more and one-third of how they import more). Acemoglu and Ventura (2002) extend Armington's view, adding endogenous capital accumulation and endogenous number of varieties. In equilibrium the number of varieties produced by a country is proportional to its employment so countries with more workers produce and export more varieties, but there is no room for extensive imports margin (all countries import all varieties).

[6] Dixit and Stiglitz assumed a single representative consumer demanding many varieties of differentiated good ('love of variety'); in contrast to the Lancaster's approach where consumers differed in their 'ideal variety' of a differentiated good.

on the size of the market itself. Assuming fixed cost of production, hampering the explosion of available varieties, the number of varieties produced in a country depends on the country size, so in Krugman's view higher GDP increases the number of varieties produced (extensive margin) rather than quantity produced per variety (intensive margin). Krugman (1979a) demonstrated that gains from trade can occur even if it takes place between similar countries, contrary to the 'traditional' view explaining trade as a result of international differences in technology (Ricardo) or factor endowments (H-O). Trade may be simply a way of extending the market, allowing exploitation of scale economies and provoking welfare gains to consumers as the number of available products increases. Additionally, 'home market effect' (Krugman, 1980) foresees that with two countries trading, the larger market will produce a greater number of products and will be a net exporter of the differentiated good.[7]

Several other authors enriched the basic workhorse Krugman's (1980) model, adding or modifying one or more of its features.[8] As an example, Melitz (2003) introduced heterogeneity in firms' productivity; in his model total range of produced varieties is allowed to vary with the exposure to trade while the subset of varieties that are consumed in a given country is endogenously determined. He shows that even though the exposure to trade forces less productive firms to exit the market, the decrease in the number of firms in a country after the transition to trade is actually dominated by the number of new foreign exporters and trade – despite being costly – generates a welfare gain in terms of greater product variety available to the consumer. Chaney (2008), extending Melitz (2003) model, apart from firm heterogeneity added asymmetries in trade barriers among countries. Hummels and Lugovskyy (2009) proposed a model useful in explaining cross-country differences in the rate of variety expansion. Building on the results of Hummels and Klenow (2002), who showed that large countries import more varieties, but that varieties differences are less than proportional to market size, they put forward a model in which when a new variety enters the market, goods become more substitutable. Marginal utility of new varieties falls with the size of market (i.e. with the number of varieties): differently from the Krugman model, where marginal utilities of the existing goods are insensible to the introduction of new goods, the introduction of new varieties "crowds" the market.

Turning to the growth literature, increase in product variety (implying major diversification of available products) is seen as a necessary requirement for long term development in endogenous growth theory: "we do not have just more of the same goods and services; we *also* have new ones"

---

[7] In a model a'la Armington, with fixed number of products supplied by each country, a larger market, having bigger demand, will be a net importer.

[8] The basic framework was also used in Krugman (1979b) to develop a model of international product cycle with catching up processes, where economic growth is represented by a growth in varieties, through the introduction of new goods by leader countries and imitation (catching-up) by followers. In the end, the model provides a steady-state share of varieties between the leader and the follower (with incomplete catch-up).

(Aghion and Howitt, 1999, p.1, italics is ours). The natural idea is that the history of economic development consists in the increase in the number of produced and consumed goods. Innovation in consumer goods is needed to satisfy growing consumers' needs (Kuznets, 1962). Then, technological progress directly affects the structure of production through the discovery of new materials and ways of production. These are neo-Schumpeterian models of growth that allow for the introduction into an economy of new or improved types of good either through discovery (advanced countries) or through imports (developing countries). Many endogenous growth models of expanding product variety (e.g. Grossman and Helpman, 1990, 1991 – chapters 3, 8, 9) actually rely on monopolistic competition framework used in 'new' trade theory, but instead of focusing on differentiated final products, they consider differentiated intermediate inputs. Main ideas is that an increase in the variety of differentiated inputs will result in an increase in output (it is modelled through CES production function), exactly as a raise in variety of final goods increases total consumer utility in love-of-variety models. Size effects in dynamic endogenous growth models are analogous to those concerning consumer's utility in Krugman's models: increasing a country's size leads to increased variety of intermediate inputs which results in efficiency gains in final good production and reduces fixed costs of inventing new inputs.[9]

Hence, in the light of endogenous growth theory, growth goes hand in hand with diversification both in consumption and in production. Desmet and Parente (2009) address this issue, developing a theoretical model of 'unified growth theory' to explain the first industrial take-off (i.e. the Industrial Revolution). They start from the empirical observation that modern economic growth has been linked to cost-reducing technological innovation, but also to innovations in consumer products as well. The model is based on a gradual expansion of the market, associated with an increasing variety of consumer goods – Lancaster's utility function is matched with product and process innovation where price elasticities of demand play a central role in the mechanism of diversification.

2.2 Related empirical literature

As already mentioned in the introduction, Imbs and Wacziarg (2003) paper boosted empirical research on the evolution of countries' diversification process with respect to their income per capita levels. Papers which appeared afterwards provided a natural extension of their work, either repeating

[9] Actually, fixed costs modelling is important: if, as in static monopolistic competition models, fixed cost is a fixed amount of labor, then it will prevent the proliferation of the number of developed goods. When fixed costs are assumed to be inversely proportional to the number of products already developed (as in Grossman and Helpman, 1990, 1991), then growth of inputs does not need to cease. Romer (1994) considers fixed cost of exporting to each market and explains why larger economies import in more categories. Melitz (2003) model features fixed cost of production and fixed cost of exporting.

similar empirical exercise with different data or exploring in more detail the diversification course and its determinants. Given trade orientation of our paper, we will focus on trade studies only.[10]

Given the data we use, our study is close to the one by Cadot et al. (2011a) who use highly disaggregated export statistics (4991 HS0 products) to analyse the degree of export concentration, in general confirming what was found by Imbs and Wacziarg (2003) with sector-level production data. They argue in favor of U-shaped (or hump-shaped, as they call it) pattern of export diversification and explore alternative (intensive and extensive) margins of trade. By decomposing the Theil index of export concentration in a large panel of 156 countries (141 of them used in the regressions) across the years 1988-2006, they find that diversification process takes place mainly along the extensive margin (which refers to newly traded/disappearing goods). Consequently, even though growth at the intensive margin (activity across goods already present in the export portfolio) appears to be the main component of export growth, a rising number of new active export lines dominates in terms of diversification process. Increasing part of the U curve in their analysis corresponds to high income countries which tend to close down export lines faster than they open the new ones (reconcentration of exports).

However, findings of Cadot et al. (2011a) - similar to those of Imbs and Wacziarg (2003) and Koren and Tenreyro (2007) obtained for production - are based on the application of standard inequality indices (typically Theil index) to measure the degree of product diversification in terms of the differences from the uniform distribution. We are more interested in relating single countries' product heterogeneity to the 'overall' benchmark trend so, from the point of view of adopted methodology, our study remains closer to those adopting relative measures of trade diversification. De Benedictis et al. (2009) draw on manufacturing export sector level data (539 sectors) for 39 countries (1985-2001). They find that after controlling for countries' heterogeneity through the inclusion of country fixed effects into semi-parametric GAM[11] estimation, sectoral export diversification increases with income per capita. They find that such a trend of export despecialization is more pronounced in the early phases of economic development. Parteka (2010) matches export and employment statistics for 17 manufacturing sectors in 32 countries (1980-2000) and compares relative diversification patterns with those emerging from the application of absolute concentration measures. She finds general tendency of decreasing relative specialization (so progressing diversification) at the beginning of the development process.

Apart from exploring the pure link between diversification and income per capita levels, some effort has been put to discover forces which can be important in influencing export diversification

[10] Cadot et al. (2012) provide a good survey of the empirical literature on trade diversification and its linkages with economic growth.
[11] Generalised Additive Model – see Section 4.3.

opportunities. Among key factors which seem to be driving the diversification process (concerning exports), other then GDP per capita, empirical literature points out mainly such country specific characteristics as: domestic market size and factors influencing ease of access to foreign markets (distance, barriers to trade). These are the main conclusions of the analysis by Parteka and Tamberi (2011) performed for 60 countries and twenty years (1985-2004). They are confirmed by the results obtained by Dutt et al. (2011) who explore gravity model for the two (intensive and extensive) margins of exports, finding additionally that the membership in multilateral trade agreement (WTO) has been helpful in raising the extensive margin of trade. This is line with Cadot et al. (2011b) who also demonstrate positive link between the degree of export diversification and trade liberalization.[12]

Import-rooted empirical literature on product diversification and its relationship with development process is much more scarce. Most studies concerning diversification of imports are country specific and deal with micro-level consequences of increased imported products variety in terms of welfare and/or productivity gains (among others: Broda and Weinstein, 2006; Goldberg et al., 2010; Pavcnik, 2002; Amiti and Konings, 2007; Fernandes; 2007). However, the literature on imports diversification performed in an international panel setting is not well developed. Consequently, empirically not much is known about the evolution of imports in terms of their specialization (or diversification) in the process of economic development observed in a multi-country context. The work by Cadot et al. (2010) stands out as an exception but instead of analyzing product-wise diversification patterns common in export literature (as in Cadot et al., 2011a) and which remains our main interest, they focus on geographical aspect of import diversification (i.e. diversification of suppliers of single products). They find that OECD imports are highly diversified in terms of suppliers, but since 1999, due to the growing importance of Chinese products in OECD imports, a re-concentration took place (turning point occurs around $40,000, 2005 PPP).

To the best of our knowledge there is no published empirical research presenting simultaneously evolution of both import and export diversification patterns along the growth process in an international panel data setting and estimated with the use of highly disaggregated product-level statistics. Our paper goes into this direction. In the next section we present the data we use and some crucial stylized facts.

[12] A contrasting result was obtained by Agosin et al. (2011) who, using a panel of 79 countries covering the period 1962-2000, surprisingly and at odds with theoretical predictions of product differentiation models in heterogeneous firms context (Melitz, 2003), find that trade openness induces higher specialization.

3. Empirical Setting

3.1 The data

In order to calculate synthetic measures of import and export diversification (defined below) we use trade statistics from UNComtrade database (accessed through WITS) at the highest level of disaggregation available within the Harmonized System of goods' classification (HS6 corresponding to sub-headings). We use the data for the years 1988-2010 and after performing product lines concordance exercise across various revisions of the HS data[13] and eliminating never traded goods (see Data Appendix for the details), we are left with 4963 product lines. With these statistics we calculate, for all countries and years, our main variables of interest – the number of active trade lines and new imported/exported product lines, synthetic measures of import and export diversification (defined below), as well as intensive and extensive margins of trade.

Importantly, we match import statistics with export data at the same level of disaggregation, so that our analysis can be read in parallel with similar export-focused research on diversification (e.g. Cadot et al., 2011a also drawing on HS0 product level data). In order to overcome the well known problems in self-reported export flows, we use mirrored data for exports. Country-level alternative indices of both imports *and* exports diversification (obtained for the same group of countries) are then matched with data on income per capita levels (PPP adjusted, const. 2005 int.USD) from the World Bank's World Development Indicators database.

After necessary clearing of the original statistics from missing or misreported values (see Data Appendix for the details) we dispose of 1905 country-year observations for which it was possible to match import and export data. The composition of our panel is summarized in Table 1 where countries are divided into income groups according to World Bank's definition[14]. In the end our panel consists of 163 countries (48 developed and 115 developing ones) across 23 years (1988-2010). The actual number of countries and observations used in the estimations is slightly lower because we drop from the estimations evident outliers (defined, on variable-per-variable basis, as observations below 1st or above

[13] Our analysis covers 23 years during which undoubtedly new products have appeared on the market. Concordance tables between older and newer revisions take in into account in an indirect way – usually new products, which exist in newer revisions as separate codes are included as part of more aggregate code in the older revision of trade data. For instance, data on mobile phones, classified in HS2007 revision as 'Telephones for cellular networks/for other wireless networks, other than Line telephone sets with cordless handsets' (HS2007 code 851712) is included in HS1988 nomenclature as part of the product line 'Transmission apparatus for radio-telephony, radio-telegraphy, radio-broadcasting or television incorporating reception apparatus' (HS1988 code 852520). Consequently, increasing importance of cellular phones in world trade would be visible in the data classified according to older nomenclature (HS1988) as a rise in the value of product line 852520. It indeed is the case – between 1988 and 2011 its import value rose by 300 times.

[14] Economies are divided according to 2010 GNI per capita, calculated using the World Bank Atlas method. The groups are: low income, $1,005 or less; lower middle income, $1,006 - $3,975; upper middle income, $3,976 - $12,275; and high income, $12,276 or more. Low income and middle income countries are classified by the World Bank as "developing countries".

