On the evolution of comparative advantage: path-dependent versus path-defying changes

Nicola D. Coniglio [(1)], Davide Vurchio [(2)], Nicola Cantore [(3)], Michele Clara [(3)]

[(1)] University of Bari 'Aldo Moro' (Italy), [(2)]University of Rome 3 (Italy), Centro Manlio Rossi-Doria, [(3)] UNIDO (Austria)

This version: 2 March 2018

Abstract

The diversification of production and trade is considered almost unanimously a fundamental policy goal, particularly for developing economies whose export baskets are heavily concentrated on a few products. In what direction trade diversification ought to take place is, however, subject to fierce debate. The Product Space (PS) framework (Hausmann and Klinger, 2007; Hidalgo et al. 2007) is a recent contribution in the economic literature that has proved very influential in policy circles. It argues that the endowment of production capabilities (technologies, production factors, institutions etc.) determines what countries produce today but it also constrains what they can produce in the future as it is uncommon that countries develop a comparative advantage in goods that do not draw from the same pool of capabilities (unrelated products). Contributions along such line argue that defying the initial comparative advantage can be a risky policy decision with high probability of failure. The main objective of this contribution is to use a novel methodology to investigate whether the patterns of diversification of a sample of 177 countries over the period 1995-2015 conform or not to the prediction of the PS framework. We find evidence of a high degree of path-dependence but our analysis suggests also that a significant number of new products that entered countries' export baskets were unrelated to the initial productive specialization (path-defying changes). We shed light on the determinants of these 'radical' patterns of diversification and show they are associated with higher economic growth. The results of this study have important policy implications in particular for the design of industrial policies aimed at actively shaping countries' structural transformation.

Keywords: path-dependence, product space, trade diversification, industrial policy
JEL Code: F1, O1, O3

Contact details: nicoladaniele.coniglio@uniba.it; davide.vurchio87@gmail.com; n.cantore@unido.org; m.clara@unido.org;

Aknowledgements. We acknowledge the United Nation Industrial Development Organization (UNIDO) for financial support. We wish to thank participants at seminars at UNIDO Vienna, Ritsumeikan University, Nihon University and Nanzan University for useful comments and suggestions.

1. Introduction

Governments around the globe are understandably concerned with 'what' their countries produce and export. Understanding how comparative advantage evolves over time is crucial to single out its determinants and, on such basis, to inform effective policy measures. From the latter perspective, renewed interest in industrial policies has re-ignited the debate on the role of governments in shaping specialization and in driving structural transformation towards specific directions.

New 'intellectual tools' to ground industrial policies (as well as to highlight their likely limits) have been provided by the Product Space (PS) framework, which was developed in the seminal works by economists and physicists (Hausmann and Klinger, 2007; Hidalgo et al. 2007). These authors apply network analysis to international trade data and they postulate that the evolution of export baskets is strongly characterized by path dependence. According to this framework, economies can be represented as sets of productive capabilities (technologies, factors endowments, institutions, etc.) that are combined in different ways to produce different products.[1] Countries progressively diversify the structure of their exports towards new products that are related to the current capabilities, and hence the current production structure; relatedness between current production and potential new products lies in the common set of productive capabilities that are needed. On the contrary, the emergence in the export basket of unrelated products, i.e. products that require productive capabilities that are different from those already employed in current production, is more unlikely as it happens less frequently. This framework posits that acquiring a comparative advantage in radically different products – path-defying diversification– is a rare event.

An important policy implication stemming from these contributions is that policies that try to engineer diversification towards 'related goods' are more likely to be successful because countries already have a 'latent comparative advantage' in these products. The economic intuition behind such 'smart specialization' strategy is straightforward: products that are closely related in the PS require a similar set of production capabilities. If an economy has a comparative advantage in a given product, then it is relatively simple for that economy to develop – possibly, with a little help from Government – a comparative advantage in products requiring the same set of capabilities. On the contrary, industrial policies aimed at promoting unrelated products are likely to lead to policy failure because 'big jumps' in the PS (what we label in this study 'path-defying' diversification) require capabilities that are scarce and hard to create. Thus the PS analysis suggests that policymakers should follow a step-by-step approach targeting products for which countries may have a latent comparative advantage ('small jumps in the PS') and it therefore discourages as overoptimistic policy initiatives that defy their comparative advantage.

Although these ideas have largely circulated in policy circles and several countries are starting to employ lessons from the PS analysis to design their policies, there is no rigorous empirical test to

[1] Differently from previous studies, the PS employs an 'agnostic' approach on the sources of comparative advantage and relies on the following intuition: if we observe - using international trade data - that two products are produced and exported in tandem (i.e. the products are 'related') then it must be very likely that they use a similar/common set of production factors. On the contrary, products that are seldom co-exported (unrelated products) are unlikely to be produced using a similar set of production capabilities. This outcome-based measure of comparative advantage has the advantage of abstracting from the exact identification of these productive capabilities (i.e. the true roots of comparative advantage).

date, to the best of our knowledge, that shows whether the pattern of specialization of countries follows such predictions.

In this work we build on the novel methodology developed by Coniglio et al. (2018) to analyse if the actual patterns of diversification over time were governed by the notion of relatedness as developed by the PS framework.[2] In particular, using disaggregated international trade data we address the following research questions:

(i) Did the diversification of countries' export baskets follow a path-dependent pattern during the last 20 years?;

(ii) Under what circumstances countries can diversify towards areas of the Product Space that are unrelated with the initial production basket (radical changes)?;

(iii) Do countries that successfully diversified to unrelated products have a better economic performance?

Our contribution sheds light on the policy relevant features of a country that are associated with the ability to defy the initial comparative advantage and to specialize in new products that are unrelated with the current export basket.

In fact, new products that enter countries export baskets and that defy the hypothesis of path-dependence are probably the most interesting from a policy perspective. Path-defying changes often requires (and at the same time generate) breakthroughs that are the result of a new way to combine knowledge and capabilities. These changes have a strong potential for the development of new technological trajectories and developments (see Dosi, 1982 or - in the context of regional innovation - Castaldi et al, 2015).

The current debate has largely shifted from the 'if' governments should influence the type of products/exports (i.e. their current comparative advantage) to 'which' type of goods/services should be targeted by industrial policies[3]. The PS is more and more used as a 'map' for informing the direction of these active policies but its 'mechanical' application might lead to undesirable outcomes. In several circumstances, countries have indeed defeated their static comparative advantages. Notable cases are the rise of the aircraft industry in Brazil and the automotive or electronic industries in Korea (Lin and Chang, 2009). These 'new product entries' are at odd with the PS framework and hence caution should be exercised in narrowly drawing policy prescriptions from these important contributions.

In the first part of the analysis we build counterfactual country-specific distributions of relatedness between new products that enter export baskets with Revealed Comparative Advantage (RCA) *à la* Balassa larger than 1 and those products exported with RCA larger than 1 five years before. We test, using non-parametric techniques, the hypothesis that these new products are unconstrained by the degree of their proximity, as measured by the PS. Our result on this general test rejects such hypothesis thus confirming a significant influence of the capabilities accumulated in a country over the pattern of its structural transformation. Behind this general pattern, two important results emerge. First, 'only' half of the new products can be

[2] Coniglio *et al* (2018) develop a 'dart-board' approach which allows to test in a rigorous way the degree of path dependence in the evolution of countries/regions export baskets over time. The authors apply their non-parametric analysis to investigate the evolution of the export basket of Italian provinces (NUTS 3 classification) before (2002-2006) and during the global crisis (2007-2011). This study shows that although the overall evolution of the Italian export basket shows a significant degree of path-dependence – as predicted by the PS framework – more radical changes that defy the initial comparative advantage do often occur.

[3] See for instance Naudé, W. A. (2010).

considered as statistically path-dependent.[4] This finding suggests that unrelated changes are far from being rare. Secondly, we find a very large degree of cross-country heterogeneity in the degree of path dependence. While some countries experienced a process of diversification that has been strongly constrained by the 'capabilities' available in the previous five years (*constrained diversification*) others have been able to diversify away from their initial comparative advantage (*unconstrained diversification*).

In the second part of the analysis we shed light on the determinants of these differences across countries. We find evidence that path-dependence is less marked in advanced and larger economies with a high trade diversification in unrelated variety. On the contrary, countries with a large natural resource sector are less likely to diversify away from their current comparative advantage suggesting another important dimension of the 'resource curse'. We show that better business and institutional environments are associated with a higher share on path-defying new entries. In other words, countries with better institutions find it easier to diversify away from their current comparative advantage. In this respect our results seem to suggest that an arm's length approach in some policy dimensions – such as the direct intervention of governments in the credit market – is more likely to bear fruits in relatively rich countries where, as argued by Stiglitz (2002), market failures dominate government failures.

Finally, our results show that countries with a higher share of new products that are unrelated to their previous production basket – i.e. those countries showing a pattern of 'unconstrained diversification' – perform better that those that follow a stronger degree of path dependence. According to our preferred estimates a 10% increase in the share of path-defying changes increases the yearly average per capita GDP growth in the subsequent 5-years by 0.8%; a sizable contribution to countries' well-being.

Our paper is related to the recent contribution by Bahar et al (2017) that studies how the emergence of new export specializations (extensive margin of trade) and the growth of export values (intensive margin of trade) are affected by relatedness with the pre-existing export basket. The authors show that the probability of exporting with revealed comparative advantage a new product within the following decade increases on average between 80-140 percent when its relatedness with the pre-existing export basket, measured, as in our paper, *à la* Hausmann and Klinger (2006), is one standard deviation above the mean. Bahar et al (2017) shed new light on the mechanisms behind the agnostic concept of relatedness developed in this framework. They show, using data from 144 countries over the period 1984-2014, that the emergence of new products is mostly driven by the use of related technological capabilities and by the existence of downstream industries (backward linkages).

While their paper contributes to the literature on the determinants of dynamic comparative advantages by exploring alternative measures of relatedness – from the general and agnostic measure of Hausmann and Klinger (2007) to measures related to specific demand and supply channels – we analyse path-dependence in the evolution of comparative advantage, its main

[4] Coniglio et al (2018) find a stronger degree of path-dependence, approximately 70% of new products, when looking at sub-national (NUTS 3) Italian data. The difference is likely to be due to the more limited set of production capabilities available at a small geographical scale which implies a lower ability to make larger jumps over the production space.

determinants and its growth potential. Our methodology can be applied to alternative definition of relatedness as those proposed by these authors.[5]
Our study is also highly related to a recent paper by Pinheiro et al (2018) that address the evolution of comparative advantage over time of 93 countries between 1970 and 2010. These authors find that the larger the share of 'unrelated' new entries, the better countries economic performance. In order to measure relatedness between the existing export basket and new entries, these authors employ a novel concept of relative density: a new product is labelled as 'related' if its proximity with the current export basket is higher than the average proximity of all those goods that can potentially be produced (option set). Employing this methodology, the authors find that unrelated activities are 'rare' (only 7,2 % of new products). One drawback of this approach – which our methodology aims at correcting – is the inability to identify spurious (un)relatedness in the absence of a statistical counterfactual. We believe that our empirical approach is superior in this respect.