99th percentile), corresponding to possible errors in the original data, misreported trade values or countries with extremely low/high income per capita. The panel is not balanced, mainly due to varying product level trade data availability across countries. Average number of year-observations per country is 11 (min=1, max=23)[15]. Also the set of countries (listed in Table 1A in the Appendix) varies across years and ranges from 11 in 1988 to 130 in 2006. The biggest group in our panel is represented by upper-middle income countries. Altogether, the countries present in our analysis on average (annual mean 1988-2010) correspond to 76% of world trade (imports): from 22 % (in 1988) to 92% (in 2006).

In the subsequent part of the analysis we use additional country specific characteristics potentially important in explaining diversification patterns (such as: population size, share of fuel exports - coming from World Bank's WDI database or geographical data from Gallup et al., 1999).

Table 1. Composition of country-year dataset matching trade diversification indices with income per capita data

		Dataset based on HS0 trade data (4963 product lines)					
		All	Countries divided by income group*				
		countries	LI	LMI	UMI	HI-nonOECD	HI-OECD
Total number of country-year obs.		1905**	178	333	609	195	590
Time span		1988-2010	1989-2010	1988-2010	1988-2010	1989-2010	1988-2010
Number of countries		163	24	43	48	17	31
Number of year obs.	Mean	11	7	8	12	19	11
per country	Min	1	1	1	1	1	12
	Max	23	20	19	22	22	23
Number of country obs.	Mean	82	8	14	26	8	25
per year	Min	11	1	1	1	9	3
	Max	130	18	27	44	31	14

Note:*country groups according to the World Bank's (2011) classification: LI- low income, LMI - lower middle income, UMI - upper middle income, HI- high income (OECD and non-OECD members)
**less when considering the number of new lines (due to 3-year or 5-year moving window- see explanation in text)
Source: own elaboration

3.2 Adopted trade diversification measures and emerging stylized facts

3.2.1 Number of imported/exported goods, relative extensive margin of trade and new imported/exported products

Following the contribution of Hummels and Klenow (2002, 2005), recent literature related to our study acknowledges the fact that diversification of traded goods can take place at alternative margins. *Intensive* margin of trade describes changes in diversification in a set of already traded goods, while *extensive* margin concerns newly traded (or disappearing) goods.[16] In the simplest way, imports (exports) diversification at the extensive margin can be measured by a change in the number of active

[15] These values change (average and max values are lower) when we consider the number of new products – adopting 3-year or 5-year moving windows to calculate them results in a drop in final number of country-year observations. In the regressions involving the use of country fixed effects we retain only countries with multiple (at least 2) observations.
[16] In particular, Cadot et. al (2011a) conclude that product diversification of exports tends to take place at the extensive margin.

import (export) lines - diversification at the extensive margin occurs when trade portfolio expands and the number of active lines rises. Hence, in first instance, for each country and year we calculate the number of active product lines with nonzero import flow (N_imp_{it}). Analogously, N_exp_{it} measures the number of active (non-zero) export product lines. As an alternative, we relate the number of active import lines N_imp_{it} reported by country i to the number of products effectively imported at the world level, obtaining $RelN\_imp_{it} = N\_imp_{it} / N\_imp_{WLDt}$. In the same way, $RelN\_exp_{it} = N\_exp_{it} / N\_exp_{WLDt}$ measures the *relative* number of active (non-zero) export product lines.

However, not all the goods have the same weight in the world trade, so apart from crude counting of active import/export lines (assigning them equal importance independently on their share in world imports/exports), we adopt for our purposes the definition of extensive margin of trade of Hummels and Klenow (2002), originally designed for exports. They introduce appropriate weighting schemes according to the economic 'importance' of goods so that active trade lines are weighted by their share in world trade (note that in such a way we obtain a *relative* measure). Lets denote a set of products imported by country i as K^i and a set of all traded goods in the world as K^{WLD}. Extensive Margin of imported products of country i, EM_imp_{it}, can then be measured as the share of products belonging to i's import portfolio in world trade:

$$EM\_imp_{it} = \frac{\sum_{K^i} imp_{WLDkt}}{\sum_{K^{WLD}} imp_{WLDkt}} \qquad (1a)$$

where $k = 1,\ldots,n$ refers to products and WLD stands for World[17]. Consequently, EM_imp_{it} expresses the share of products imported by country i in world structure of imports and gives the information on how much the goods which country i imports count in world imports.

Similarly:

$$EM\_\exp_{it} = \frac{\sum_{K^i} \exp_{WLDkt}}{\sum_{K^{WLD}} \exp_{WLDkt}} \qquad (1b)$$

denotes the measure of Extensive Margin of products exported by country i and expresses their share in world structure of exports.[18]

In Table 2 we report average number of traded goods and average measures of extensive margin, separately for imports and exports, in all sample and in the subgroups of countries divided by

---

[17] As in Hummels and Klenow (2002) we consider world values of trade net of country i for which extensive margin is being calculated.

[18] Corresponding Hummels and Klenow measures of Intensive Margin (IM) of trade would reflect country i's market share in what it imports (exports), by relating the value of country i's imports (exports) to the value of world imports (exports) of only those products that are imported (exported) by country i:

$IM\_imp_{it} = \frac{\sum_{K^i} imp_{ikt}}{\sum_{K^i} imp_{WLDkt}}$ in case of imports and, analogously in case of exports, $IM\_\exp_{it} = \frac{\sum_{K^i} \exp_{ikt}}{\sum_{K^i} \exp_{WLDkt}}$.

income. Additionally, we accompany the evidence on N_imp_{it}, N_exp_{it} and EM_imp_{it}, EM_exp_{it} with the information on the number of '*new* product lines'. The reason is that measuring diversification based only on the information on currently *active* traded products traded can be oversimplified, as trade relationships tend to be very dynamic, with many products appearing in country's trade statistics in one year and not traded afterwards (see Besedes and Prusa, 2006 on evidence concerning U.S. import relationships, half of which appear to be observed for a single year). In order to take into account the survival of trade flows, in the first instance for every country at time *t* we calculate $N\_new1\_imp_{it}$ in the spirit of (Besedes and Prusa, 2006): new product is defined here as such which was not imported in the prior year but is still imported in the following year (with respect to time *t*), so the definition is based on one-year cutoff and three-year moving window to define the spell. $N\_new1\_exp_{it}$ is obtained in the same way for new exported goods. Secondly, we consider relatively more strict definition as in Cadot et al. (2011a): $N\_new2\_imp_{it}$ is our second measure of the number of new imported products which are defined as those products that were not active in country's import portfolio in the preceding two years but were imported in the subsequent two years – in this definition a moving five-year window is adopted. $N\_new2\_exp_{it}$ denotes similar measure of new exported goods based on two year cutoff.

Table 2. Number of imported/exported products, extensive margin of trade and new import/export lines – averages (overall and by income group)

	All countries	Countries divided by income group*				
	(1)	LI (2)	LMI (3)	UMI (4)	HI-nonOECD (5)	HI-OECD (6)
GDPpc [PPP, const.2005 int.USD]	14291	922.52	2971.01	8173.75	26384.50	27031.40
Population [1000]	48401	25302	80446	67029	3717	32987
Share of fuel exports [%]	13.2	3.2	14.4	16.6	29.2	6.3
Imports:						
N_imp_{it}	3917	3063	3476	3931	3746	4467
$RelN\_imp_{it}$	0.80	0.63	0.71	0.80	0.77	0.91
EM_imp_{it}	0.85	0.77	0.82	0.87	0.81	0.90
$N\_new1\_imp_{it}$	89	154	127	99	99	49
$N\_new2\_imp_{it}$	35	47	50	39	38	23
Exports:						
N_exp_{it}	2713	1015	1685	2547	2303	4066
$RelN\_exp_{it}$	0.56	0.21	0.35	0.52	0.47	0.83
EM_exp_{it}	0.70	0.42	0.57	0.70	0.69	0.87
$N\_new1\_exp_{it}$	182	156	199	223	208	131
$N\_new2\_exp_{it}$	68	49	71	85	87	49

Note: average values across country-year observations (163 countries, 1988-2010) *country groups according to the World Bank's (2011) classification: LI- low income, LMI - lower middle income, UMI - upper middle income, HI- high income (OECD and non-OECD members)

Source: own elaboration based on trade data from UNComtrade (2011); GDP per capita, population and share of fuel exports (as % of all merchandise exports) from WB WDI (2011).

As for the average number of active import lines in the whole sample of 163 countries (column 1 of Table 2), it is relatively high with 3917 non-zero import lines per country per year (theoretical max=4963 HS0 lines) which corresponds to 80% of products imported at the world level. In comparison, average number of exported products is much lower (only 2713 per country per year) and reaches only 56% of the set of goods actively exported in the world. Also the average measure of extensive margin of exports is lower than in case of imports (0.70 versus 0.85). Consequently, there is room for activity at extensive margin of exports and the process of adding new lines to the export portfolio is much more dynamic than in case of imports: depending on the definition of new products used (clearly, bigger numbers are obtained with the measure based on one-year cutoff: $N\_new1\_imp_{it}$ and $N\_new1\_exp_{it}$) on average countries from our sample are characterized by 182 (or 68) new exported products annually, compared to only 88 (or 35) in case of imports portfolio.

Looking at different countries according to their level of development (columns 2-6, groups are presented in ascending order of average income per capita), it is evident that, especially in case of exports, stages of diversification can occur. The number of exported products tends to grow with income per capita: average N_exp_{it}, ranges from 1015 (21% of world set of exported products) in case of low income countries to 4066 (83% of world set of exported products) in case of high income – OECD members countries. Similarly, looking at variation in $EM\_\_exp_{it}$ across income groups, it can be seen that on average the share of products belonging to low income countries export portfolio in world exports is approximately two times lower (0.42 versus 0.87) than in case of high income (OECD) countries. Note that the considerable difference in the number of exported products and extensive margin of exports between high income-OECD and non-OECD members, despite their similar average income per capita levels, is due to the fact that in the latter group there are many petrol-abundant countries typically having very concentrated exports structures. As reported in Table 2, average share of fuel exports in all merchandise exports in each of the 5 income groups is as follows: LI 3.2%, LMI 14.4%, UMI 16.6%, HI-OECD 6.3%, HI-nonOECD 29.2% (average values for countries and years present in our dataset, data from WB WDI, 2011). The latter group is also composed of some very small countries.

Actually, the comparison between product variety of exports and imports is very informative – while there is a substantial room for exports diversification at the extensive margin (thus by increasing the number of exported products) in case of developing countries, the degree of import diversification is much more stable across income groups. Quite surprisingly, even low income countries (column 2 of Table 2) cover in their import portfolio already 63% of products imported at the world level (3063 product lines) while they export only 21% of all goods exported internationally. Unsurprisingly, the set of imported products expands as countries move to higher categories of income and high income

OECD countries are characterized by the biggest average number of imported products (4467 active import lines corresponding to 91% of international set of imported goods), but the process is not as evident as in case of exported goods.

Turning back to the number of new products, clearly developing countries expand their import and export portfolio by adding new lines. Sticking to less stringent definition of new lines in line with Besedes and Prusa (2006), $N\_new1\_imp_{it}$ and $N\_new1\_exp_{it}$, low income countries (column 2 of Table 2) add on average 154 new import lines and 156 new export lines annually. For middle income countries these values are, respectively, equal to 127 and 199 (column 3 of Table 2). The comparison between the process of adding new product lines to export and import portfolio for countries at higher stages of development is interesting. Of course, high income OECD countries (column 6 of Table 2), which are already very close to the maximum number of goods traded, tend to be characterized by relatively low number of *new* imported and exported goods. Differences between the expansion of imports and exports are evident in case of upper-middle income and high-income non OECD countries (columns 4 and 5 of Table 2). They continue to expand rapidly the set of exported goods (more than 200 new export lines annually), but having already well diversified imports, add less new import lines annually than developing countries. This is quite logical if we recall that since the very beginning of the development process product heterogeneity of imports is much higher than that of exports: already in case of low income countries their set of imported goods is three times more heterogeneous than the set of exported products.

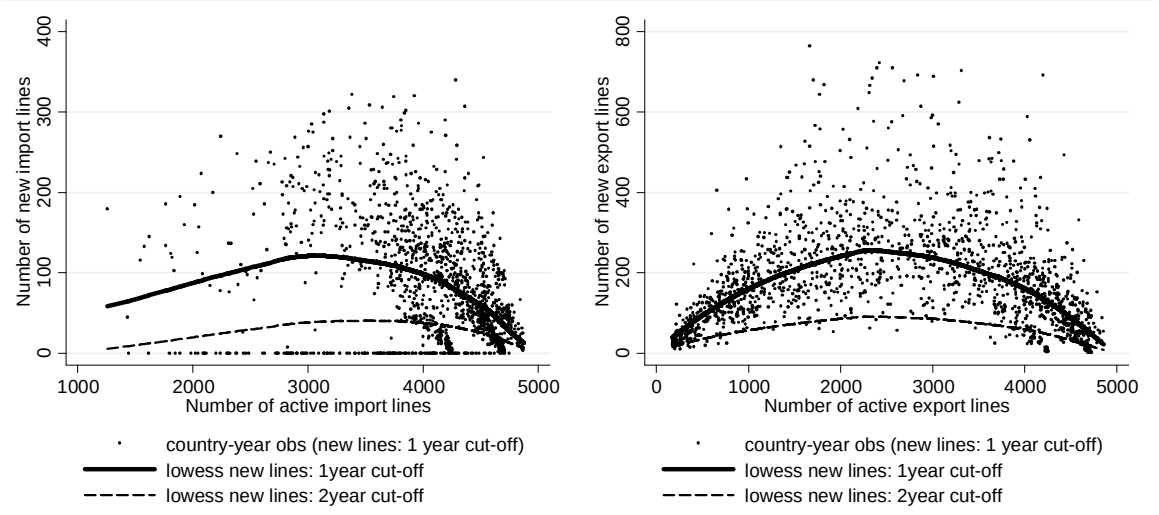

Figure 2. Number of new imported (exported) products versus number of active import (export) lines (left plot: imports, right plot: exports)

Note: lowess - span=0.8, sample without outliers defined as observations below 1st or above 99th percentile (number of obs.=1828)

Source: own elaboration based on trade data from UNComtrade, 2011 (HS0, 4963 product lines) and GDP per capita from WB WDI (2011).

This is confirmed by the relationship between active import/export lines and new import/export lines plotted in Figure 2 (showing lowess[19] approximations based on country-year data points). The distribution of the number of exported products is quite symmetric, with only 171 products exported by Maldives in 1995 up to 4858 active export lines in case of Netherlands in 1994 (see also Table 3). On the contrary, overwhelming mass of countries in our sample already have quite diversified import structures, with minimum value of 1257 imported products in case of Central African Republic in 1999 and maximum of 4868 imported products (Canada, 1990).[20] This suggests that import and export diversification processes along the course of economic growth, albeit having something in common (e.g. correlation between N_imp_{it} and N_imp_{it} equals to 0.77), can be actually quite different phenomena – we will go back to this point in the following sections.

3.2.2 Synthetic measures of relative diversification of imported and exported products

In order to measure the degree of trade diversification (or its opposite: overall specialization), researchers usually rely on inverse indices of diversification based on inequality/concentration indicators (inverse because diversification and concentration act as antonyms)[21]. Standard ones such as: Herfindahl index, Gini index or Theil index give the information on how far country's trade flows are from the uniform distribution (so they indicate if the country's imports/exports are concentrated or not, in terms of product shares). At the same time, being *absolute* measures of diversification, they have the disadvantage of isolating country-specific trade patterns from those typical for the whole sample of countries. Our choice, instead, is to apply the variation of standard measures of product-wise diversification of trade flows in the form of *relative* measures: Relative Theil index, Relative Gini index and Dissimilarity Index.[22] They involve measuring country's degree of trade flows' diversification with

---

[19] Lowess curves represent graphical nonparametric approximation of the relationship of interest estimated on the base of data points and obtained with lowess smoother (Clevelend, 1979), useful in cases of suspected nonlinearity in the data. Given a relationship of the form $Y=f(X)$, a smoother is a nonparametric tool used for estimating the trend and the estimate of unknown function f(.) produced by the smoother is known as smooth. We implement locally-weighted smoother (also known as lowess or loess). It is computed in the following steps (notation from Hastie and Tibshirani, 1990: 30). Smooth $s(x_0)$ uses k nearest neighbours (closest points to x_o) denoted by $N(x_0)$. The number of nearest neighbours, usually expressed as a percentage of the data points is the smoothing parameter called span. Next, the distance of the furthest near-neighbour from x_o, $\Delta x_{0} = max_{N(x0)} |x_0 - x_i|$, is computed. Weights w_i are given to each point in $N(x_0)$ using the tri-cube weight function: $W\left[\left(\left|x_0 - x_i\right|\right)/\Delta x_0\right]$, where $W[u]=(1- u^3)^3$ for $0 \leq u < 1$ and 0 otherwise. Such weighting scheme provides decreasing weights (and less relative importance) on observations which are more distant from x_0. Finally, $s(x_0)$ is a fitted value at x_0 coming from the weighted least squares fit of Y to X confined to $N(x_0)$. The procedure is repeated for each observation (the number of regressions is equal to the number of observations) and the fitted values are used for the construction of the nonparametric curve, representing the relationship between Y and X).

[20] Minimum and maximum values in a sample without outliers defined as observations below 1st or above 99th percentile.

[21] The discussion on potential characteristics of "ideal" measures and real limits of existing ones can be found in (among others): Bickenbach et al., 2010; Palan, 2010.

[22] For the comparison between the application of absolute and relative diversification measures to the study of economic diversification see the discussion in Parteka (2010).

respect to the overall pattern of diversification in the world structure of trade (denoted as *WLD*) and thus account for changes in relative importance of products.

Our preference for relative measures of diversification has several justifications. The main advantage is that by using relative measures we account for trends common to all countries (i.e. changes in prices of commodity goods) and do not isolate country-specific trade diversification patterns from those typical for the whole sample. Moreover, we deal with highly disaggregated trade data (HSO, 6-digit data, subheadings) across 23 years, when changes in the relative importance of products undoubtedly occur at the world level - up to the disappearance of goods which would not be captured by the use of 'absolute' indices retaining *n* constant as the number of products theoretically being subject to international trade and present in the classification scheme.[23]

Let's indicate the share of imports of product *k*=1, 2,…*n*[24] in total imports of country *i* at time *t* as: $s\_imp_{ikt} = imp_{ikt} / IMP_{it}$, where IMP_{it} denotes the value of country *i*'s total imports. Similarly, the 'typical' share of product *k* in total imports reported at the world level is given by: $w\_imp_{kt} = imp_{WLDkt} / IMP_{WLDt}$, where imp_{WLDkt} denotes the overall value of imports of product line *k* reported at the world level and IMP_{WLDkt} measures total world imports.[25] Consequently, the ratio of s_imp_{ikt} to w_imp_{kt} would reflect the relative importance of product line *k* in country *i*'s import portfolio vis-à-vis the same product's importance in world imports. In the same way we calculate $s\_\exp_{ikt} = \exp_{ikt} / EXP_{it}$ and $w\_\exp_{kt} = \exp_{WLDkt} / EXP_{WLDt}$ in case of exports.

Relative Theil index is an entrophy measure and in case of imports is given by[26]:

$$\mathrm{Re}lT\_imp_{it} = \sum_{k=1}^{n}\left(s\_imp_{ikt} \cdot \ln\left(\frac{s\_imp_{ikt}}{w\_imp_{kt}} \right) \right) \qquad (2a)$$

and, in case of exports, by:

$$\mathrm{Re}lT\_\exp_{it} = \sum_{k=1}^{n}\left(s\_\exp_{ikt} \cdot \ln\left(\frac{s\_\exp_{ikt}}{w\_\exp_{kt}} \right) \right) \qquad (2b)$$

---

[23] Actually, Bickenbach et al. (2010) show that only the relative measures can account for general long-term changes in the size distribution across industries (in our case – products).

[24] We consider theoretical number of all products listed within our level of disaggregation (HS subheadings) in our cleared dataset (see Data appendix), so that n=4963.

[25] As values of world trade of any particular good we use direct data from UNComtrade on imports of this product where reporter =all countries. This value does not always coincide with the sum of imports of all countries treated separately which could influence the values of w_imp_{kt} As an alternative to taking direct data on world imports from UNComtrade we could have summed imports of every product across reporting countries. However, we have compared series of w_imp_{kt} obtained with these two alternative ways: they are highly correlated (0.99) and the differences between them are negligible.

[26] The formula is based on Cowell (1995): 49. Note that the computation of Theil index involves the use of logarithm so in cases in which the term in internal brackets resulted to be equal to zero (due to the presence of zero trade flow – products not present in country's trade portfolio) we substituted it with a very small number (10^{-10}), which allowed us to compute correctly the index affecting only negligibly its value.

The theoretical lower bound of relative Theil indices is 0 while the upper limit is equal to ln*(n),* where *n* is the number of all product lines. Higher values of Theil indices are associated with more specialised (thus less diversified) trade structures of countries under examination.

Then, for each country *i* and time *t* we calculate Relative Gini index of imports' diversification, based on the computation of Lorenz-type curve of relative diversification and approximated by the following formula (as in Parteka and Tamberi, 2011):

$$\mathrm{Re}lG\_imp_{it} = \sum_{j=0}^{n-1}(p\_imp_{jt} - q\_imp_{ijt}) / \sum_{j=1}^{n-1}\left(p\_imp_{jt}\right) \qquad (3a)$$

where $p\_imp_{jt} = \sum_{k=1}^{j} w\_imp_{kt}\ ,\vee j = 1,...,n$ is the cumulative of the denominator of Balassa indexes (calculated for each country *i* and imported product *k=1,…,n* as $BI\_imp_{ikt}=s\_imp_{ikt}/w\_imp_{kt}$ and ranked in the ascending order) and $q\_imp_{ijt} = \sum_{k=1}^{j} s\_imp_{kjt}\ ,\vee j = 1,...,n$ is the cumulative of the numerator of BI_imp_{ikt}.[27] In the same manner we obtain Relative Gini index of exports' diversification:

$$\mathrm{Re}lG\_\exp_{it} = \sum_{j=0}^{n-1}(p\_\exp_{jt} - q\_\exp_{ijt}) / \sum_{j=1}^{n-1}\left(p\_\exp_{jt}\right) \qquad (3b)$$

The relative Gini index may vary from 0 (maximum diversification) to 1 (the opposite).

Finally, Dissimilarity Index of relative imports product diversification[28] views the relationship between the numerator and the denominator of the Balassa index as the difference (and not the ratio) and is obtained as:

$$DI\_imp_{it} = \sum_{k=1}^{n} \left| s\_imp_{ikt} - w\_imp_{jt} \right| \qquad (4a)$$

Similarly, in case of exports we calculate the measure of relative product diversification of country *i*'s exports from the typical world structure as:

$$DI\_\exp_{it} = \sum_{k=1}^{n} \left| s\_\exp_{ikt} - w\_\exp_{jt} \right| \qquad (4b)$$

---

[27] Graphically, relative Gini index can be calculated in the following steps (Amiti, 1999). Given a set of Balassa indexes ($BI\_imp_{ikt}=s\_imp_{ikt}/w\_imp_{kt}$ ranked in the ascending order, for each country *i* the cumulative of the denominator of BI_imp_{ikt} (denoted in text as p_imp_{it}) is represented on the horizontal axis while the cumulative of the numerator of BI_imp_{ikt} (denoted in text as q_imp_{ijt}) is represented on the vertical axis. The relative Gini index can then be calculated as twice the area between the 'Lorenz curve' and the 45 degree line which is associated with a case when country *i* has the same pattern of revealed comparative advantage as the benchmark (in our case: world). The same procedure can be done for exports using $BI\_exp_{ikt}=s\_exp_{ikt}/w\_exp_{kt}$.

[28] In order to avoid confusion, note that the term "Dissimilarity Index" is sometimes used in the literature with respect to the measure of trade dissimilarity between country's exports and imports.

All the three types of relative indices are positively related to the degree of relative product-wise specialisation and inversely related to the degree of relative diversification. Consequently, higher values of $RelT\_imp_{it}$, $RelG\_imp_{it}$, DI_imp_{it} (or $RelT\_exp_{it}$, $RelG\_exp_{it}$, DI_exp_{it}) are associated with less diversified import (or export – respectively) structures with respect to the overall benchmark trend visible in the world composition of trade.

In Table 3 we show summary statistics of these three synthetic indices of relative diversification, (accompanied by statistics referring to the number of imported/exported products), calculated in a whole sample of countries (excluding only observations below 1st or above 99th percentile). The comparison between mean values referring to imports and exports clearly proves that in general imports are characterised by much higher degree of relative product heterogeneity than exports. Average country in our sample (outliers excluded) covers in its import portfolio 81% of all goods actively traded in the world (min=26% and max=99%, corresponding to 1257 and 4868 active import lines, respectively). On the export side, an average country exports around half (55%) of all goods actively exported in the world; the range is from very restricted set of exported goods (only 171 lines) to almost full export portfolio (4858 lines corresponding to 99% of all products exported in the world). Similarly, average values of the synthetic indices of relative imports diversification are always lower than that of exports.