This paper is organized as follows. In *Section 2*, we survey contributions that focus on the determinants of trade diversification as well as selected studies related to the Product Space framework. In *Section 3*, we describe the data and the methodology used for testing the hypothesis of path-dependence. In *Section 4*, we employ econometric analysis to study the main determinants of path-dependence and shed light on those country level characteristics – including policy relevant variables – that affect the ability to introduce path-defying new products in the export basket. The final step of our analysis is reported in *Section 5* where we investigate the nexus between the share of path-defying changes and economic growth. *Section 6* concludes.

2. On the evolution of comparative advantage: a selective literature review

This study on the evolution of comparative advantage is related to the literature on export diversification. While studies on diversification generally look at the determinants of 'new entries', our analysis is interested in the relationship between 'old' and 'new entries' (i.e. the direction of trade diversification).
Several studies have emphasized the importance of trade diversification as a strategy for hedging against the risks of overspecialization (di Giovanni and Levchenko, 2009), for knowledge spillovers (Hausmann et al 2007) and as an engine of structural change and economic growth (Agosin 2009; Cadot et al 2013; Gozgor and Can 2017). In particular, the nexus between trade

[5] The paper by Bahar et al (2017) has the merit to be the first one which explores the black box of the concept of relatedness proposed in this literature. In fact, Hidalgo *et al* (2007) use an agnostic measure which rests on the theoretical idea that if any two products are co-exported with high probability by the same countries then they use a 'related' set of (undefined) production capabilities (technologies, factors endowment, institutions etc.). Bahar et al (2017) introduce five specific measures of relatedness in order to assess the relative importance of different channels (technological relatedness, labour force relatedness, linkages with suppliers and with costumers). One important limit of this approach is based on the way these specific relatedness measures are built. While generic measures of relatedness are built using world trade data, these specific measures due to data limitations are built using USA data with the underlying assumption that USA can be applied also to other countries. The potential bias in measurement is likely to be large as the sample includes many developing and emerging economies. For instance 'labour relatedness' is measured on the basis of US Current Population Survey on the "fraction of separating workers from each industry *i* that move to firms in each industry *j*". The assumption that this pattern of labour mobility applies to other countries is rather strong. Similarly, backward and forward linkages and technological relatedness are based on respectively US Input-Output tables and R&D and patent citation data.

diversification and growth has been the subject of scrutiny by several studies. One of the most comprehensive and rigorous analysis has been conducted by Mau (2016), who finds robust support to the hypothesis that trade diversification causes an increase in GDP per capita. The author also finds weak evidence of the reverse causality but GDP growth has more limited effects on trade diversification and, in addition, the effect takes time to emerge.

In what follows we firstly review some studies on the main determinants of trade diversification and then we discuss the evidence on path-dependence in the inclusion of new products in export baskets.

What determines trade diversification?

Income per capita

Several studies have investigated the relationship between trade diversification and economic development, as proxied by income per capita. In the seminal theoretical paper by Hausmann and Rodrik (2003), economic development is assimilated to a 'cost discovery' process where entrepreneurs develop new products or processes by means of risky and uncertain investment activities. Grossman and Helpman (1993) show that economic growth is associated to an increase in diversification through innovation and the development of new products. While most theoretical contributions suggest a positive association between the degree of diversification and economic development, empirical studies report contradicting results.

Imbs and Wacziarg (2003) and Cadot et al (2011) find evidence of an 'inverted-U' relationship where trade diversification rises up to a certain level of development and then declines.[6] According to these authors, when countries develop, market failures that limit the 'cost-discovery' process are gradually mitigated and this lead to higher diversification *via* an expansion of the extensive margin of trade (number of goods and services produced and exported). Progressively, as countries develop, a dynamic process of removal of less profitable varieties and specialization in goods and services where the comparative advantage is stronger leads to a lower degree of export diversification (re-specialization through growth at the intensive margin).

More recent evidence questions this conclusion. Using highly disaggregated trade data for 110 countries over the period 1998-2009, Mau (2016) finds evidence of a robust and positive correlation between income per capita and trade diversification and rejects the hypothesis of an inverse U-shaped relationship. The author, using dynamic panel estimates, shows that causality between income per capita and trade diversification runs both way but that the effect of trade diversification on countries' growth performance is stronger than the reverse effect.

De Benedictis et al (2009) use relative rather than absolute measures of trade diversification to reject the hypothesis of re-specialization as GDP per capita reaches a certain threshold. This result is further supported by Parteka and Tamberi (2013), which also deals with the potential endogeneity of GDP per capita and confirm a positive (but rather slow) effect of development on trade diversification.

Country size, human capital and geography

---

[6] Klinger and Lederman (2004) find evidence of a reduced pace of trade diversification for higher level of income per capita. As argued by Mau (2016) this finding might be explained by the fact that new products might become indistinguishable from old ones using standard trade data classifications.

Population size and the quality of human capital are channels through which a country can add new products to its production and export basket (Hausmann et al 2007; Parteka and Tamberi 2013). Using data on a large panel of countries in the period 1962-2000, Agosin et al (2012) find evidence of a positive effect of human capital accumulation on export diversification. Interestingly, these authors show that countries with a better endowment of human capital are more successful in expanding their trade baskets when positive terms-of-trade shock occurs. This result suggests that a better endowment of human capital allows countries to react more quickly and effectively to opportunities for diversification.

A recent study by Jetter and Ramirez-Hassan (2015)[7] finds that the two most important predictors of export diversification among 36 possible determinants that include political, macroeconomic, cultural and geographical factors are net enrolment in primary education and natural resource rents over GDP, respectively with positive and negative effects. Secondary and tertiary education are found to be less relevant compared to primary education. This result suggests that in the medium to long-term it is a larger base of the educational pyramid that matters the most for promoting a well-diversified economy.

Geography as well may matter as it affects the fixed as well as the variable costs of exporting a new product or service and, in turn, the number of export varieties.[8] Agosin et al (2012) and Parteka and Tamberi (2013) find that distance from global markets – measured with proxies of remoteness or distance from main world markets – reduces the degree of trade diversification. Also Basile et al (2017) using a spatial dynamic model show that proximity to large countries has a positive effect on trade diversification.

Trade policy and other policy or institutional determinants of trade diversification.

Access to foreign market is key to attain comparative advantage in new products and services. International trade costs are determined both by geography and by economic policies; technological advancements in the transport sector have pushed the balance more and more toward a greater relevance of artificial (policy-driven) barriers to trade. Mau (2016) finds evidence of an important effect of the dismantlement of trade barriers on the degree of diversification, in particular in developing countries. Dennis and Sheperd (2011) find a robust and positive effect of trade facilitation - i.e. the set of policy measures that reduce barriers to international trade, such as transaction and bureaucratic costs - on export diversification. These authors use as measures of trade facilitation Wold Bank data on export costs - the total official costs associated to the shipping of a standardized container - and on market entry costs[9] - the costs of starting a business (Doing Business database). Using data on a large cross section of countries in the period 1991-2003, Feenstra and Ma (2014) show that a 10% bilateral

---

[7]Jetter and Ramirez-Hassan (2015) use a Bayesian Model Averaging approach which consists in averaging all the possible combinations of 36 covariates (that is 2^{36} possible model combinations) in order to infer which of them are consistently good predictors of the dependent variable. This methodology allows to overcome the drawback of model uncertainty which is common in this literature. The authors perform their analysis using a cross-section of 105 countries over the period 2000-2010.

[8]The theoretical mechanisms are well-captured by Melitz (2003) and subsequent literature on heterogeneous firms in international trade where exporting activities incurs in fixed entry costs as well as variable costs. In this class of models, trade costs – such as geographical barriers or trade policies - affects both the intensive and extensive margins of trade.

[9] Market entry costs are measured using a composite and standardized index which considers a wide variety of costs such as those related to administrative procedures, transportation of goods to relevant sea ports, custom clearance and procedures etc.

improvement of their port efficiency measure – an important channel of trade facilitation – has a sizable positive impact on export diversity (in the range of +1,5 - +2,4% according to model specifications).

In the last two decades, the proliferation of Regional and Preferential Trade Agreements (RTAs/PTAs) has been a driving force in the reduction of barriers to trade. On the role played by these agreements in boosting the extensive margin of trade the findings are not univocal. Cook and Jones (2015) find evidence of a positive effect of the African Growth and Opportunity Act (AGOA)[10] on the diversity of exports from eligible Sub-Saharan African countries to the USA market while, using 6-digit bilateral trade data, Dutt et al (2013) estimate that WTO membership increases the extensive margin of exports by up-to 25 percent. Contrary to the above-cited studies, Dingemans and Ross (2012) find no evidence on the role of FTAs in promoting export diversification in Latin America. The authors argue that these important international trade infrastructures do not automatically affect what countries produce and export. In this respect they cite the emblematic case of Chile where most diversification occurred in the 1970s-80s, before the surge of FTAs signed by this country. Similar evidence is provided – for French firms trading in agricultural products - by Buono and Lalanne (2011) that show how the implementation of the Uruguay Round Agreement had a negative impact on the extensive margin (and positive on the intensive margin) of trade.

Helpman et al (2008), using country-level data, show that 'common approaches' to gravity models attempting at estimating FTAs effects on trade are significantly biased when omitting control for the extensive margin of trade. The authors find that when trade costs related to distance fall, the response of the extensive margin of trade is larger for less developed countries.

Also the study by Parteka and Tamberi (2013), using a panel of 60 countries in the period 1985-2004, finds that lower barriers to trade and RTAs have a positive effect on diversification.

Interestingly, the type of political regime seems to matter in shaping both export diversification and the sophisticatedness of the export basket. Using data from a large panel of 116 countries over the period 1970-2005, Makhlouf et al (2015) find evidence of an heterogeneous effect of trade openness on export specialization (the other side of the coin of diversification) and sophistication (proxied by the EXPY measure first introduced in Hausmann et al 2007). The authors show that trade openness enhances trade diversification as well as sophisticatedness only in democratic regimes while the opposite is found for autocracies. These results shed an important light into the importance of institutional settings in shaping the effects of global interactions.

New export discoveries and path-dependence

The development of 'new export products' depends on the existing export basket because new products can only originate from a re-combination of the current set of production capabilities. Current production capabilities are the key link between what a country produces today and what it will produce tomorrow, in other words the essence of the mechanism of path-dependence that this study aims to investigate.

---

[10] The AGOA is a PTA started in 2000 which provides a preferential treatment to a wide range of products exported from a large pool of eligible African countries. Two provisions were designed. One is the AGOA Generalized System of Preferences which extends the list of products that benefit from duty-free and quota-free access to the US market. The second is the AGOA apparel provision targeted to apparel and textile products.

In recent years, an increasing number of studies based on the PS framework have investigated the existence of path-dependence in the process of structural transformation. As in the original contribution by Hidalgo et al (2007), these studies generally use trade specialisation – measured by revealed comparative advantage - as a proxy of production specialisation and analyse its evolution across the PS over time.

Hausmann and Klinger (2010) and Hidalgo (2012) show how the export baskets of Ecuador and a pool of African countries (Kenya, Mozambique, Rwanda, Tanzania and Zambia) mostly consist in peripheral products[11] and highlight a rather strong persistence of the position of these countries over the PS through time.

Some studies have focused on the nexus between centrality in the PS and trade diversification. Minondo (2011) in a study on a set of 91 countries shows that the average connectedness of countries' export baskets (i.e. the degree of centrality in the PS) is a strong predictor of diversification levels. In a related study, Boschma and Capone (2016) analyse the process of trade diversification for EU-27 and European Neighbourhood Policy (ENP) countries between 1995 and 2010. The authors find evidence of path-dependence as countries develop their revealed comparative advantage in products related to those in which they were already specialized. Bahar et al (2017) find that the probability that a product enters the export basket of a country increases by 140% (from 1,9 to 4,6 per cent) if its relatedness to the export basket ten years before is one standard deviation higher than the mean value. These authors employ the 'density' measure developed by Hidalgo et al (2007) as main proxy for relatedness in their parametric analysis, which is computed as the average proximity (relatedness) of a new potential product to a country's current production basket. A positive and significant effect of these product-specific proxies of relatedness on the probability for a product to enter the export basket is evidence of path-dependence.

Although suggestive of the strong links between what is produced today and what will be produced tomorrow, these studies cannot be considered as a formal test of the PS hypothesis of path-dependence as they cannot discriminate between relatedness due to shared production capabilities (as the framework suggests) and spurious relatedness, which is the result of a random (unconstrained) process of diversification.

Our methodology allows us to test whether new products that enter in the export basket of a country are significantly related to those previously exported with revealed comparative advantage. We can therefore measure to what extent the evolution of comparative advantage is path-dependent and, in turn, shed light on those characteristics that are significantly associated with big leaps or jumps over the PS (i.e. path-defying changes).

3. The pattern of structural change: path-dependence versus path-defying changes

According to the PS framework, countries develop new products following a path-dependent process driven by the set of capabilities available in an economy. Thus, economies export a good with revealed comparative advantage because they already possess sets of production capabilities that can be easily redeployed from 'related' productions. Such relatedness between each couple

[11] One regularity that emerges from the network analysis of trade data is the fact that while industrialized countries are mainly specialized in the production of 'central goods' i.e. goods with higher average connections to others and higher sophisticatedness, low income countries have most of their export baskets located in the 'periphery' of the Product Space.

of products is proxied by the minimum of the pair wise conditional probability of being co-exported. Relatedness represents a sort of inverse measure of distance on the PS.

Products that enter the export basket of a country[12] should be those sharing most capabilities with - and therefore with highest relatedness to - the previously exported goods and services. Following Coniglio *et al* (2018), we test the validity of the path-dependency hypothesis by observing whether the actual distribution of the relatedness of new products' significantly differs from a randomly generated process (counterfactual distribution of relatedness).

Methodology

3.1. The general test of path-dependence

The first step of our analysis requires us to define 'new entries' as those goods that are not part of the production basket at time t_0 (*option set*) and enter the countries' export basket at time t_1. We recur to the standard definition of revealed comparative advantage (RCA) and define the set of goods in the export basket as those with a Balassa index that is larger than 1. Relevant new entries are represented by products with a RCA lower than 0.5 at time *t* and higher than unity at *t* + *T*.[13] For each country $c \in C$ and each time interval, we identify the set of new entries $n \in N_{c,t,t+T}$.

In the second step, we compute an *M x M* matrix containing the relatedness measures between any pair of goods *ij* exported in the world, ($i, j \in M_t$ where *M* is the set of goods exported in each year *t*).[14] For each country in the world, *c*, and for each of year, we denote x_{ict} as 1 if country *c* has a RCA in the production of good *i* at time *t* and 0 otherwise:

$$x_{ict} = \begin{cases} 1 \text{ if } RCA_{ict} > 1 \\ 0 \text{ otherwise} \end{cases} \quad (1)$$

where RCA_{ict} is the standard Balassa (1965) index employed as a measure of export specialisation. Thus, after creating the country-product matrixes of *RCAs*, following Hausmann and Klinger (2007), we compute the proximity (or inverse distances) between each couple of goods *i* and *j* as the minimum of the pair-wise conditional probability of being co-exported:

$$\varphi_{i,j} = \min\left\{P\left(x_i \mid x_j\right), P\left(x_j \mid x_i\right)\right\} \quad (2)$$

Where $\varphi_{i,j}$ represents the proximity between any good *i* and *j*.

In the third step, we denote with $B_{c,t}$ the set of goods exported with *RCA* by country *c* at time *t* (pre-existing export basket). We then define $D_{c,t}$ an *M x C* matrix of relatedness between the new

---

[12] When referring to export basket, we intend the set of products exported with revealed comparative advantage (Balassa, 1965) higher than unity.

[13] Since this choice of *RCA* thresholds is inevitably arbitrary, for robustness we identify a new entry using two additional alternative thresholds. We use a less restrictive threshold ($RCA_t < 1$ and $RCA_{t+T} \geq 1$) and a more restrictive one ($RCA_t < 0.2$ and $RCA_{t+T} \geq 1$). In terms of the time span employed we use *T = 5* years. For robustness purposes, also time intervals *T=10* years is taken into consideration.

[14] We obtain a 1,241-by-1,241matrix of products that are exported in the world economy in the period 1995 to 2010.

N_c products (entering the export basket between t and $t+T$) and the pre-existing export basket, for each country $c \in C$, as follows:

$$D_{ic} = \begin{cases} d_{ic}\left(\varphi_{ij}\right) = \max\left(\varphi_{ij}\right) \text{ when } j \in B_{ct},\ i \in N_c \\ \text{no value} \qquad\qquad\qquad\qquad \text{if } j \notin B_{ct} \end{cases} \tag{3}$$

where $d_{ic}\left(\varphi_{ij}\right) = \max\left(\varphi_{ij}\right)$ captures the proximity of new good i at time $t + T$ with the most related good already in the product space of country c at time t.[15]

Our formal test is based on the idea that if new entries follow a path-dependent process we should observe that the distribution of relatedness of observed new entries ($n \in N_{c,t,t+T}$) statistically differs from that obtained by randomly generated counterfactual entries and, more precisely, that actual data are significantly more concentrated than random data for high levels of proximity. The intuition behind this test is that, in a world where capabilities did not exert a significant constraint on the development of new products, we should observe a greater degree of diversification towards unrelated products, that is those with lower levels of proximity.

We build for each country c and each time interval (t, $t + T$) a counterfactual distribution of relatedness using 1000 random draws of size equal to the actual number of new entries ($N_{c,t,t+T}$) from the option set, i.e. products not exported with comparative advantage at t. We then reject the null hypothesis when the pattern of relatedness of actual new entries is statistically different from the randomly generated counterfactual.

The last step of the analysis is that of comparing the distribution of actual data with that of our counterfactual. Following Duranton and Overman (2005) and Coniglio *et al* (2018), we perform the analysis by implementing a Kernel smoothed density estimation of new entries' relatedness[16]. More precisely, for any level of proximity, d, we estimate the smoothed Kernel density function of relatedness as:

$$\overline{K}(d) \equiv \frac{1}{\left(\sum_{i=1}^{M}\sum_{t=1995}^{2010} I_{it}\right)h} \sum_{i=1}^{M}\sum_{t=1995}^{2010} f\left(\frac{d-d_{it}}{h}\right) \text{for all countries } c \tag{4}$$

with densities calculated non-parametrically using a Gaussian Kernel function with bandwidth h set according to Silverman's optimal rule of thumb (Silverman, 1986). $d_{i,t}$ is measured using eq.

[15] We employ two alternative measures for robustness: i) *average proximity*, a measure of relative distance of each new product with all pre-existing products; ii) *weighted average proximity*, relative distance of each new product with all pre-existing products weighted by the relative export share of the latter set of goods. These alternative measures provide qualitatively similar results (available upon request) but as argued in Coniglio et al (2018), our measure of maximum proximity is preferable as more in line with the theoretical idea of relatedness in the product space due to the sharing of a common set of capabilities between any two products *i*,*j*.

[16] A vector of distances for each of the three definitions of new entries is created in order to ensure the robustness of our results to the definition of this key element. In the paper, we only present, for the sake of brevity, the results for one definition of a new entry ($RCA_{t+T} < 0.5$ and $RCA_{t+T} \geq 1$).

(3) while $\sum_{i=1}^{M}\sum_{t=1995}^{2010} I_{it}$ is the total number of 'new' products in the considered time interval. The Kernel density functions are computed for each country *c*.[17]

Finally, we build a counterfactual distribution of relatedness and compare it with the actual one obtained from eq. (4). The counterfactual density function is based on simulated relatedness computed from 1,000 random draws of size identical to the actual one (i.e. $\sum_{i=1}^{M}\sum_{t=1995}^{2010} I_{it}$) for each country in our sample.[18]

Plotting the distributions with proximity as the variable on the horizontal axis ranging from 0 (highly unrelated) to 1 (highly related), we can imagine three possible scenarios emerging from the comparison of the two Kernel distributions. First, if the Kernel distribution of proximities of actual new entries lies to the right of the counterfactual, we can reject the null hypothesis of randomness for any level of proximity in the actual data (*full path-dependence*). A second possible scenario is when the Kernel distribution of proximities of actual new entries lies below the counterfactual one. In this case we cannot reject the null hypothesis of randomness as the two distribution of relatedness cannot be statistically distinguished (*no path-dependence* in the evolution of the comparative advantage). These two extreme can be interpreted as symmetrically opposed cases where the evolution of the export basket is either fully constrained or fully unconstrained by the initial set of productive capabilities. A third (more likely) scenario is a mix between the two reported above whereby the Kernel of actual data is partly to the right of the counterfactual one. In this scenario we can reject the null hypothesis of random relatedness only for those new entries for which the first Kernel lies above the counterfactual, i.e. only for relatively high level of proximities which cannot be the result of a random process. This scenario would give a general support to the hypothesis of path-dependence but also tell us that long-distance and random jumps over the product space can occur.