Table 3. Summary statistics of import and export diversification measures – all countries (without outliers)

Import diversification measures							Export diversification measures						
index	group	obs	mean	sd	min	max	index	group	obs	mean	sd	min	max
N_imp_{it}	all countries	1867	3939	693	1257	4868	N_exp_{it}	all countries	1867	2690	1461	171	4858
$RelN\_imp_{it}$	all countries	1867	0.81	0.14	0.26	0.99	$RelN\_exp_{it}$	all countries	1867	0.55	0.30	0.03	0.99
$RelT\_imp_{it}$	all countries	1867	1.18	0.64	0.23	3.38	$RelT\_exp_{it}$	all countries	1867	2.63	1.37	0.44	6.24
$RelG\_imp_{it}$	all countries	1867	0.66	0.13	0.33	0.91	$RelG\_exp_{it}$	all countries	1867	0.88	0.12	0.49	1.00
DI_imp_{it}	all countries	1867	1.06	0.27	0.49	1.63	DI_exp_{it}	all countries	1867	1.53	0.31	0.73	1.95

Note: Sample without outliers defined as observations below 1st or above 99th percentile.
Source: own elaboration

Table 4. Summary statistics of import and export diversification measures – countries by income group

Import diversification measures							Export diversification measures						
index	group	obs	mean	sd	min	max	index	group	obs	mean	sd	min	max
N_imp_{it}	LI	173	3127	661	1257	4197	N_exp_{it}	LI	168	1067	772	178	3482
	LMI	325	3538	593	1444	4669		LMI	326	1695	1134	177	4641
	UMI	605	3952	599	1819	4826		UMI	608	2532	1296	171	4767
	HI-nonOECD	193	3782	787	1419	4778		HI-nonOECD	194	2302	1279	193	4622
	HI-OECD	571	4454	290	2435	4868		HI-OECD	571	4035	714	377	4858
$RelN\_imp_{it}$	LI	173	0.64	0.14	0.26	0.86	$RelN\_exp_{it}$	LI	168	0.22	0.16	0.04	0.71
	LMI	325	0.72	0.12	0.30	0.95		LMI	326	0.35	0.23	0.04	0.95
	UMI	605	0.81	0.12	0.37	0.98		UMI	608	0.52	0.27	0.03	0.97
	HI-nonOECD	193	0.77	0.16	0.29	0.97		HI-nonOECD	194	0.47	0.26	0.04	0.94
	HI-OECD	571	0.91	0.06	0.50	0.99		HI-OECD	571	0.83	0.15	0.08	0.99
$RelT\_imp_{it}$	LI	167	2.08	0.52	1.18	3.31	$RelT\_exp_{it}$	LI	172	4.58	0.83	2.30	6.22
	LMI	330	1.65	0.44	0.95	3.38		LMI	327	3.57	1.06	1.42	6.13
	UMI	606	1.17	0.45	0.46	3.19		UMI	602	2.72	1.14	0.86	6.24
	HI-nonOECD	193	1.40	0.59	0.57	2.83		HI-nonOECD	195	2.49	0.92	1.11	6.06
	HI-OECD	571	0.58	0.25	0.23	1.94		HI-OECD	571	1.47	0.81	0.44	4.64
$RelGini\_imp_{it}$	LI	173	0.83	0.05	0.69	0.92	$RelGini\_exp_{it}$	LI	167	0.98	0.02	0.93	1.00
	LMI	325	0.77	0.06	0.62	0.92		LMI	331	0.95	0.05	0.75	1.00
	UMI	607	0.68	0.08	0.48	0.92		UMI	607	0.91	0.08	0.64	1.00
	HI-nonOECD	191	0.73	0.10	0.50	0.92		HI-nonOECD	191	0.92	0.07	0.73	1.00
	HI-OECD	571	0.52	0.09	0.33	0.77		HI-OECD	571	0.76	0.12	0.49	0.99
DI_imp_{it}	LI	175	1.41	0.12	1.09	1.63	DI_exp_{it}	LI	165	1.85	0.08	1.56	1.95
	LMI	324	1.28	0.13	0.99	1.60		LMI	328	1.75	0.15	1.28	1.94
	UMI	607	1.09	0.18	0.71	1.58		UMI	608	1.59	0.22	1.02	1.94
	HI-nonOECD	190	1.17	0.20	0.75	1.63		HI-nonOECD	195	1.59	0.21	1.18	1.92
	HI-OECD	571	0.76	0.15	0.49	1.38		HI-OECD	571	1.22	0.26	0.73	1.92

Note: country groups according to the World Bank's (2011) classification: LI- low income, LMI - lower middle income, UMI - upper middle income, HI- high income (OECD and non-OECD members). Sample without outliers defined as observations below 1st or above 99th percentile.
Source: own elaboration

4. Stages of imports' and exports' relative diversification in economic development process

4.1 Revealed relative diversification curve - unconditional nonparametric results (lowess)

In order to verify how trade structures change along the path of economic growth, in the first instance we match synthetic inverse measures of relative diversification of countries from our sample (defined in Section 3.2.2) with their corresponding GDP per capita levels. *A priori*, we do not want to impose any functional form on the relationship between the variables of interest. We plot unconditional nonparametric lowess curves, separately for imports and exports diversification, where each plot corresponds to the nonparametric equation of the following form:

$$DIV\_imp_{it} = f(GDPpc_{it}) \quad (5a)$$

and, similarly, in case of exports:

$$DIV\_\exp_{it} = f(GDPpc_{it}) \quad (5b)$$

where *DIV _imp* and *DIV _exp* denote one of the synthetic inverse measures of relative diversification (respectively for import and exports), *i* refers to countries and *t* to time period, *GDPpc* is a proxy of the development level (real income per capita) while *f(.)* is an unspecified function estimated through the use of the lowess smoother (Clevelend, 1979) and represented graphically[29]. All the plots are corrected for the presence of evident outliers - defined as observations below 1st or above 99th percentile (possible errors in the original data, misreported trade values etc.).

In Figure 3 we demonstrate nonparametric relationship (lowess curves) between the degree of imports product-wise relative diversification and levels of economic development (left plots), compared to analogous patterns obtained with export data (right plots). Here we employ our three baseline relative measures of diversification: Relative Theil ($RelT\_imp_{it,}$ *and* $RelT\_exp_{it,}$), Relative Gini ($RelG\_imp_{it,}$ *and* $RelG\_exp_{it,}$) and Dissimilarity Index ($DI\_imp_{it,}$ *and* DI_exp_{it}). Figures showing the evolution of diversification of imported products in the process of economic growth represent entirely our contribution, as to the best of our knowledge such an exercise has not been done so far.

Each plot shows two lowess lines: grey one obtained with all country-year observations (throwing out only evident outliers) and the black one obtained with observations corresponding to income per capita levels below 40,000 USD (PPP, const.2005). Why do we choose to split the sample in such a manner?

We were interested to check whether the U shaped pattern of diversification, found by Imbs and Wacziarg (2003) for production or Cadot et al. (2011a) for exports and obtained with absolute measures of diversification is confirmed in our sample matching import and export data in a relative setting. In fact, when we consider the whole group of countries, we also obtain the U curve of trade diversification (in Figure 3 it is represented by the grey line, which partly overlaps with the black one). Independently on the index, the line is first decreasing (which suggests relative diversification course at lower levels of economic development) and then, after income per capita level around 30,000 USD (PPP, const.2005) it starts to increase (which suggests relative re-specialisation of trade structures at higher stages of development). This pattern is similar both in case of exports (right plots) and imports (left plots).

[29] See note 19 for the description of lowess procedure.

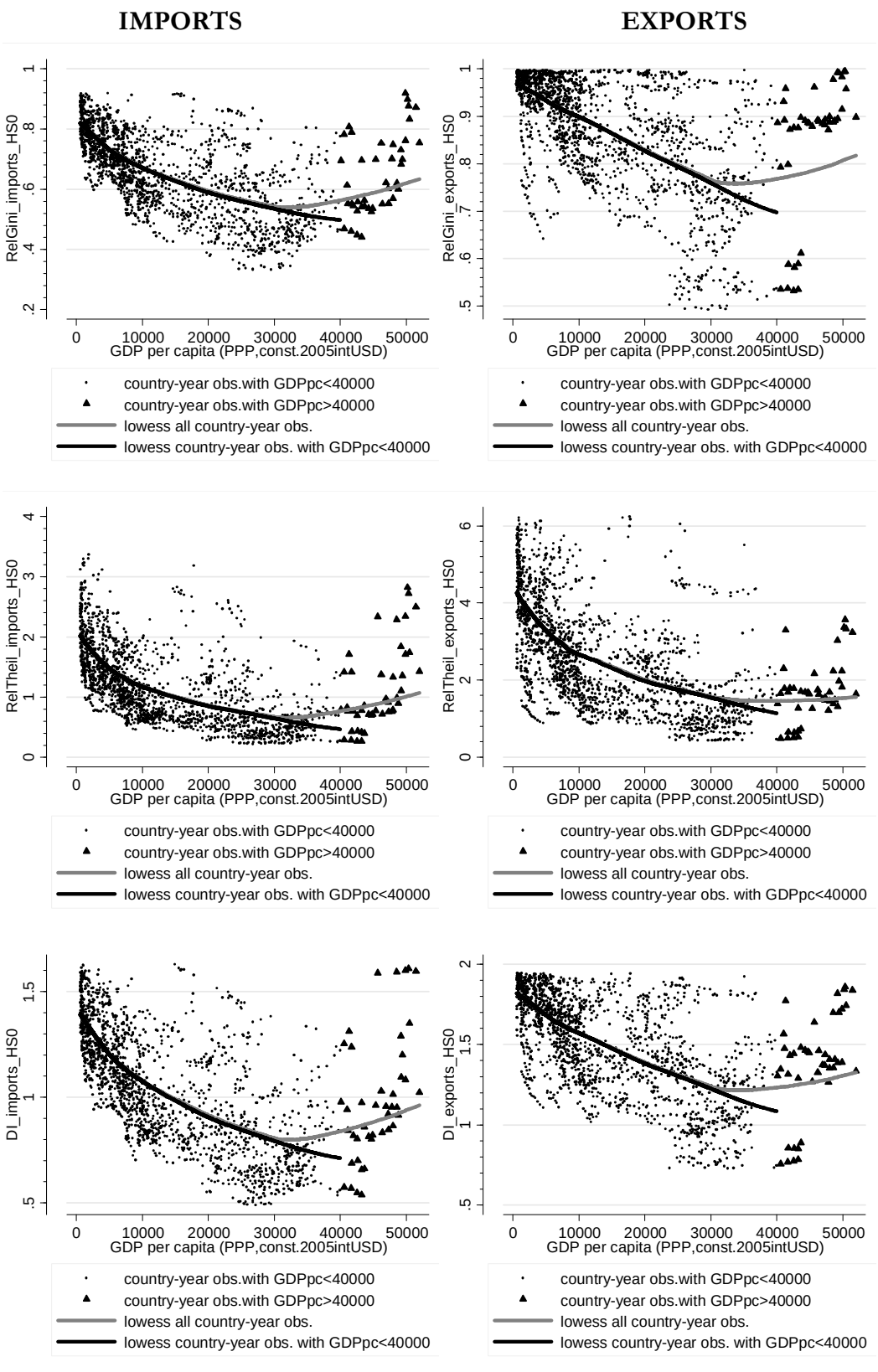

Figure 3. Degree of diversification of imported and exported products versus development level (left plots: imports, right plots: exports), alternative indices

Source: own elaboration based on trade data from UNComtrade, 2011 (HS0, 4963 product lines) and GDP per capita from WB WDI (2011).

Note: lowess - span=0.8, sample without outliers defined as observations below 1st or above 99th percentile (number of obs.=1828)

Table 5. Observations potentially driving upward part of U-shaped lowess curve of relative diversification

Country name	Income group	No of obs.	GDPpc [PPP, const.2005 int.USD]*	Share of fuel exports [%]*	Population [1000]*
Ireland	HI-OECD	1	41025	0.8	4420
Hong Kong SAR	HI-nonOECD	2	41147	4.4	7055
UnitedStates	HI-OECD	8	42415	4.1	300040
Norway	HI-OECD	14	45460	60.9	4556
Macao SAR	HI-nonOECD	2	45901	0.0	512
Singapore	HI-nonOECD	8	46574	12.9	4509
United Arab Emirates	HI-nonOECD	4	48322	21.1	2566
Kuwait	HI-nonOECD	1	49411	96.5	2317
Brunei Darussalam	HI-nonOECD	4	50294	96.6	293
For comparison:					
Rest of the sample**		1784	13100	12.6	48934

Note: *average values across country-year observations identified as potentially driving Upward part of U-curve (marked with triangles in Figure 3 and being above GDPpc of 40,000 [PPP, const.2005 int.USD]. **outliers excluded. Income groups according to the World Bank's (2011) classification: LI- low income, LMI - lower middle income, UMI - upper middle income, HI- high income (OECD and non-OECD members)
Source: own elaboration based on trade data from UNComtrade (2011); GDP per capita, share of fuel exports (as % of all merchandise exports) and population from WB WDI (2011).