3.2 A single product's test of path-dependence

In addition to the test described above, we introduce a test of random relatedness for each new product that enters the export basket of a country in a given period. This strategy allows us to measure how much path-dependence (or path-defying changes) we observe in the data for each country in our sample by computing the share of statistically path-dependent new entries over total new entries in a given period.

Our counterfactual distribution of proximities is built using Monte-Carlo methodology as follows. For each country $c \in C$ and for each 5-year time interval, we randomly draw, from the option set at time *t*, a number of products equal to the actual number of new export entries at time *t+ T*, *Jc*, compute proximities using eq. (3) and generate an average value per draw. The random draw is carried out 1,000 times in order to compute a distribution of random average

---

[17] I_{it} is a product by year matrix of size*M x S* which has values of 1 for each new entrant product for each country and 0 otherwise, being *M* the number of products in the PS and *S* the number of periods included in the analysis (*i.e.* 16 time- periods when T=5 and 11 periods when T=10).

[18] In every simulation, for each country, we randomly draw a number of new entries from the products not in the export basket at time *t* which is identical to the number of effective ones. In other words, our counterfactual exercise takes explicit account of the country-time-specific distribution of new entries even though the time dimension is then pooled in the Kernel estimates since in most countries – and mainly for the most industrialized ones – the number of new entries per period is quite low and mines the reliability of the non-parametrical analysis. The country-time peculiarities are then investigated with a Monte-Carlo methodology later presented which allows to use them in the parametrical analysis.

proximities which represent our country-time-specific statistical counterfactual. From these counterfactual distributions of proximities, we identify the 95th percentile value, which represents our threshold to detect statistically path-dependent new entries. In other words, a new product is labelled as (non) path-dependent when its proximity with the pre-existing export basket is (lower) higher than the (5th) 95th percentile of counterfactual average proximities.

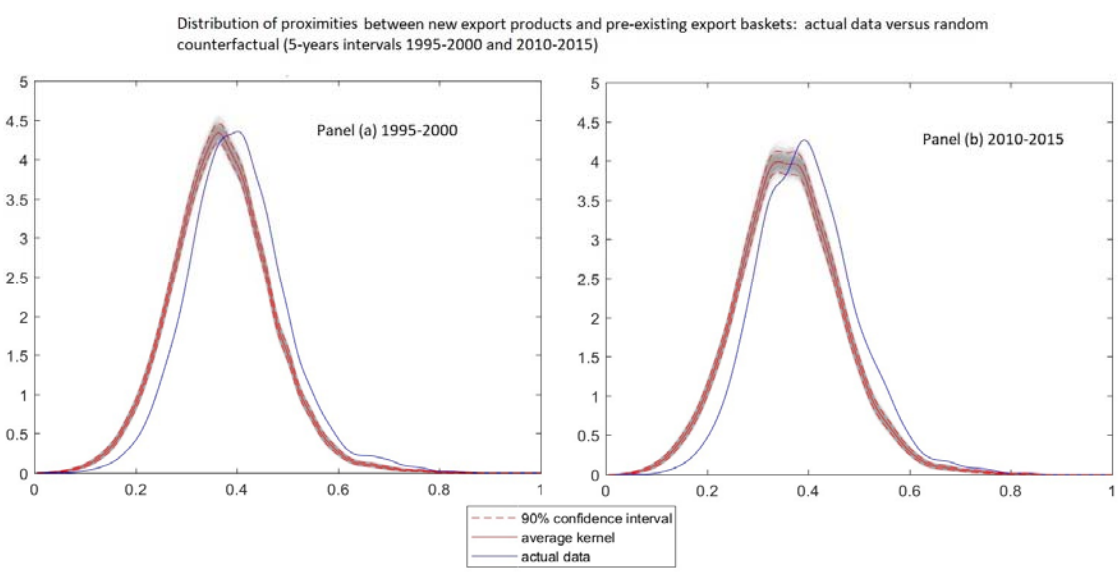
Data

Our analysis is carried out using 4-digit Harmonized System exports data for the period between 1995 and 2015 (CEPII BACI database). Such disaggregation allows us to have a (year specific) Product Space composed by approximately 1240 goods globally exported for up to 221 countries.[19] Export data are used to compute all countries revealed comparative advantages and then to obtain the year-specific matrixes of relatedness between each couple of goods as well as to identify the new entries for each country in every moving time interval. Since the length of our time periods of analysis is 5 years – 10 years as robustness check – our t goes from 1995 to 2010 while $t + T$ goes from 2000 to 2015.

Results

Figure 1 represents the (kernel) distributions of proximities between new entries at time $t + 5$ and pre-existing export baskets at time t for all countries in our sample in two time intervals (1995-2000 and 2010-2015, respectively the first and last in our analysis). Distance is defined as the maximum proximity (eq. 3) and a new export is identified as new entry when the RCA at $t + T$ is lower than 0.5 and RCA at $t + T$ is higher than unity. The comparison between the Kernel of the actual data (in blue) and the Kernel of randomly generated data (counterfactual) is a general test on the (null) hypothesis of unconstrained evolution of comparative advantage against the hypothesis of path-dependence.

Figure 1



[19] Hidalgo et al (2007) compute a network of relatedness – i.e. the Product Space – using SICT-4 digits trade data in the time interval 1998-2000; their matrix of relatedness includes 775 products.

We can reject the null hypothesis (of unconstrained diversification) for any level of relatedness (horizontal axes, *d*, ranging from 0 - max distance or min proximity - to 1 - min distance or max proximity), when the actual data's Kernel lies above the upper limit of the 90 percent confidence interval.

In both panel of *Figure 1* we can state that, above proximity value of 0.4, there is a non-random concentration of new entries. Using the simile of the dart-board, we find that actual darts (actual new products) tend to concentrate closer to the target (pre-existing export basket) that we would expect if they were not constrained by the initial set of capabilities (randomly generated new products). Evidence suggests that the pattern does not significantly change when shifting the focus on a different period and that path dependence has driven the overall process of structural transformation in both the periods 1995-2000 (*Figure 1a*) and 2010-2015 (*Figure 1b*).

The general pattern of path dependence is confirmed using two alternative definition of relatedness between new entries at $t + T$ and export basket at t, weighted average proximity and average proximity (see *Appendix 1*).[20]

Our finding support the 'mixed' scenario outlined above. Since the Kernel distribution of actual new entries lies above the counterfactual only for relatively high level of proximities, we cannot reject the null hypothesis of random relatedness for the entire distribution of relatedness. The results confirm the idea that, in general, diversification is strongly constrained by capabilities: what you export today strongly influences what you will export tomorrow (*constrained diversification*). *Figure 1* also suggests that a non-negligible share of new products entering the export basket are concentrated at distance level where the actual data kernel lies under the counterfactual distribution; these products are evidence that *unconstrained diversification* occurs frequently and that countries do defy their initial comparative advantage.

The results from the general test presented graphically in *Figure 1* mask large cross country heterogeneity. The methodology presented above can be used to test the general pattern of path-dependence for each of the countries in our sample.[21]

Figure 2 presents the distribution of distances of actual new entries over the period 1995-2015 compared to the counterfactual ones for four economies: Cambodia (*panel a*), Republic of Korea (*panel b*), United States (*panel c*) and Yemen (*panel d*). The results highlight heterogeneous patterns. The evolution of Cambodia comparative advantage shows a strong degree of path-dependence; the south-Asian economy recorded, for level of relatedness higher than 0.38, a non-random concentration of actual data. A similar pattern is found in Yemen where the test of randomness is rejected at 95 percent confidence for level of relatedness higher than 0.35. *Panels a* and *d* show that the process of structural transformation involving developing economies such

---

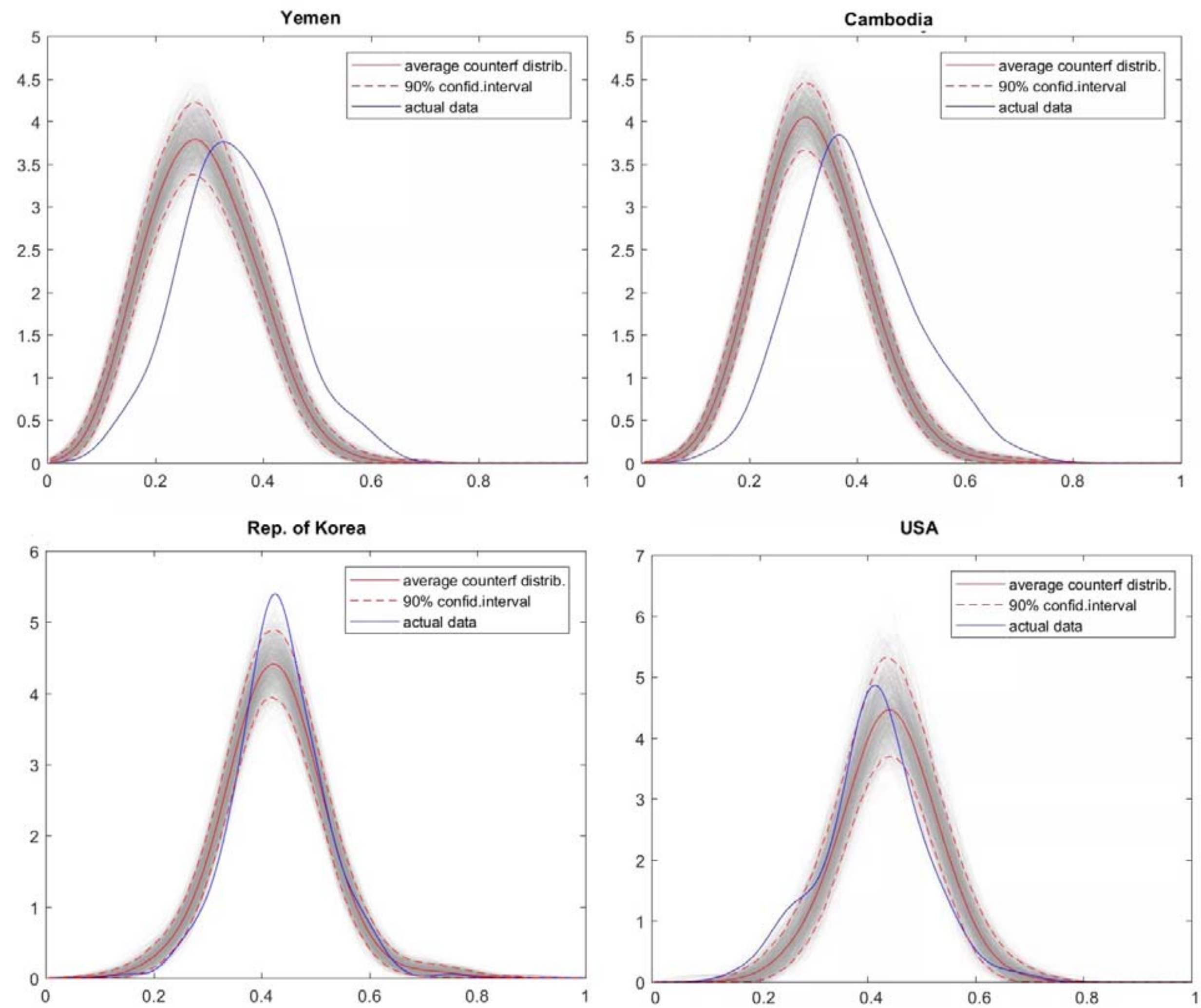
[20] We employ for alternative measures of relatedness, one-sided Kolmogorov-Smirnov tests for first order stochastic dominance where we compare the relatedness distribution of actual data with the randomly generated one. This parametric test the null hypothesis that the two distributions can be generated by the same (random) process and that the relatedness of actual data are significantly more concentrated at higher level of proximities (i.e. the cumulative distribution function of the actual data distributions lies below the mean values counterfactual one). These tests have been performed for all new entries in the world in 5-years time intervals. The results strongly reject the hypothesis of the two distributions and confirm that there is a significantly higher concentration of actual data at high leel of proximity. These results are available in *Appendix 3*.

[21] We derive country-specific and year specific counterfactual distributions, hence our randomly generated distances account for the different pre-existing export baskets in terms of number of variety and their relative position in the Product Space. It is important to notice that the frequency of new entries is also highly heterogeneous, for some countries many new products enter the export basket in the 5-years intervals while for other countries – in particular relatively more advanced economies – new export discoveries are observed less often.

as Cambodia and Yemen are strongly influenced and constrained by the production capabilities locally available.

On the contrary, Republic of Korea has a kernel distribution of actual data statistically similar to the one generated by the random process: the inclusion of new products in the export basket of the Korean economy does not follow the PS predictions of path-dependence. The pre-existing position of US comparative advantage over the PS does not seem to condition its evolution in the subsequent 5 years; the kernel densities for the period under scrutiny do not show a statistically significant concentration of new entries for relatively high levels of relatedness.[22] The results represented in *Figure2* are robust to the new entries' identification strategy and to the measure of relatedness taken into account.

Figure 2 – *New export products: distribution of actual and counterfactual distances in selected economies (all new entries, 1995-2015)*



Notes: Kernel densities of actual and counterfactual distances

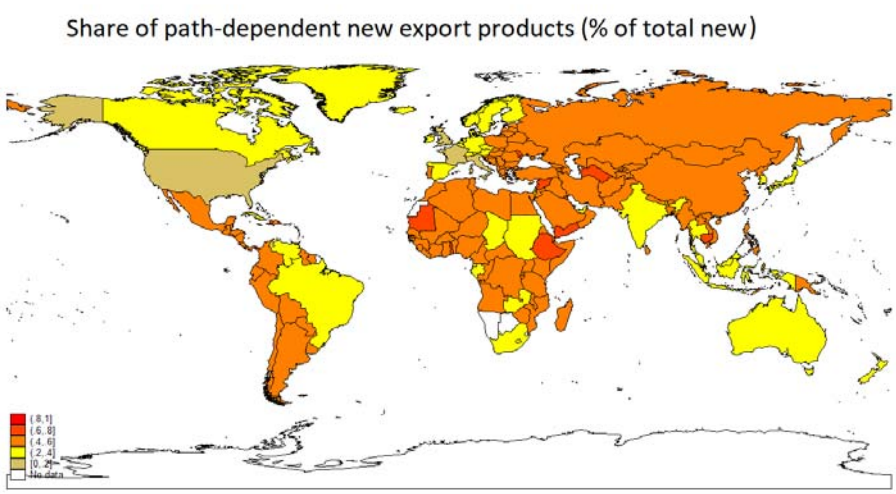
[22] Most of kernel estimates of actual data are drawn inside the 90 percent confidence interval and the only distances for which the distribution of observed relatedness lies above the 95th percentile of the kernel counterfactual's distribution are the farthest from the initial export basket, denoting a non-random concentration of new entries for very low levels of relatedness.

In general, from the results of the general test on path-dependence it appears that small, least advanced and resource abundant countries tend to confirm the prediction of the PS framework and evolve their export baskets including mainly (but not only) those goods which are strongly related to the initial export basket.

When considering the ‘single product’ test we move from the analysis of the entire bundle of new export products to that of each single ones. A new product is labelled as path-dependent (path-defying) change when its proximity with the pre-existing export basket is higher (lower) than the 95^{th} percentile (5^{th} percentile) of counterfactual proximities obtained using the Monte Carlo methodology described in *Section 3.2*.

Over the full sample, approximately 49.2% of the total number of new entries in the world between 1995 and 2015 are significantly related (path-dependent) to the initial export basket. The other side of the coin of this finding is that more than half of the new products that enters the export basket represents relatively more radical ‘jumps’ over the Product Space. A result that suggests that it is fundamental to take the normative implications stemming from this framework with a grain of salt.

Figure 3



Note: average values in the period 1995-2015

Figure 3 represents the average values of the share of path-dependent new entries in the period 1995-2015 when relatedness is measured as the maximum proximity between new entries and already exported products[23]. Among countries where we observe diversification to have been

[23] Maps with alternative definitions of distance are available upon request from the authors. Using alternative definition of distance/relatedness does not change substantially the relative ranking of countries in terms of path dependence shares.

least constrained by path-dependence we find developed countries such as US, UK, France and Italy. Most western European countries, as well as very different countries such as Japan, Canada, Brazil, Australia, India, and South Africa, show very low path dependence in the period. On the opposite, economies whose diversification seems characterized by high degree of path-dependence include Yemen, Ethiopia, Mauritania, Syria, Turkmenistan and Cambodia. Most of these countries are developing economies highly relying on natural resources.

In order to analyse more systematically which characteristics drive the observed heterogeneity in the process of structural change - as well as its importance in determining economic growth – we employing panel data analysis. These empirical exercises are presented in the two next sections.

3. When does the apple fall near (or far) from the tree? The determinants of path-dependence over the Product Space

In the previous section we found evidence of a large cross-country heterogeneity in the degree of path-dependence of new entries in the export basket. In this section we analyse the country-level characteristics that drive these heterogeneous patterns. The emergence of new sectors that are unrelated to the pre-existing production basket signals the ability of countries to effectively free themselves from the constraints posed by the initial set of production capabilities or, alternatively, to redeploy them with relative ease. This ability is essential to promote structural transformation and economic growth.

We estimate a *panel tobit model* where our dependent variable, *path-dependent entries (PD share)*, is the share of new entries in the export basket of country *i* at time *t* that are statistically related to the pre-existing economic specialisation at time *t – 5*. As in the previous section, we use, as our preferred measure of distance/relatedness, *maximum proximity* (eq. 3) and define new entries as those products with a RCA≥1 at time *t* and an RCA<0.5 at time *t – 5*. In other words new entries are products that enter in a statistically significant way into export baskets.[24] Our empirical model is the following:

$$PDshare_{i,t} = \alpha + \beta X_{i,t-5} + \phi_i + \gamma_t + \varepsilon_{it} \qquad (5)$$

where $X_{i,t-5}$ includes our main country-level covariates that are measured five years earlier (*time t – 5*), ϕ_i includes time-invariant country level controls, γ_t time fixed effects and ε_{it} is the error term.

A full description and summary statistics of our dependent variable and of all covariates is reported in *Table 1*. As a proxy for the relative level of development we use *GDP per capita (in log)*. We expect this variable, which reflects the degree of factors' productivities and competitiveness, to be negatively related with the dependent variable. More advanced economies have a better endowment and a wider set of productive capabilities and are, in turn, more able to

[24] In this section we only report results based on our preferred measure of path-dependence. For robustness we employ alternative definitions of the dependent variable using different criteria for both measuring relatedness and for identifying new entries (i.e. alternative thresholds for the RCA index at time t; RCA<0.1, RCA<0.2 and RCA<1). The results reported using these alternatives are qualitatively similar and available upon request.

re-combine these capabilities in order to produce goods that are unrelated to the existing specialization. As described in *Section 2*, the existing literature has emphasized the existence of a positive relationship between GDP per capita and trade diversification. While these studies look at the 'quantitative' dimension of trade diversification (i.e. the number of new entries in the export basket), to our knowledge, this is the first study which looks at the 'direction' that the process of trade diversification takes.

We control for country size in terms of *population* since we expect that larger countries have, ceteris paribus, a higher potential to diversify away from the current comparative advantage. For this variable as well we expect a negative (positive) association with the number of path-dependent (path-defying) new entries.

Trade openness measured as the sum of total export and import over GDP (in log) is included among our covariates. The effect of this variable on the degree of path-dependence is expected to be ambiguous. On one side, we expect that, controlling for other countries characteristics as size, more open economies are less likely to observe a path-dependent evolution as a higher degree of internationalization should facilitate the access to a wider set of productive capabilities and to trade opportunities in unrelated sectors. On the other side, when countries have a large tradable sector they might be less likely to attract the necessary resources (from the relatively smaller non-tradable sector) for kick-starting new line of export, in turn export discoveries are more likely to be related to those areas in which countries already have a comparative advantage (potential lock-in effect).

We include a measure of export diversification, *variety* (in log), that is defined as in Frenken (2007). We expect that more differentiated export baskets – a signal of large and diversified sets of production capabilities – will allow an economy to diversify away from the initial production basket. We also expect that trade diversification in products that draw from the same pool of production capabilities will provide relatively less opportunities to diversify away from the current comparative advantage. We employ two variables, *related variety* and *unrelated variety*, which measure respectively the degree of trade diversification within a specific sector or across sectors. While more unrelated varieties by enriching the portfolio of production capabilities might increase the ability to diversify in unrelated areas of the PS, a higher number of related varieties might induce a 'lock-in' effect which hampers diversification.[25]

Economic diversification is a compelling policy goal for resource rich countries which are often trapped in a 'resource curse' (Humphreys et al, 2007). Abundance in *natural resources* might negatively affect trade diversification through several channels; among the most debated in the literature are the negative effects of resource rents on institutional quality and real exchange rate appreciation due to a 'Dutch Disease' effect. We expect that this variable is positive associated with our dependent variable as an economy that is strongly dependent on natural resources is less likely to diversify away from the current export basket.