However, after a bit of data mining, we actually found out that the upward rising part of the U curve might be guided by a limited number of observations (namely: only 44 out of 1828 in our sample without outliers), corresponding to few specific countries with income per capita levels above roughly 40,000 USD (PPP, const.2005). They are marked in Figure 3 with triangles. We list them, in ascending order of income per capita, along with their crucial characteristics in Table 5. U.S. is one of them: as we will show later on, the country has quite distinct pattern of trade as it clearly follows a path of relative specialization with respect to the overall benchmark. Among other countries, responsible for the upward part of the nonparametric U curve, there are mainly quite specific economies, being either petrol exporting countries (such as Norway, Kuwait or Brunei Darussalam) and/or small countries (in terms of population) which in general are likely to have more concentrated (specialized) trade structures than other countries. Lowess curves of relative diversification based on the remaining observations (97% of all observations in our sample) are clearly decreasing, showing that in general vast majority of countries tend to diversify both their imports and exports portfolio as they develop.

4.2. Country specific patterns of relative imports' and exports' diversification

In Section 4.1 we demonstrated that few, well characterized data-points are responsible for the upward part of the non-parametric relative diversification curve, corresponding to eventual re-specialization path at higher stages of development. This leads us to the importance of country fixed

effects in the correct estimation of the revealed diversification curve. Countries greatly differ from each other because of: size, institutions, geographical characteristics (influencing, among other things, the kind and variety of agricultural products), natural resources, endowments, economic policies (openness policies), etc. and all of these factors can eventually influence the shape of curve revealed from the data.[30]

Examination of country specific patterns of relative trade diversification[31] indeed demonstrates that countries specificity indeed can play a role in the diversification process. To illustrate thus, in Figure 4 we show an example based on two developed (AUS – Australia, and USA) and two developing (ROM – Romania and CHN – China) countries present in our sample. We show simultaneously how the degree of relative imports diversification (measured inversely by Relative Theil index: *RelT_imp*) and relative exports diversification *(RelT_exp*) evolved with respect to each country's income per capita level. We maintain the same scale for all countries to show changes in diversification indices in a comparative setting. Recall that a decrease in the relative Theil index implies progressing relative diversification and thus diminishing specialization of trade basket. Increase in the index marks up progressing dissimilarity (re-specialization) with respect to typical world structure of trade.

It is evident that specific countries may display specific patterns of changes in trade composition and just the level of income per capita is not enough to explain it. In case of ROM, CHN and AUS imports are more diversified than exports (*RelT_exp* > *RelT_imp*; the pattern already revealed in the whole sample), in case of the USA the difference is negligible (very low values of both *RelT_exp* and *RelT_imp* indicate already very high degree of exports and imports diversification). Another highly developed country –AUS - is characterized by very diversified import structure, but relatively concentrated exports (a hint that its location may influence trade costs and hamper export diversification opportunities). In terms of evolution along the path of growth, Romania demonstrates quite evident diversification trend of both imported and exported products as its income per capita grows. Another developing country, China, was initially diversifying, then (after income per capita of approx. $4000) a very slight respecialization of exports began. Very slight U-shaped pattern is typical also for AUS, but clearly with a very different turning point than CHN. Changes in product diversity in

[30] An analogy can be found with another very famous U-curve in economics; the inverted-U 'Kuznets curve'. There are two different 'Kuznets curves' of this kind: the original one, depicting the relationship between income distribution and economic development; and so-called environmental Kuznets curve, where the environmental impact of production is related with economic development. One of the strong objection to the existence of the inverted-U shaped curve derived from the use of country fixed effects in the estimation. The inverted U-turn was first empirically questioned by Deininger and Squire (1998), who found that the use of country specific effects largely wiped-out the bell shape of the inequality-development relationship; after that, the use of country specific effects became unavoidable.

[31] Given unbalanced nature of our panel we have examined the cases for which we dispose of more than 18 observations within the years 1988-2010.

case of USA are negligible when compared to other countries, and only a slight increase in *RelT_exp* and *RelT_imp* after $40,000 took place - note eventual respecialization is very small in magnitude.

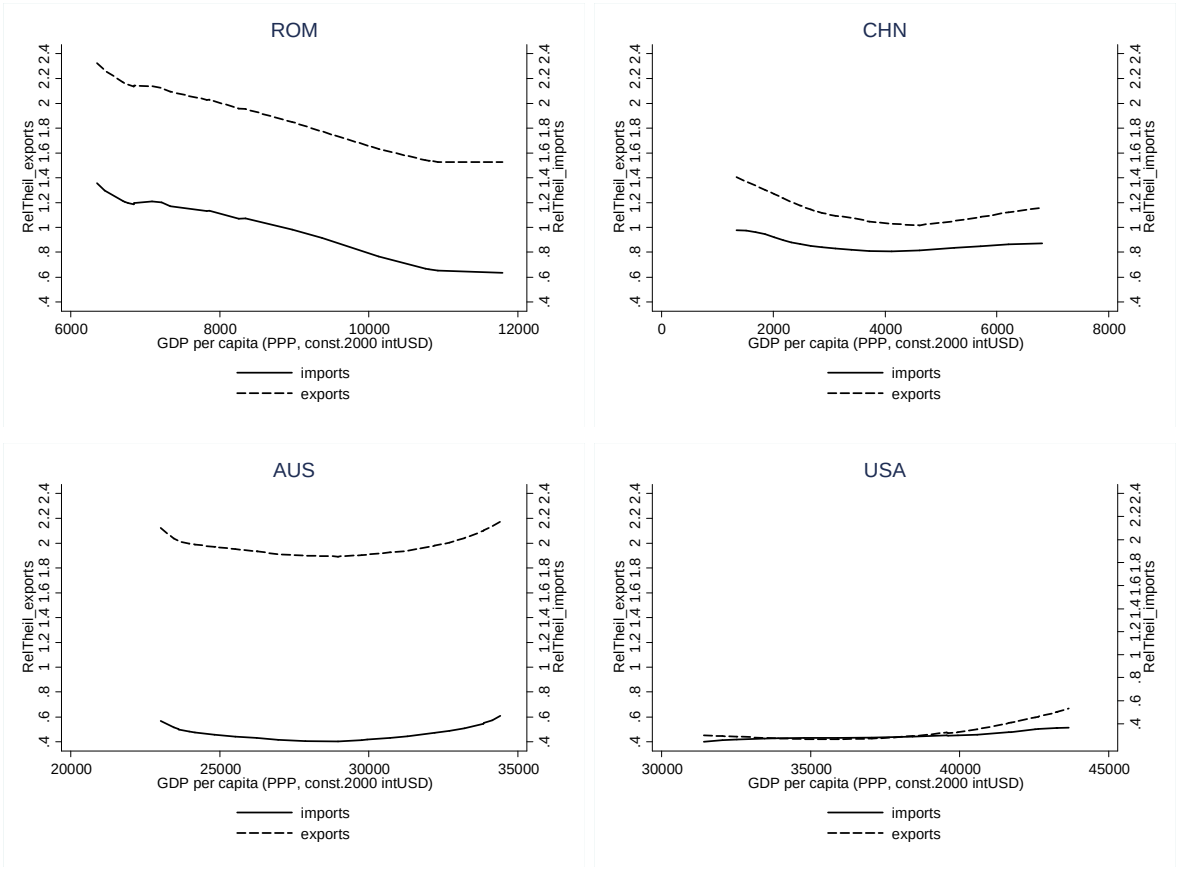

Figure 4. Degree of relative diversification of imported and exported products versus development level – selected country specific trends

Note: lowess - span=0.8. Left axis: relative Theil index of exports concentration; Right axis: relative Theil index of imports concentration.

Source: own elaboration based on trade data from UNComtrade, 2011 (HS0, 4963 product lines) and GDP per capita from WB WDI (2011).

4.3 Incorporating country fixed effects into general relative diversification curve - semiparametric results

Given just shown importance of country specificity, which must be taken into account when revealing general diversification curve (demonstrating 'typical' trend in the sample), we first consider the most natural extension of the unconditional lowess model: incorporation of country fixed effects into the estimation of nonparametric diversification curves (as in: de Benedictis et al., 2009 and Parteka, 2010). In order to do so, we apply semi-parametric (thus linking non parametric flexibility of lowess

with the inclusion of parametric components, in this case country dummies) Generalised Additive Model (GAM)[32] of the following form:

$$DIV\_imp_{it} = f(GDPpc_{it}) + \sum_{i}^{k} D_i \quad (6a)$$

and, similarly, in case of exports:

$$DIV\_\exp_{it} = f(GDPpc_{it}) + \sum_{i}^{k} D_i \quad (6b)$$

where D_i denotes country dummies and *f(.)* is an unspecified function estimated through the use of the lowess smoother, estimated from the data through a "backfitting" procedure[33] formally represented as Gauss-Seidel algorithm and consisting of iteratively smoothed partial residuals (see Buja et al., 1989 for formal details on the procedure). Nonlinear component of the GAM model embedded in function *f(.)* - in our case the relationship between synthetic inverse index of relative diversification and the development level - is represented in a graphical form as partial residuals plot. GAM model, being halfway between nonparametric methods (such as unconditional lowess) and multivariate linear regression, gives us the possibility of matching simultaneously the flexibility of the analysis (GDP per capita enters into a model in the form of an unspecified function) with the inclusion of additional determinants of the diversification curve (here: country-fixed effects).

In Figure 5 we show plots of partial residuals of GAM model obtained with alternative synthetic measures of relative imports and exports diversification. Both in case of imports and exports, semiparametric curves are decreasing (with only minor differences in shapes depending on the measure used) – an increase in income per capita goes hand in hand with a decrease in the indices of relative specialization. Hence, the general result is that when we correct unconditional nonparametric lowess curves for the presence of country fixed effects, trade diversification is the predominant tendency in our wide sample of countries, with only a negligible possibility of trade reconcentration at higher levels of development (more visible in case of imports). In other words: we do not exclude the possibility of re-specialization course in case of single countries but the trend based on predominant mass of country-year observations argues in favor of treating imports and exports relative diversification as a 'typical' tendency in the course of economic growth

---

[32] Generalised Additive Models (Hastie and Tibshirani, 1987, 1990) offer a generalisation of univariate smoothing techniques such as lowess to the multivariate situation with several predictors. In particular, with the use of GAM it is possible to estimate additive model of the form: $E(Y) = s_0 + s_1(X_1) + ... + s_p(X_p)$ where $s_i(X_i)$, $i=1, \ldots p$ denote smooth functions which shapes are unrestricted (so a mixture of nonparametric and parametric components is allowed).

[33] Backfitting procedure was introduced as an iterative fitting method within the framework of nonparametric multidimensional regression (Friedman and Stuetzle, 1981). It determines estimates for the covariates in a successive manner, using the currently available information from all covariates except the one for which estimates have just been computed.