Since foreign investors might be important agents of structural transformation we include *FDI inflows* (net, share of GDP) and we expect that countries attracting a relatively larger share of FDI are more likely to defy their static comparative advantage and diversify the economy adding to the export basket unrelated varieties (thus we expect a negative effect on the dependent variable).[26]

---

[25] Castaldi *et al* (2015) use similar covariates in their analysis on the emergence of technological breakthroughs.

[26] We include also FDI stocks as an alternative to inflows and results (not reported but available upon request) are similar.

Human capabilities are a crucial element for the 'cost discovery' process that leads to the introduction of new export varieties. Given the notorious lack of panel data on formal education, and more in general on the stock of human capital, we use two proxies of human capital: the number of *scientific and technical journal publications* (a measure of output) and *educational expenditure* as a share of a country GDP (a measure of input).[27]

Our specifications include two additional control variables. The total number of new entries in the export basket and the relative importance of new entries measured as their share in total exports; both variables are controls for size effects.
Finally we include time dummies that account for common trends in the data (for instance due to global economic shocks) and macro-area dummies which account for other time invariant characteristics like remoteness or participation in regional trading blocks.

In *Table 2* we report the estimates of the random-effect tobit models on the determinants of the degree of path-dependence in the evolution of countries comparative advantage. We start from a parsimonious model which includes GDP per capita and population as well as control variables and gradually include additional covariates in further specifications.
The coefficient on our proxy for the level of development is always negative and highly significant suggesting that more advanced economies experience a change in the export basket that is less path-dependent – a 1% increase in GDP per capita at time *t* is associated to a decrease (increase) between 2.3 and 5.3 percentage points in the share of path-dependent (radical) entries in the export basket 5 years later. A similar effect is related to the size of population although this result is less robust.
Interestingly, we do not find evidence of a significant effect of the volume of international trade on the direction of trade diversification. As expected, when more trade varieties are present in countries export baskets, it is more likely that diversification in the following 5 years would be unrelated, i.e. that countries will be able to 'jump' further over the Product Space (*Table 2, column 2*). This result is entirely driven by unrelated variety while higher diversification within the same sector has no significant effect (*columns 3 to 7*). A 1% increase in our measure of unrelated trade varieties has a large negative (positive) effect on the share of path-dependent (radical) new entries (approximately 5% in the specification reported in *column 3*).
We find mild evidence that a higher share of natural resources in the economy is associated to a higher degree of path-dependence in the subsequent 5 years. This result is suggestive of the fact that 'extractive' rents do not only limit a country's ability to diversify but also 'lock' them into a highly related diversification over the Product Space.
Countries with higher FDI inflows (or stocks) are more likely to 'defy' their static comparative advantage (*column 5*); this result suggests that MNEs might play an important role as agents of structural change[28] and lead to the acquisition of a wide range of production capabilities. This can be considered as an important (and new) channel of spillovers from foreign investments.

---

[27] Note that for those countries for which we have detailed information on the distribution of population by educational level, the number of scientific and technical articles in journals have a high level of correlation with the share of primary and secondary education which are found to be among the most important driver of trade diversification by Jetterand Ramirez-Hassan (2015).
[28]Neffke et al (2014) show that firms with a higher degree of internationalization represent crucial agents of structural change since non-local firms and entrepreneurs tend to diversify in sectors that are less related to pre-existing regional production bundles.

Human capital has a limited effect on our dependent variable. We find that the number of scientific and technical journal articles is positively related with the share of path-dependent new entries. One should note that the number of observations is significantly reduced as we include this variable as some countries drop out of the sample.
It is interesting to note that most of the variance in the dependent variable is due to the cross-sectional difference between countries (more than 80% of total variance) rather than the temporal variation.

Table 2 - *When does the apple fall near the tree? The determinants of path-dependent new entries in the export basket*

Dependent variable: % share of path-dependent new entries over total new entries (5 years time periods)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GDP per capita (log)	-4.020***	-3.109***	-2.348***	-2.789***	-2.724***	-5.305***	-5.384***
	(0.727)	(0.751)	(0.806)	(0.834)	(0.838)	(1.405)	(1.891)
Population (log)	-1.126**	-0.364	0.227	-0.376	-0.381	-2.282*	-1.255
	(0.474)	(0.502)	(0.553)	(0.608)	(0.612)	(1.192)	(1.654)
Trade openness (log)	0.231	0.0203	0.442	-1.371	-0.820	-0.605	3.011
	(1.292)	(1.288)	(1.289)	(1.432)	(1.446)	(1.612)	(3.014)
Trade variety (log)		-7.615***					
		(1.730)					
Related varieties (log)			-0.290	0.849	0.508	2.153	1.589
			(1.290)	(1.326)	(1.337)	(1.622)	(2.367)
Unrelated varieties (log)			-4.992***	-4.185**	-3.892**	-5.551**	-8.420**
			(1.679)	(1.794)	(1.803)	(2.219)	(3.566)
Natural resources (% of GDP)				0.177*	0.162*	0.272**	0.00470
				(0.0910)	(0.0917)	(0.106)	(0.161)
FDI inflows (net, % of GDP)					-0.0777**	-0.0550	-0.0267
					(0.0383)	(0.0417)	(0.0551)
Scientific & tech. publications (log)						1.891**	1.925
						(0.921)	(1.262)
Educational expenditure (% of GDP)							-0.353
							(0.223)
Total new entries (last 5 years; log)	8.471***	10.47***	10.91***	11.34***	11.40***	11.19***	9.904***
	(0.931)	(1.024)	(1.050)	(1.086)	(1.098)	(1.341)	(1.781)
Share of new entries over total export value	-19.67***	-20.06***	-21.10***	-23.80***	-23.65***	-16.68**	-22.78**
	(4.826)	(4.811)	(4.822)	(4.900)	(4.968)	(6.652)	(9.167)
Constant	71.87***	59.50***	47.80***	61.71***	59.20***	101.7***	91.58**
	(14.44)	(14.59)	(15.21)	(16.01)	(16.07)	(27.14)	(39.34)
sigma_u	*7.650****	*7.581****	*7.599****	*8.083****	*8.057****	*8.062****	*9.318****
	(0.772)	*(0.762)*	*(0.763)*	*(0.786)*	*(0.789)*	*(0.951)*	*(1.197)*
sigma_e	*27.22****	*27.13****	*27.11****	*26.76****	*26.73****	*27.17****	*25.54****
	(0.393)	*(0.391)*	*(0.391)*	*(0.387)*	*(0.389)*	*(0.487)*	*(0.589)*
Observations	*2,715*	*2,715*	*2,715*	*2,696*	*2,657*	*1,827*	*1,137*
Macro-area and year dummies	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*	*YES*
Number of countries	*177*	*177*	*177*	*177*	*177*	*173*	*154*

Note: Estimations using Panel Tobit method; Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Business and Institutional Environment and path-dependence

Path-defying changes in the export basket are the results of investments carried out by economic agents in risky and highly uncertain cost-discovery activities. The quality of the business and institutional environment is fundamental for boosting the incentives to undertake such potentially high-reward but also high-risk investments. In this sub-section we analyse the role that these policy relevant factors have in shaping the degree of path-dependence by using country measures of business and institutional quality developed by the Fraser Institute (2017).
In *Table 3* we report the results of our estimates where we test the importance of a *general index of economic freedom* and its main sub-components. In order to test the relative importance of these institutional variables for countries at different level of development we have interacted the covariates reported in each of the 7 columns with income level dummies which follow the World Bank classification. Note that these indexes range from a maximum of 10 (highest level of freedom or lower degree of Government intervention) to a minimum of 0 (lowest level of freedom or highest degree of Government intervention), hence an increase of the variable implies that there is a reduced intervention or more 'economic freedom'.

Table 3 - *The determinants of path-dependent new entries in the export basket: the role of the Institutional Environment*

Dependent variable: % share of path-dependent new entries over total new entries (5 years time periods)

	(1)	(2)	(3)	(4)	(5)	(6)
I_Var = Institutional Variable reported in Column	*Economic Freedom Index*	*Size of Government*	*Legal system and property rights*	*Sound money*	*Freedom to trade internatio nally*	*Regulatio n*
Institutional Variable (baseline = High Income Country)	-5.611**	-1.008	-0.451	-3.15**	-2.151	-1.648
	(2.795)	*(1.229)*	*(1.536)*	*(1.435)*	*(1.949)*	*(1.739)*
Low Income Country * I_Var	5.453	0.451	2.129	1.754	2.122	5.655**
	(3.745)	*(1.880)*	*(2.359)*	*(1.832)*	*(2.767)*	*(2.614)*
Lower-Middle Income Country* I_Var	3.186	-1.653	-0.364	2.817	1.944	0.663
	(3.526)	*(2.044)*	*(2.268)*	*(1.808)*	*(2.499)*	*(2.734)*
Upper Middle Income Country* I_Var	5.732	-0.392	3.650	3.402**	3.464	0.427
	(3.639)	*(1.822)*	*(2.404)*	*(1.707)*	*(2.699)*	*(2.442)*
Obs.	*1,504*	*1,503*	*1,514*	*1,504*	*1,499*	*1,504*
N. of countries	*143*	*143*	*143*	*143*	*143*	*143*

The results suggest the existence of some degree of heterogeneity across income groups. A higher *Economic Freedom Index* is negatively (positively) associated with path-dependence (path-defying) changes in the export basket five years later but only for high income countries while the effect is not significant for countries with lower level of income. The *size of the Government* is negatively but not-significantly associated with the dependent variable but not for low income countries. This result lends some weak support to the hypothesis that an active role of the State can be less distortive in very poor countries where market failures are pervasive and probably more severe than government failures (Stiglitz 2002, Bjorvatn and Coniglio, 2012). A similar pattern is found when introducing composite indexes of the quality of the *legal system* *and of property rights protection*, the *freedom to trade internationally* and *regulation* although the estimated coefficients are not statistically significant (columns 3, 5 and 6). The effectiveness and coherence of monetary policies seem to matter in shaping the likelihood of observing path-defying changes in rich countries but not in relatively poor ones (index of '*sound money'*, column 4).

Overall the results suggests that a less pervasive and distortive Government intervention leads to a higher ability to move dynamically the comparative advantage away from the current one in relatively rich countries where market failures and barriers to entry in new sectors are likely to be low. On the contrary we do not find strong evidence of these effects in less developed economies.

Table 4 - *Role of the Government and path-dependent new entries in the export basket: some specific channels*

Dependent variable: % share of path-dependent new entries over total new entries (5 years time periods)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Government consumption	*Transfers and subsidies*	*SOEs and investments*	*Top marginal tax rate*	*Non tariff trade barriers*	*Credit market regulations*	*Ownership of banks*	*Business regulations*
Richest countries (top 35%)	-0.529 (0.95)	-1.085 (1.01)	0.682 (0.66)	-0.696 (0.78)	-4.537*** (1.40)	-3.058** (1.37)	-1.711** (0.69)	0.350 (1.76)
Obs.	*578*	*571*	*570*	*572*	*530*	*573*	*570*	*544*
N. of countries	*64*	*60*	*59*	*61*	*60*	*62*	*59*	*62*
Poorest countries (bottom 35%)	2.23* (1.20)	-0.532 (2.76)	-0.616 (0.55)	-2.21*** (0.80)	-0.0537 (2.36)	-0.0772 (1.20)	0.431 (0.65)	-4.196* (2.15)
Obs.	*425*	*331*	*444*	*279*	*272*	*418*	*383*	*333*
N. of countries	*52*	*46*	*52*	*39*	*42*	*51*	*46*	*49*

Note: Estimations using Panel Tobit method; Includes all covariates and controls included in Table 1 (model 6); Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

In *Table 4* we report the estimation results on more specific channels where we consider separately rich countries (top 35%) and poor ones (bottom 35%).[29] This exercise allows us to further explore the heterogeneity already emphasized in *Table 3*. When we consider the four sub-component of the aggregate index 'Size of Government' – columns 1 to 4 – we find, in coherence with *Table 3*, that the intensity of government intervention in the economy (for instance via industrial policy) does not affect the direction of the evolution of comparative advantage. One exception is Government consumption and top marginal tax rate in poor countries. More precisely a decrease (increase) in government consumption is associated to less (more) path-dependence. On the contrary a lower top marginal tax rate in the poorest countries in the sample is associated with a higher degree of path-defying changes.

[29] We employ sub-components of the more aggregated indexes reported in previous estimates.

Overall the results suggest that a less pervasive Government intervention are associated with higher ability to move dynamically the comparative advantage away from the current one in relatively rich countries where market failures and barriers to entry in new sectors are likely to be low. In rich countries, Government intervention seems to hamper diversification in unrelated and new areas of the PS. On the contrary we do not find strong evidence of these effects in less developed economies with one important exception: in relatively poor countries improvements in business market regulations (bureaucratic costs, cost of starting a business, bribes and corruption, licensing restrictions and tax compliance) improve the ability of defying the static comparative advantage.

4. Path-defying changes and economic performance

In this section we explore the following research question: do unrelated new entries in the export basket – those which defeat the static comparative advantage – lead to higher economic growth? We analyse this relationship in panel growth regressions using alternative estimation approaches as in Hausmann et al (2007).[30]

In *Table 5* we report the results of our growth estimates based on the full sample of countries. Our *dependent variable* is defined as *5-year average growth rate of GDP per capita*. We use non-overlapping periods in the time span 1995-2015 for a sample of 177 countries. The variables employed in the model and the list of countries covered are described in *Appendix 2*.

***Table 5**- Do path-defying changes (non path-dependent entries) lead to higher growth?*

Dependent variable: GDP per capita growth, yearly average (5-year panel; % change)	Pooled OLS (1)	LSDV (2)	FE (3)	FE (4)	IV (5)	System GMM (6)
Path-defying new entries (log)	0.271***	0.279**	0.337***	0.354***	0.947*	0.301**
	(0.105)	(0.126)	(0.128)	(0.115)	(0.489)	(0.120)
Initial GDP per capita (log)	-0.344***	-4.04***	-5.88***	-9.85***	-10.9***	-3.77***
	(0.078)	(0.672)	(0.976)	(1.475)	(1.334)	(0.796)
Initial Export sophisticatedness, ExpY			0.515	-0.338	-0.545	0.756
(log)			(0.575)	(0.632)	(0.556)	(0.305)
Initial Economic Freedom Index			1.003***	-0.275	-0.148	-0.500
			(0.234)	(0.417)	(0.356)	(0.468)
Initial Scientific & tech. publications (log)				0.887***	0.0667	0.591
				(0.287)	(0.337)	(0.434)
Constant	2.51***	28.66***	39.59***	84.20***	84.29***	26.554***
	(1.166)	(4.712)	(9.062)	(11.40)	(9.279)	(7.931)
Observations	*692*	*692*	*504*	*384*	*384*	*383*
R2	*0.031*	*0.554*	*0.188*	*0.356*	*0.745*	
Number of countries	*177*	*177*	*143*	*141*	*141*	*141*

Note: Least Square Dummy Variable (LSDV) models in column (1)(2) include country fixed effects; Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 *Panel IV regression:* path-defying new entries (log) are instrumented with the 5 year lag of the variable and the log of population. *GMM is the Arellano-Bond / Blundell-Bond system estimator and includes the lagged dependent variable and year dummies as suggested by Roodman (2009) / (n° of instruments = 13).*

[30] Given the proximity of our research question to the seminal paper of Hausmann et al (2007) entitled 'What you export matters' that investigates using panel regressions the relationship between export complexity (measured with the variable ExpY) and economic performance, we follow as much as possible their approach in this section.

The *main variable of interest* is our measure of *the share of path-defying new entries* in countries' product space. This variable is defined following our original methodology presented in *Section 3* and captures the share of products that enters a country's export basket with a RCA index ≥1 that are unrelated to the pre-existing export basket. The main hypothesis we would like to test is whether we observe a relatively better growth performance in those countries that have introduced new export discoveries in areas of the Product Space (statistically) unrelated to the pre-existing export basket. We expect a positive effect of this variable on economic growth as diversification in unrelated variety implies that a country has a wider set of productive capabilities that can be combined in an efficient way in order to boost productivity and in turn economic performance.

We employ, as in Hausmann et al (2007), a rather parsimonious growth model which includes the following covariates: the *initial (log of) GDP per capita* (a proxy of the initial level of development which is expected to be negatively related to subsequent growth, i.e. convergence hypothesis), the *initial level of export basket sophisticatedness* measured as (log of) *ExpY* (the main covariate in Hausmann et al 2007), a comprehensive measure of the quality of business environment (*Economic Freedom Index*) and a proxy for human capital (for boosting the number of observations we employ our 'output' measure used and described in the previous section, i.e. *the log of the initial number of scientific and technical journal articles*).

We find robust support for our hypothesis of a positive effect of path-defying changes in the country export basket and subsequent economic growth. In *columns (1)* and *(2)* we use respectively pooled OLS and Least Square Dummy Variable (LSDV) model. For both models the effect of radical changes on growth is positive and significant. A fixed effect panel regression is employed in *columns (3)* and *(4).* As our main variable of interest could be in principle endogenous, we report in *columns (5)* and *(6)* the results of IV regression and system GMM respectively. In the IV regression as instruments for the share of path-defying new entries we use its 5 year lag and country size proxied with population. As the theoretical hypothesis of scale effects in economic growth has not been robustly confirmed by the existing empirical evidence (see Rose, 2006 and a similar argument in Hausmann et al, 2007) we believe that our choice of instruments is not problematic.[31]

Given the weak evidence of endogeneity of our measure of path-defying changes, our preferred specifications are those employing Fixed Effects as these estimates allows us to control for time-invariant country effects and better identify the impact of an unrelated evolution of the export basket on subsequent growth netting out the within country differences. The results reported in columns (3) and (4) suggest that a 10% increase in the share of path-defying new entries leads to a 0,78-0.82% increase in yearly growth of GDP per capita in the next 5 years; the growth effect is sizable. The result obtained with the system GMM estimator is in line with the FE model while the IV estimates suggest an even higher effect (although weakly significant).

As a final step of our analysis we estimate the same specification on a smaller sample which excludes high-income countries. The estimation results, reported in *Table 6*, confirm the findings

[31] It should be noted that when performing standard tests we reject the null hypothesis of endogeneity of our key co-variate. For completeness and given the well-known weakness of these tests we decided to include anyway the IV estimates.

highlighted above. The magnitude of the estimated coefficients is comparable to that obtained from the full sample when country time invariant characteristics are considered (see FE models in columns 3 and 4).

Table 6 - *Do path-defying changes (non path-dependent entries) lead to higher growth? (excluding high income countries)*

Dependent variable: GDP per capita growth, yearly average (5-year panel; % change)

	Pooled OLS (1)	LSDV (2)	FE (3)	FE (4)	IV (5)	System GMM (6)
Non path-defying new entries (log)	0.361***	0.307**	0.324**	0.338***	0.0066	0.287*
	(0.124)	(0.146)	(0.140)	(0.125)	(0.498)	(0.159)
Initial GDP per capita (log)	-0.138	-2.905***	-4.086***	-8.512***	-10.98***	-7.451***
	(0.142)	(0.796)	(1.229)	(2.059)	(1.616)	(1.104)
Initial Export sophisticatedness, ExpY (log)			0.744	-0.0291	0.915	1.32*
			(0.773)	(0.833)	(0.740)	(0.771)
Initial Economic Freedom Index			0.933***	-0.639	-0.588	-1.326***
			(0.274)	(0.574)	(0.420)	(0.456)
Initial Scientific & tech. publications (log)				0.971***	-0.369	0.666*
				(0.349)	(0.343)	(0.398)
Constant	2.51***	21.59***	19.73*	65.26***	79.96***	51.52***
	(1.166)	(5.466)	(11.66)	(13.91)	(10.71)	(9.38)
Observations	*479*	*479*	*325*	*252*	*252*	*251*
R2	*0.0141*	*0.551*	*0.113*	*0.324*	*0.795*	
Number of countries	*122*	*122*	*96*	*95*	*95*	*95*

Notes: Least Square Dummy Variable (LSDV) models in column (1)(2) include country fixed effects; Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Panel IV regression: Path-defying new entries (log) are instrumented with the 5 year lag of the variable and the log of population

GMM is the Arellano-Bond / Blundell-Bond system estimator

5. Conclusion

In this work we study the evolution of countries comparative advantages in the period 1995-2015 to test and measure how this evolution conforms with the notion of path-dependence, which is a cornerstone of the Product Space framework (Hausmann and Klinger 2007; Hidalgo *et al* 2007). We find evidence that a large share of new products entering countries export baskets is significantly related to the pre-existing comparative advantage or, in other words, that the extent of *diversification* was statistically *constrained*. Our study lends partial empirical support to the hypothesis of path-dependence: we find that countries dynamically acquire a comparative advantage in productions that require a set of productive capabilities that are already available or easily accessible. Our analysis however reveals that a significant share of 'apples' (new product entering the product space of countries) did indeed fall far from the tree: almost half of new products are not significantly related with those already present in countries export basket.

We employ a novel methodology that allows us to rigorously test, using non-parametric techniques, the hypothesis of (non-random) relatedness. To our knowledge, this is the first study that moves beyond a simple description of the dynamics of changes in the bundle of goods produced by countries over time or the use of measures of relatedness can be the result of random processes (as the recent work by Pinheiro et al 2018).