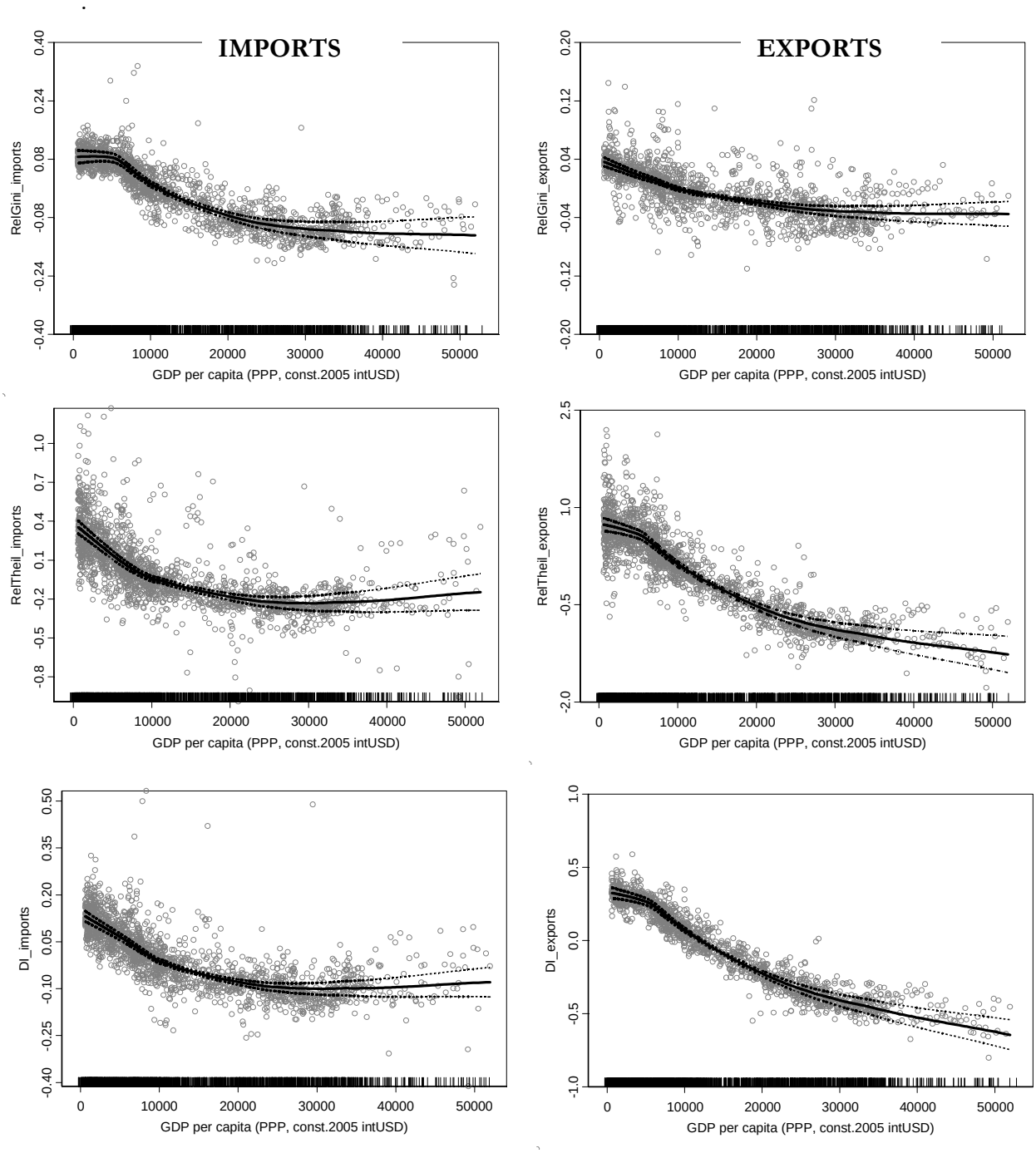

Figure 5. Degree of relative diversification of imported and exported products versus development level (left plots: imports, right plots: exports) – semiparametric curves controlling for country fixed effects, alternative indices

Source: own elaboration based on trade data from UNComtrade, 2011 (HS0, 4963 product lines) and GDP per capita from WB WDI (2011).

Note: yaxis: inverse diversification measure - partial residuals of GAM model (eq. 6a and 6b), span=0.5; sample without outliers defined as observations below 1st or above 99th percentile (number of obs.=1828). Due to FE inclusion only countries with multiple observations retained.

4.4 Parametric results – country specific factors and relative diversification process

In order to provide further confirmation for the robustness of our finding of progressing imports and exports relative diversification in the process of economic growth, in the following step we confront revealed diversification curves obtained through GAM, with parametric estimates of the following model:

$$\ln DIV\_imp_{it} = \alpha + \beta(\ln GDPpc_{it}) + D_i + D_t + \nu_{it} \quad (7a)$$

and, similarly, in case of exports:

$$\ln DIV\_\exp_{it} = \alpha + \beta(\ln GDPpc_{it}) + D_i + D_t + \nu_{it} \quad (7b)$$

where *DIV* (one of our alternative inverse indices of imports/exports diversification) and *GDPpc* (income per capita) are introduced in natural logarithms in order to mimic the shape of semiparametric diversification curves. D_i stands for country specific fixed effects and, additionally (in alternative formulations) time dummies D_t are included to control for common business cycle effects (for example affecting world prices of imported/exported products).

Table 6 reports results of FE estimates[34] (in lower panel - to be compared with OLS ones reported in the upper panel), referring to relative diversification imports. Table 7 reports analogous results referring to exports. Alternative models employ the use of different inverse indices of diversification; specifications (4) to (5) include time effects. In both cases – imports and exports – estimated β coefficient is negative and statistically significant which confirms the trend visible in relative diversification curves revealed through semiparametric estimation.

In order to prove robustness of the result, we have checked if it holds in a subsample of countries – eliminating from eq. 7a and eq.7b microstates (countries with population below 1mln) and petrol abundant countries (with petrol exports above 50% of total exports) – results are reported, respectively, in columns (2) and (3) of Table 8 (referring to imports) and Table 9 (exports)[35].

[34] Note that a drop in the number of countries in FE models is due to the unbalanced nature of our panel – only countries with multiple country-year observations could have been included into the model. Outliers are excluded from all of the specifications.

[35] Here we employ Relative Theil index; results obtained with alternative measures are obtainable upon request (no significant change in the conclusions drawn).

Table 6. Regression estimation results – relative *imports* diversification and income per capita

	Dependent variable: inverse index of relative imports diversification					
	(1)	(2)	(3)	(4)	(5)	(6)
	ln_RelT_imp$_{it}$	*ln_RelG_imp*$_{it}$	*ln_DI_imp*$_{it}$	*ln_RelT_imp*$_{it}$	*ln_RelG_imp*$_{it}$	*ln_DI_imp*$_{it}$
	OLS	OLS	OLS	OLS	OLS	OLS
lnGDPpc	-0.359***	-0.129***	-0.172***	-0.358***	-0.129***	-0.171***
[PPP, const.2005 int.USD]	[-40.27]	[-37.89]	[-41.63]	[-39.75]	[-37.64]	[-40.92]
No of observations	1828	1828	1828	1828	1828	1828
adj R2	0.47	0.43	0.48	0.48	0.45	0.49
year effect	no	no	no	yes	yes	yes
country effect	no	no	no	no	no	no
No. of countries	155	155	155	155	155	155
years	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010
	Dependent variable: inverse index of relative imports diversification					
	(1)	(2)	(3)	(4)	(5)	(6)
	ln_RelT_imp$_{it}$	*ln_RelG_imp*$_{it}$	*ln_DI_imp*$_{it}$	*ln_RelT_imp*$_{it}$	*ln_RelG_imp*$_{it}$	*ln_DI_imp*$_{it}$
	FE	FE	FE	FE	FE	FE
lnGDPpc	-0.238***	-0.024	-0.123***	-0.382***	-0.077**	-0.150***
[PPP, const.2005 int.USD]	[-4.78]	[-1.15]	[-6.26]	[-5.05]	[-2.54]	[-4.96]
No of observations	1820	1821	1822	1820	1821	1822
adj R2	0.47	0.43	0.48	0.47	0.43	0.48
year effect	no	no	no	yes	yes	yes
country effect	yes	yes	yes	yes	yes	yes
No. of countries	149	149	149	149	149	149
years	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010

Note. *, ** and *** denote significance at 10%, 5% and 1% level, respectively. Sample without outliers defined as observations below 1st or above 99th percentile. In FE estimates only countries with multiple observations retained. Robust t-statistics in parenthesis under coefficients. Constant included-not reported.
Source: own elaboration based on trade data from UNComtrade, 2011 (HS0, 4963 product lines); GDP per capita from WB WDI (2011).

Table 7. Regression estimation results – relative *exports* diversification and income per capita

	Dependent variable: inverse index of relative exports diversification					
	(1)	(2)	(3)	(4)	(5)	(6)
	ln_RelT_exp$_{it}$	*ln_RelG_exp*$_{it}$	*ln_DI_exp*$_{it}$	*ln_RelT_exp*$_{it}$	*ln_RelG_exp*$_{it}$	*ln_DI_exp*$_{it}$
	OLS	OLS	OLS	OLS	OLS	OLS
lnGDPpc	-0.339***	-0.074***	-0.123***	-0.338***	-0.075***	-0.123***
[PPP, const.2005 int.USD]	[-36.57]	[-26.32]	[-33.05]	[-35.78]	[-25.73]	[-32.30]
No of observations	1828	1831	1833	1828	1831	1833
adj R2	0.41	0.29	0.37	0.42	0.3	0.38
year effect	no	no	no	yes	yes	yes
country effect	no	no	no	no	no	no
No. of countries	155	155	155	155	155	155
years	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010
	Dependent variable: inverse index of relative exports diversification					
	(1)	(2)	(3)	(4)	(5)	(6)
	ln_RelT_exp$_{it}$	*ln_RelG_exp*$_{it}$	*ln_DI_exp*$_{it}$	*ln_RelT_exp*$_{it}$	*ln_RelG_exp*$_{it}$	*ln_DI_exp*$_{it}$
	FE	FE	FE	FE	FE	FE
lnGDPpc	-0.236***	-0.047***	-0.119***	-0.152**	-0.029*	-0.061***
[PPP, const.2005 int.USD]	[-5.56]	[-3.75]	[-6.66]	[-2.57]	[-1.80]	[-2.82]
No of observations	1822	1825	1827	1822	1825	1827
adj R2	0.41	0.3	0.38	0.39	0.25	0.34
year effect	no	no	no	yes	yes	yes
country effect	yes	yes	yes	yes	yes	yes
No. of countries	150	151	151	150	151	151
years	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010

Note. *, ** and *** denote significance at 10%, 5% and 1% level, respectively. Sample without outliers defined as observations below 1st or above 99th percentile. In FE estimates only countries with multiple observations retained. Robust t-statistics in parenthesis under coefficients. Constant included-not reported.
Source: own elaboration based on trade data from UNComtrade, 2011 (HS0, 4963 product lines); GDP per capita from WB WDI (2011).

Additionally, as reported in columns (4) to (7) of Table 8 and Table 9, we have also checked whether trade diversification course in the development process holds once other country-specific characteristic are included into the model. We estimated the extended equation of the following form:

$$\ln DIV\_imp_{it} = \alpha + \beta(\ln GDPpc_{it}) + \sum_{k=1}^{K} \delta_k X_{ik(t)} + D_t + \nu_{it} \quad (8a)$$

and, similarly, in case of exports:

$$\ln DIV\_\exp_{it} = \alpha + \beta(\ln GDPpc_{it}) + \sum_{k=1}^{K} \delta_k X_{ik(t)} + D_t + \nu_{it} \quad (8b)$$

where $X_{ik(t)}$ denotes set of *k* country-specific factors (time variant such as GDP or time invariant, such as geographical features). In line with the empirical literature (see Cadot et al., 2012, section 4) and suggestions of the theoretical models we consider: country size - measured alternatively in terms of gross output - *GDP* and population - *POP* (bigger countries should have more diversified economic structures), *DISTANCE* being the distance from the main markets (directly affecting foreign market access) and *PETROL* - the importance of petrol in countries' economic structure (expecting it to be positively linked to trade concentration, at least on the side of exports). Note that in this case we do not include country fixed effects as they wipe out most of the cross-country variability and, additionally, are collinear with time invariant characteristics, such as market distance.

All of the coefficients have the expected signs and the general conclusion holds even when additional controls are included into the model: there is a negative relationship between income per capita and inverse measure of diversification, thus as countries move towards higher stages of development, they tend to diversify both their import and export structures. Increase in domestic market size fosters the diversification of trade activity (note negative and significant relationship between size measures and inverse index of both imports and exports diversification). Additionally, geography and endowments play a role: countries further away from the main markets (which means higher cost of trade), as well as petrol abundant ones, have more concentrated trade baskets.

Table 8. Regression estimation results – relative *imports* diversification versus income per capita and additional country specific factors

Dependent variable: inverse index of exports diversification (ln_RelTheil_imp)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Country sample#	All	No microstates	No Petrol abundant	All+controls	All+controls	All+controls	All+controls
LnGDPpc	-0.358***	-0.390***	-0.372***	-0.198***	-0.393***	-0.213***	-0.411***
[PPP, const.2005 int.USD]	[0.009]	[0.008]	[0.009]	[0.009]	[0.007]	[0.011]	[0.008]
LnGDP				-0.152***		-0.116***	
[PPP, const.2005 int.USD]				[0.004]		[0.005]	
LnPOP					-0.149***		-0.161***
[1000]					[0.004]		[0.005]
LnDISTANCE						0.093***	
[km]						[0.011]	
LnPETROL							0.007***
[%total exports]							[0.001]
No of observations	1828	1627	1706	1821	1825	1613	1825
adj R2	0.47	0.6	0.52	0.71	0.69	0.73	0.70
year dummy	yes	yes	yes	yes	yes	yes	yes
No. of countries	157	133	146	155	156	128	156
years	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010

Note. *, ** and *** denote significance at 10%, 5% and 1% level, respectively. OLS log-log estimates, #sample without outliers defined as observations below 1st or above 99th percentile. Model (2) – without countries with population below 1mln (microstates). Model (3) – without countries with fuel exports above 50% of total exports. Robust standard errors in parenthesis under coefficients.

Source: own elaboration based on trade data from UNComtrade, 2011 (HS0, 4963 product lines); GDP per capita, GDP, population and fuel exports from WB WDI (2011), Distance (km from the closest major market: New York, Rotterdam or Tokyo) from Gallup et al.(1999).

Table 9. Regression estimation results – relative *exports* diversification versus income per capita and additional country specific factors

Dependent variable: inverse index of exports diversification (ln_RelTheil_exp)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Country sample#	All	No microstates	No Petrol abundant	All+controls	All+controls	All+controls	All+controls
LnGDPpc	-0.338***	-0.371***	-0.357***	-0.141***	-0.389***	-0.093***	-0.396***
[PPP, const.2005 int.USD]	[0.009]	[0.009]	[0.010]	[0.007]	[0.006]	[0.007]	[0.007]
LnGDP				-0.192***		-0.159***	
[PPP, const.2005 int.USD]				[0.004]		[0.004]	
LnPOP					-0.194***		-0.198***
[1000]					[0.004]		[0.004]
LnDISTANCE						0.203***	
[km]						[0.008]	
LnPETROL							0.003**
[%total exports]							[0.001]
No of observations	1828	1632	1705	1820	1825	1618	1825
adj R2	0.41	0.53	0.43	0.76	0.76	0.83	0.76
year dummy	yes	yes	yes	yes	yes	yes	yes
No. of countries	155	132	145	155	155	127	155
years	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010	1988-2010

Note. *, ** and *** denote significance at 10%, 5% and 1% level, respectively. OLS log-log estimates, #sample without outliers defined as observations below 1st or above 99th percentile. Model (2) – without countries with population below 1mln (microstates). Model (3) – without countries with petrol exports above 50% of total exports. Robust standard errors in parenthesis under coefficients.

Source: own elaboration based on trade data from UNComtrade, 2011 (HS0, 4963 product lines); GDP per capita, GDP, population from WB WDI (2011), Distance (km from the closest major market: New York, Rotterdam or Tokyo) from Gallup et al.(1999).

5. Conclusions

Main aim of the research was to contribute to empirical trade diversification literature by adding a missing piece – i.e. imports - to already analyzed export trends. In particular, we compare the evolution of imports and exports diversification in the process of economic development assessed with respect to changes in world composition of trade. Consequently, our study presents the natural extension of literature on relative trade diversification (de Benedictis et al., 2009; Parteka, 2010) in the sense that: *(i)* apart from providing evidence for exports we analyze also relative diversification of imports, *(ii)* we perform our analysis with highly disaggregated product level (and not sector level) data *(iii)* not limited to manufacturing only and *(iv)* for much larger panel of international economies than it used to be done.

We merge country-level income per capita data with a wide set of alternative relative diversification indices calculated with highly disaggregated import and export statistics (4963 product lines) for 163 countries across time period 1988-2010. To the best of our knowledge, this is the first attempt to explore *jointly* patterns of diversification of imported and exported goods in the context of countries' development process. We first provide descriptive evidence on product diversity visible in baskets of imported and exported goods, and later on pass to the empirical estimation of relative diversification curve.

We find that diversification of imports and exports differ in levels, but not in terms of evolution along the economic growth path. As far as differences in levels are concerned, since the very beginning of the development process, product heterogeneity of imports is much higher than that of exports. Set of goods imported by low income countries is already three times more heterogeneous than the set of products they export – they cover in their import portfolio already 63% of products imported at the world level while they export only 21% of all goods exported internationally. Consequently, at low levels of development there is room for activity at extensive margin of exports and the process of adding new lines to the export portfolio is much more dynamic than in case of imports, being relatively diversified even at lower levels of income per capita.

In terms of evolution along the economic growth path, we find robust tendency towards progressing relative diversification of both import and export structures. Even though slight re-specialization course is possible in case of some specific countries, predominant tendency revealed from the data is different: as income per capita grows, countries' import and export structures tend to become less specialized (more diversified) with respect to the world structure of trade. This finding is not sensitive to changes in the methodological setting. We have addressed the robustness of our finding in several ways: employing alternative relative measures of diversification, different estimation methods (nonparametric, semiparametric and parametric) and controlling not only for country fixed effects, but

also for the importance of additional determinants of the diversification process other than income per capita level.

References

Acemoglu D. and Ventura J. (2002) The World Income Distribution. *Quarterly Journal of Economics*, 117: 659-694.

Aghion P., Howitt P. (1999), *Endogenous Growth Theory*, MIT Press

Agosin M.R., Alvarez R., Bravo-Ortega C. (2011). Determinants of export diversification around the world: 1962-2000. *The World Economy* (forthcoming). DOI: 10.1111/j.1467-9701.2011.01395.x

Amiti M. (1999). Specialisation patterns in Europe. *Review of World Economics,* 135(4): 573-593.

Amiti M. and Konings J. (2007). Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia. *American Economic Review*, 97(5): 1611-1638.

Armington P. (1969). A theory of demand of for products distinguished by place of production. IMF Staff Papers, XVI: 159-178.

Besedes T. and Prusa T. (2006). Ins, outs and the duration of trade. *Canadian Journal of Economics*, 39: 266-295.

Bickenbach F., Bode E. and Krieger-Boden C. (2010). Closing the Gap between Absolute and Relative Measures of Localization, Concentration or Specialization. Kiel Working Paper No.1660.

Broda C., Weinstein D.E. (2006). Globalization and the gains from variety, *Quarterly Journal of Economics*, 121(2): 541-585.

Buja A., Hastie T., Tibshirani R. (1989). Linear smoothers and additive models (with discussion). *Annals of Statistics*, 17, 453-555.

Cadot O., Carrere C., Kukenova M., Strauss-Kahn V. (2010). OECD Imports. Diversification and quality search. The World Bank Policy Research Working Paper No.5258.

Cadot O., Carrere C., Strauss-Kahn V. (2011a). Export diversification: What's behind the hump? *The Review of Economics and Statistics*, 93(2): 590-605.

Cadot O., Carrere C., Strauss-Kahn V. (2011b). Trade diversification: drivers and impacts. [in:] *Trade and Employment. From myths to facts.* Jansen H., Peters R., Salazar-Xirinachs J.M. (eds.), ILO, Geneva, 253-307.

Cadot O., Carrere C., Strauss-Kahn V. (2012). Trade diversification, income, and growth: what do we know? *Journal of Economic Surveys* (forthcoming).

Chaney T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review*, 98(8): 1707-1721

Cleveland W. (1979). Robust locally weighted regression and soothing scaterplots. *Journal of American Statistic Association*, 74(368): 829-836.

Cowell F. (1995). *Measuring inequality.* Prentice Hall/Harvester Wheatsheaf, London (2nd edition).

de Benedictis L., Gallegati M., Tamberi M., (2009). Overall trade specialization and economic development: countries diversify, *Review of World Economics*, 145(1): 37-55.

Deininger and Squire (1998). New ways of looking at old issues: inequality and growth, *Journal of Development Economics*, 57: 259-287

Desmet K., Parente S. (2009). The Evolution of Markets and the Revolution of Industry: A Quantitative Model of England's Development, 1300-2000, CEPR Discussion Papers No. 7290

Dixit A. K., Stiglitz J. E. (1977). Monopolistic Competition and Optimum Product Diversity. *American Economic Review*, 67(3): 297-308.

Dutt P., Mihiov I., van Zandt T. (2011). Does WTO matter for the extensive and the intensive margins of trade? INSED Faculty and Research Working Papers No. 2011/47/EPS

Fernandes A. (2007). Trade policy, trade volumes and plant-level productivity in Colombian manufacturing industries, *Journal of International Economics,* 71(1): 52-71.

Friedman, J.H., Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association,* 76, 817-823.

Gallup J.L., J. Sachs and A.D. Melinger (1999). Geography and Economic Development. *International Regional Science Review,* 22 (2): 179-232; database available at www.cid.harvard.edu/ciddata

Goldberg, P.K.; Khandelwal, A.; Pavcnik, N.; Topalova, P. (2010). Imported intermediate inputs and domestic product growth: Evidence from India. *Quarterly Journal of Economics*, 124(4): 1727–1767.

Grossman and Helpman (1990). Comparative advantage and long run growth. *American Economic Review*, 80: 796-815.

Grossman and Helpman (1991). *Innovation and growth in the global economy*. Cambridge: MIT Press.

Hastie, T.J., Tibshirani, R.J. (1987). Generalized Additive Models. Some Applications. *Journal of American Statistical Association*, 82, 371-386.

Hastie T.J. and R.J. Tibshirani (1990). *Generalised Additive Models.* Chapman and Hall, London.

Helpman E., Krugman P. (1985). *Trade policy and market structure*. MIT Press

Hummels D. and Klenow P. (2002). The variety and quality of a nation's trade. NBER Working Paper No. 8712.

Hummels D. and Klenow P. (2005). The variety and quality of a nation's exports. *American Economic Review,* 94(1): 704-723.

Hummels D., Lugovskyy V. (2009). International Pricing in a Generalized Model of Ideal Variety, *Journal of Money, Credit and Banking*, 44(s1): 3–33,

Imbs, J., Wacziarg, R. (2003). Stages of diversification, *American Economic Review*, 93(1): 63-86.

Klinger B.; Lederman D. (2006). Diversification, innovation, and imitation inside the global technology frontier, World Bank Policy Research Working Paper No. 3872.

Koren, M.; Tenreyro, S. (2007). Volatility and development, *Quarterly Journal of Economics*, 122(1): 243-287.

Krugman P. (1979a). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics*, 9(4): 469-479.

Krugman P. (1979b). A model of innovation, technology transfer, and the world distribution of income, *Journal of Political Economy*, 87(2): 253-266.

Krugman, P.R. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *American Economic Review***,** 70 (5): 950-959.

Krugman, P.R. (1981). Intraindustry Specialization and the Gains from Trade, *Journal of Political Economy***,** 89(5): 959-973.

Kuznets S. (1962). The share and structure of consumption. *Economic Development and Cultural Change*, 10(2): 1-92

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

Palan N. (2010). Measurement of specialisation - the choice of indices, FIW working paper No.62.

Parteka A. (2010) Employment and export specialisation along the development path: some robust evidence**.** *Review of World Economics,* 145(4): 615-640.

Parteka A. and M. Tamberi (2011). Export Diversification and Development - Empirical Assessment. Universita' Politecnica delle Marche, Dipartimento di Scienze Economiche e Sociali Working Papers No. 359.

Pavcnik (2002). Trade liberalization, exit, and productivity improvement: Evidence from Chilean plants, *Review of Economic Studies*, 69(1): 245-276.

Romer, P. (1994). New Goods, Old Theory, and the Welfare Costs of Trade Restrictions, *Journal of Development Economics***,** 43: 5-38.

Appendix

Table 1A. List of countries in the dataset

Country code	Country Name	Income group*	GDP per capita [PPP, const.2005 int.USD]**	Years covered by the analysis	
				First	Last
BDI	Burundi	LI	403.02	1993	2010
BEN	Benin	LI	1321.28	1999	2006
BFA	BurkinaFaso	LI	967.256	1995	2010
BGD	Bangladesh	LI	929.165	1989	2006
CAF	CentralAfricanRepublic	LI	738.693	1993	2006
COM	Comoros	LI	1110.42	1995	1995
ETH	Ethiopia	LI	715.259	1995	2010
GIN	Guinea	LI	899.902	1995	2006
GMB	Gambia,The	LI	1112.19	1995	2001
KEN	Kenya	LI	1342.44	1992	2010
KGZ	KyrgyzRepublic	LI	1665.18	1995	2010
KHM	Cambodia	LI	1712.71	2004	2010
MDG	Madagascar	LI	892.184	1990	2010
MLI	Mali	LI	906.449	2004	2010
MOZ	Mozambique	LI	625.294	2001	2006
MWI	Malawi	LI	656.596	1990	2010
NER	Niger	LI	622.345	1995	2010
NPL	Nepal	LI	942.249	1994	2010
RWA	Rwanda	LI	911.698	2005	2009
TCD	Chad	LI	775.475	1995	1995
TGO	Togo	LI	834.185	1994	2005
TZA	Tanzania	LI	994.958	1995	2010
UGA	Uganda	LI	879.5	1994	2010
ZWE	Zimbabwe	LI	169.291	1995	2010
ARM	Armenia	LMI	3749.78	1997	2010
BLZ	Belize	LMI	5380.63	1992	2010
BOL	Bolivia	LMI	3668.07	1992	2010
BTN	Bhutan	LMI	3501.01	1993	2010
CIV	Coted'Ivoire	LMI	1691.27	1995	2009
CMR	Cameroon	LMI	1877.97	1995	2006
COG	Congo,Rep.	LMI	3243.59	1993	1995
CPV	CapeVerde	LMI	2637.41	1997	2010
DJI	Djibouti	LMI	2087.17	2009	2009
EGY	Egypt,ArabRep.	LMI	4658.63	1994	2010
FJI	Fiji	LMI	4220.25	2002	2009
GEO	Georgia	LMI	3425.32	1999	2010
GHA	Ghana	LMI	1310.47	2005	2010
GTM	Guatemala	LMI	3984.07	1993	2010
GUY	Guyana	LMI	2447.61	1991	2010
HND	Honduras	LMI	3092.84	1994	2009
IDN	Indonesia	LMI	2475.73	1989	2001
IND	India	LMI	1785.88	1988	2010
KIR	Kiribati	LMI	2119.82	1995	2005
LKA	SriLanka	LMI	3101.67	1990	2010
MAR	Morocco	LMI	3145.68	1993	2006
MDA	Moldova	LMI	2173.6	1994	2010
MNG	Mongolia	LMI	2884.68	2003	2007
NGA	Nigeria	LMI	1725.66	1999	2010
NIC	Nicaragua	LMI	2149.23	1993	2010

PAK	Pakistan	LMI	2203.22	2003	2010
PHL	Philippines	LMI	2707.4	2000	2001
PNG	PapuaNewGuinea	LMI	1858.78	2001	2004
PRY	Paraguay	LMI	4090.58	1989	2010
SDN	Sudan	LMI	1532.39	1995	2009
SEN	Senegal	LMI	1584.04	1996	2010
SLV	ElSalvador	LMI	5425.54	1994	2010
STP	SaoTomeandPrincipe	LMI	1018.62	1999	2000
SWZ	Swaziland	LMI	4091.36	2000	2001
SYR	SyrianArabRepublic	LMI	4199.88	2001	2008
TKM	Turkmenistan	LMI	2151.75	1999	2000
TMP	Timor-Leste	LMI	692.021	2004	2005
TON	Tonga	LMI	4150.35	2008	2010
UKR	Ukraine	LMI	4077.3	2001	2001
VNM	Vietnam	LMI	2157.9	2000	2009
WSM	Samoa	LMI	3601.49	2001	2010
YEM	Yemen,Rep.	LMI	2251.41	2006	2009
ZMB	Zambia	LMI	1148.38	1995	2010
ALB	Albania	UMI	5413.32	1996	2010
ARG	Argentina	UMI	10838.6	1993	2010
ATG	AntiguaandBarbuda	UMI	13932.7	1999	1999
AZE	Azerbaijan	UMI	5446.31	1999	2010
BGR	Bulgaria	UMI	8857.79	1996	2010
BIH	BosniaandHerzegovina	UMI	6638.91	2003	2010
BLR	Belarus	UMI	8244.79	1998	2010
BRA	Brazil	UMI	8065.28	1989	2010
BWA	Botswana	UMI	9624.23	2000	2001
CHL	Chile	UMI	10505.7	1990	2010
CHN	China	UMI	3359.38	1992	2010
COL	Colombia	UMI	7052.69	1991	2010
CRI	CostaRica	UMI	8610.85	1994	2010
CUB	Cuba	UMI	9370.14	2001	2006
DMA	Dominica	UMI	7278.84	1993	2010
DOM	DominicanRepublic	UMI	7367.72	2001	2010
DZA	Algeria	UMI	6442.37	1992	2010
ECU	Ecuador	UMI	6065.84	1991	2010
GAB	Gabon	UMI	13917.1	1993	2006
GRD	Grenada	UMI	5841	1993	2001
IRN	Iran,IslamicRep.	UMI	8140.25	1997	2006
JAM	Jamaica	UMI	6777.94	1991	2010
JOR	Jordan	UMI	4141.21	1994	2010
KAZ	Kazakhstan	UMI	7563.11	1998	2009
KNA	St.KittsandNevis	UMI	11084.1	1993	2001
LBN	Lebanon	UMI	9868.63	1997	2010
LCA	St.Lucia	UMI	7823.73	1992	2001
LTU	Lithuania	UMI	12000.1	1994	2010
LVA	Latvia	UMI	10659.3	1994	2010
MDV	Maldives	UMI	3729.26	1995	2008
MEX	Mexico	UMI	11286.3	1990	2010
MKD	Macedonia,FYR	UMI	7185.25	1994	2008
MUS	Mauritius	UMI	9374.96	1993	2010
MYS	Malaysia	UMI	9673.83	1989	2010
NAM	Namibia	UMI	5053.94	2000	2008
PAN	Panama	UMI	8685.63	1995	2008

PER	Peru	UMI	6126.72	1992	2010
ROM	Romania	UMI	8257.19	1989	2010
RUS	RussianFederation	UMI	10663.5	1997	2009
SUR	Suriname	UMI	4876.47	1994	2010
SYC	Seychelles	UMI	16674.2	1994	2002
THA	Thailand	UMI	5655.72	1988	2010
TUN	Tunisia	UMI	5753.52	1991	2010
TUR	Turkey	UMI	9886.54	1989	2010
URY	Uruguay	UMI	9731.02	1994	2009
VCT	St.VincentandtheGrenadines	UMI	6019.17	1993	2010
VEN	Venezuela,RB	UMI	10001.5	1994	2006
ZAF	SouthAfrica	UMI	8177.64	1992	2010
ARE	UnitedArabEmirates	HI-OECD	46482.5	1991	2008
BHR	Bahrain	HI-OECD	26071.2	2002	2010
BHS	Bahamas,The	HI-OECD	24975	1997	2009
BMU	Bermuda	HI-OECD	35062.4	1995	1995
BRB	Barbados	HI-OECD	17973.7	2000	2006
BRN	BruneiDarussalam	HI-OECD	50294.4	1992	2004
CYP	Cyprus	HI-OECD	21986.6	1989	2010
HKG	HongKongSAR,China	HI-OECD	32559	1993	2010
HRV	Croatia	HI-OECD	13256.2	1992	2010
KWT	Kuwait	HI-OECD	49411.3	2006	2006
MAC	MacaoSAR,China	HI-OECD	32125.3	1991	2010
MLT	Malta	HI-OECD	19966.7	1994	2010
OMN	Oman	HI-OECD	18298.1	1989	2007
QAT	Qatar	HI-OECD	67156.3	2000	2006
SAU	SaudiArabia	HI-OECD	20044.4	1991	2010
SGP	Singapore	HI-OECD	37397.2	1989	2010
TTO	TrinidadandTobago	HI-OECD	16309	1991	2009
AUS	Australia	HI-nonOECD	28625.7	1988	2010
AUT	Austria	HI-nonOECD	31864.6	1994	2010
BEL	Belgium	HI-nonOECD	31735.3	1999	2010
CAN	Canada	HI-nonOECD	31088.3	1989	2010
CHE	Switzerland	HI-nonOECD	34228.6	1988	2010
CZE	CzechRepublic	HI-nonOECD	18435.4	1993	2010
DEU	Germany	HI-nonOECD	29069.6	1988	2010
DNK	Denmark	HI-nonOECD	30139.5	1989	2010
ESP	Spain	HI-nonOECD	24020.7	1989	2010
EST	Estonia	HI-nonOECD	13504.8	1995	2010
FIN	Finland	HI-nonOECD	26349.9	1988	2010
FRA	France	HI-nonOECD	28154.1	1994	2010
GBR	UnitedKingdom	HI-nonOECD	29747.5	1993	2010
GRC	Greece	HI-nonOECD	21101.5	1988	2010
HUN	Hungary	HI-nonOECD	13499.7	1992	2007
IRL	Ireland	HI-nonOECD	30952.2	1992	2010
ISL	Iceland	HI-nonOECD	29545.1	1988	2010
ISR	Israel	HI-nonOECD	23044	1995	2010
ITA	Italy	HI-nonOECD	27161.6	1994	2010
JPN	Japan	HI-nonOECD	28353.2	1988	2010
KOR	Korea,Rep.	HI-nonOECD	17867.3	1988	2009
LUX	Luxembourg	HI-nonOECD	66777.1	1999	2010
NLD	Netherlands	HI-nonOECD	31695.1	1992	2006
NOR	Norway	HI-nonOECD	43558.9	1993	2010
NZL	NewZealand	HI-nonOECD	22381.9	1989	2010

POL	Poland	HI-nonOECD	12676.2	1994	2010
PRT	Portugal	HI-nonOECD	19329.4	1988	2010
SVK	SlovakRepublic	HI-nonOECD	14937.6	1994	2010
SVN	Slovenia	HI-nonOECD	21107.5	1994	2010
SWE	Sweden	HI-nonOECD	29122.1	1992	2010
USA	UnitedStates	HI-nonOECD	38398.1	1991	2010

Note:
Countries ordered alphabetically within income groups
*country groups according to the World Bank's (2011) classification: LI- low income, LMI - lower middle income, UMI - upper middle income, HI- high income (OECD and non-OECD members)
**year average

Data Appendix

Trade data

Harmonized System is an international nomenclature for the classification of products which allows participating countries to classify traded goods on a common basis for customs purposes. For the purpose of international comparisons, HS is a six-digit code system. Currently the HS comprises approximately 5,300 article/product descriptions that appear as headings and subheadings, arranged in 99 chapters, grouped in 21 sections. The six digits can be broken down into three parts. The first two digits identify the chapter the goods are classified in, the next two digits identify groupings within that chapter and the last two digits are even more specific. Up to the HS six digit level, all countries classify products in the same way. Deeper disaggregation series are not harmonized and thus are inappropriate for cross-country studies.

We run our analysis since the first year HS was launched (1988) until 2010. In this period there were three major revisions following the first release, corresponding to following data series: HS0 – 1988/92 (5016 products), HS1 – 1996 (5113 products), HS2 – 2002 (5224 products), HS3 – 2007 (5053 products). We combine data from these revisions, for each year taking the most accurate one (i.e. matching HS0 data for the years 1988-1995 with HS1 data for the years 1996-2001, and so on). In order to obtain consistent long time data series we use automatic conversion into HS0 product division (based on conversion tables from WITS).

Rough trade data reported by COMTRADE does not contain the information on every product line for every country-year. This might be a problem if one considers potential mismatch between inactive trade lines (zero trade value) and missing data on existing flows. To overcome this issue, following Cadot et al. (2011a) we fill-in the database, adding missing product lines and assign them zero trade values (we thus assume that unreported trade flow is zero flow). In case of imports this shall not be a major weakness as import statistics are believed to be much more complete and of a better quality than export data. However, we exclude from the sample 53 HS0 codes which correspond to never-traded ('silent') product lines. Within the years 1988-2010 the imports of these 53 goods were never reported as positive neither by any of the countries, nor at the aggregate world level. In the end, we are left with 4963 product lines and with these statistics we calculate product-wise diversification indices for each country and year.

HS system is very detailed for some sectors (such as textiles) and less detailed for other (such as machinery). However, this is not necessarily a problem, because the share of every HS section in total number of lines is largely correlated with its relative importance in world trade.

Some problem may be caused by the fact that not every product present in newer HS revisions can be associated with basic HS0 1988/92 product code (this difficulty is present especially in case of the latest revision: HS3-2007). On country-year basis we have calculated the share of trade value (in total) that would be missed due to non perfect matching between HS0 1988/92 and newer revisions. The average non-counted imports share is equal to 4% of total country imports (s.d =0.05, min=0, max=0.5). We exclude from the final sample these country-year pairs in which more than 15% of total country trade value would not be accounted for by HS0 product classification (corresponding to roughly 2% of all our country-year observations). Finally, we obtain a set of diversification indices and after matching them with income per capita data we are left with 1905 observations for 163 countries (unbalanced panel) with 11 average annual observations per country.

When we calculate the number of new product lines, on country-year basis we exclude from the analysis observations corresponding to the first year in which country's imports are reported in COMTRADE database (artificially this would result in a huge number of "new" products added to the country portfolio). Obviously, adopted definitions of 'new products" (3-year or 5-year moving window) mean also the truncation of observed period and total number of observations is lower than in case of synthetic measures of diversification such as Theil or Herfindahl index.

Additional data

GDP per capita data comes from World Bank's WDI and is expressed in PPP terms (const 2005 international $). The data is not available for all country-year observations on trade that we have, thus in 15 observations we complete WDI's data on income per capita with that from PWT (version 7.0, June 2011), also expressed in PPP(2005) terms.