These findings have important implications for industrial and innovation policies. As a matter of fact, governments around the globe are more and more seduced by the Product Space idea of 'latent comparative advantage', which suggests that policy effort should be 'smartly' targeted to those products that are not yet in countries export baskets but are related to it (i.e. small jumps over the Product Space are those that are likely to be effective). Our paper shows that the evolution of comparative advantage of the world economy is characterized by significant jumps away from the diversification pattern postulated by PS. Using such literature too narrowly as a map or guide to identify 'latent comparative advantage' might hence be undesirable since structural transformation has often taken very different and largely unpredictable directions. Furthermore, and most significantly in the eyes of governments, we find that path-defying diversification – i.e. unrelated new entries – are associated to higher growth rates.

Our analysis of the determinants of the share of unrelated new entries shows the importance of public policies: better regulatory and business environments and (although the evidence is less robust) better endowment of human capital allows countries to develop their comparative advantages in new areas of the Product Space.

Although this approach allows us to shed some lights on the observed patterns, admittedly we are not able to fully control for the unobserved heterogeneity at the country level and at sectoral level. It would be interesting to analyse cross-country spatial spillovers in the evolution of comparative advantage as international trade, capital mobility and migration are likely to affect the set of production capabilities that countries have and, in turn, their production baskets. These interesting analyses are left to future research.

References

Agosin Manuel (2009), Export Diversification and Growth in Emerging Economies, *CEPAL Review*, 97: 115–31.

Agosin, Manuel. R., Alvarez, Roberto and Claudio Bravo-Ortega (2012), Determinants of export diversification around the world: 1962–2000, *The World Economy*, 35(3): 295–315.

Bahar, D., Stein E., Wagner R. And S. Rosenow (2017), The Birth and Growth of New Export Clusters: Which Mechanisms Drive Diversification?, *CID Working Paper No. 86, Center for International Development at Harvard University,* September 2017.

Balassa, B. (1965), Trade Liberalisation and "Revealed" Comparative Advantage. *The Manchester School, 33: 99–123.*

Basile R., Parteka A. and R. Pittiglio (2017), Export diversification and economic development: a dynamic spatial data analysis, forthcoming in *Review of International Economics*.

Bjorvatn K. and N. D. Coniglio (2012), Big push or big failure? On the effectiveness of industrialization policies for economic development, *Journal of the Japanese and International Economies*, 26(1): 129-141.

Boschma, R. and Capone, G. (2016), Relatedness and diversification in the European Union (EU-27) and European Neighbourhood Policy countries, *Environment and Planning C: Government and Policy June 2016 34: 617-637.*

Buono, I. and Lalanne, G. (2012), The effect of the Uruguay round on the intensive and extensive margins of trade. *Journal of International Economics*, *86*(2), 269–283.

Cadot, Olivier, Carrère, Celine and Vanessa Strauss-Kahn (2011), Export diversification: What's behind the hump?, *Review of Economics and Statistics*, 93(2): 590–605.

Cadot, Olivier, Carrère, Celine, and Vanessa Strauss-Kahn (2013), Trade diversification, income, and growth: What do we know?, *Journal of Economic Surveys*, 27(4): 790–812.

Castaldi, C., Frenken K. and B. Los (2015), Related variety, Unrelated Variety and Technological Breakthroughs: An analysis of US State-Level Patenting, *Regional Studies,* 49(5): 767-781.

Cook Nathaniel P.S. and Jason Cannon Jones (2015), The African Growth and Opportunity Act (AGOA) and export diversification, *The Journal of International Trade & Economic Development*, 24(7): 947-967.

Coniglio , N.D., Armenise M., Lagravinese R. and D. Vurchio (2018), The pattern of structural change: testing the Product Space framework, *forthcoming Industrial and Corporate Change*.

Dennis, A. and Shepherd, B. (2011), Trade Facilitation and Export Diversification. *The World Economy, 34: 101–122.*

De Benedictis, Luca, Gallegati, Marco, and Massimo Tamberi (2009), Overall trade specialization and economic development: Countries diversify, *Review of World Economics*, 145(1): 37–55.

Di Giovanni J. and A.A. Levchenko (2009), Trade Openness and volatility, *Review of Economics and Statistics*, 91(3): 558-585.

Dingemans Alfonso, César Ross (2012), Free Trade Agreements in Latin America since 1990: an evaluation of export diversification, *CEPAL Review n. 108*, pp. 27-48, December 2012.

Dingemans Alfonso, César Ross (2016), Trying to stay ahead of the curve in Chile's economic development: exploring a way out of the middle-income trap through pragmatic export development, *Development Policy Review*, 34 (5): 643-669;

Dosi G. (1982), Technological paradigms and technological trajectories, *Research Policy*, Vol. 11: 147-162.

Duranton, G. and Overman, H.G. (2005), Testing for Localization Using Micro-Geographic Data, *Review of Economic Studies*, 75: 1077-1106.

Dutt, P., Mihov, I. and Van Zandt, T. (2013), The effect of WTO on the extensive and the intensive margins of trade, *Journal of International Economics*, *91*(2), 204–219.

Feenstra Robert C. and Hong Ma (2014), Trade Facilitation and the Extensive Margin of Exports, *Japanese Economic Review*, 65(2): 158-177.

Fraser Institute (2017), *The Economic Freedom of the World*, Report 2017.

Frenken, K (2007), Entropy statistics and information theory. *In: Hanusch H, Pyka A (eds), The Elgar companion to Neo-Schumpeterian economics. Edward Elgar, Cheltenham, UKand Northampton MA, pp 544–555*

Gozgor Giray and Muhlis Can (2017), Causal Linkages among the Product Diversification of Exports, Economic Globalization and Economic Growth, *Review of Development Economics*, 21(3): 888–908.

Grossman, G. M., and E. Helpman (1991), Trade, knowledge spillovers, and growth, *European Economic Review*, 35(2–3): 517–526.

Hausmann R., Hwang J. and D. Rodrik (2007), What you export matters, *Journal of Economic Growth*, 12(1): 1-25.

Hausmann, R., and B. Klinger (2007), The Structure of the Product Space and the Evolution of Comparative Advantage. Harvard University Center for International Development Working Paper #146.

Hausmann, R. and B. Klinger, (2010), Structural Transformation in Ecuador, Inter-American Development Bank.

Hausmann, R., and D. Rodrik (2003), Economic Development as Self-discovery. *Journal of Development Economics, 72:603–633.*

Helpman, E., Melitz, M. and Rubinstein, Y. (2008), Estimating trade flows: trading partners and trading volumes. *Quarterly Journal of Economics 123 (2), 441–487.*

Hidalgo, C. (2012), Discovering East Africa's Industrial Opportunities, *ArXiv.org papers*

Hidalgo, C., Klinger, B., Barabasi, A. and Hausmann, R. (2007), The Product Space Conditions the Development of Nations, *Science Magazine* 317(5837): 482-487.

Humphreys, M., Sachs, J. and J. Stiglitz (2007), *Escaping the Resource Curse*, Columbia University Press, New York.

Jetter M. and A. Ramirez-Hassan (2015), Want export diversification? Educate the kids first, *Economic Inquiry*, 53(4): 1765-1782.

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

Klinger, B. and D. Lederman (2004), Discovery and development: An empirical exploration of ''new'' products, *World Bank Policy Research Working Paper Series 3872*, Washington, DC: The World Bank.

Lin, J. and Chang, H.-J. (2009), Should Industrial Policy in Developing Countries Conform to Comparative Advantage or Defy it? A Debate between Justin Lin and Ha-Joon Chang, *Development Policy Review*, 27 (5): 483-502.

Makhlouf, Y. N. M. Kellard and D. Vinogradov (2015) Trade openness, export diversification, and political regimes, *Economics Letters*, 36: 25-27.

Mau, K. (2016), Export diversification and income differences reconsidered: the extensive product margin in theory and application, *Review of the World Economy*, 152 (2): 351-381.

Melitz, M. (2003), The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, *Econometrica*, 71(6): 1695–1725.

Minondo, A. (2011), Does comparative advantage explain countries' diversification level?, *Review of World Economics (WeltwirtschaftlichesArchiv), Springer, vol. 147(3), pages 507-526, September.*

Naudé, W. A. (2010), Industrial Policy: Old and New, *WIDER Working Paper 106, Helsinki: UNU-WIDER.*

Neffke, F., Hartog, M., Boschma, R. and Henning, M. (2014), Agents of structural change. The role of firms and entrepreneurs in regional diversification, *Papers in Evolutionary Economic Geography (PEEG) 1410*, Utrecht University, Section of Economic Geography, revised Apr 2014.

Parteka A. and M. Tamberi (2013), What Determines Export Diversification in the Development Process? Empirical Assessment, *The World Economy*, 36: 807-826.

Pinheiro F., Alshamsi A., Hartmann D., Boschma R. and C. Hidalgo (2018), Shooting Low or High: Do Countries Benefit from Entering Unrelated Activities?, *Papers in Evolutionary Economic Geography 18.07*. Uthrecht University, Urban and Regional Research centre Utrecht.

Rose, A. (2006). "Size Really Doesn't Matter: In Search of a National Scale Effect." Journal of the Japanese and International Economies 20, 4 (December 2006): 482-507.

Silverman, B. W. (1986), Density Estimation for Statistics and Data Analysis, *New York: Chapman and Hall.*

Stiglitz, J. (2002), *Globalization and its Discontents*. The Penguin Press.

Table 1 - The determinants of path-dependence: variable descriptions and summary statistics

Variables	Description	Source	Obs.	Mean	St. Dev	Min	Max
(***Dependent variable***) Path-dependent new entries (% share over total new entries; 5 years time periods)	Share of new entries (i.e. products that are exported with RCA > at time *t* that where not exported or exported with a RCA<0.5 at time t - 5) that are 'path-dependent' (see Section 3) over total new entries.	Own calculation based on CEPII-BACI database	2928	49.19	29.22	0.00	100.00
GDP per capita (log)	GDP per capita measured in constant US$ (year 2010; in log)	World Bank (WDI)	2867	8.30	1.53	4.75	11.46
Population (log)	Total population (log of)	World Bank (WDI)	2912	15.53	2.09	9.13	21.01
Trade openness (log)	Total value of export plus imports over GDP (log of)	World Bank (WDI)	2776	4.30	0.64	-3.86	6.28
Trade variety (log)	Herfindal index measuring the degree of export diversification by means of an entropy measure at the 6-digit level (see Boschma & Iammarino 2009). (log of)	Own calculation based on CEPII-BACI database	2902	1.43	0.59	-2.33	2.14
Related varieties (log)	A measure of diversification within sectors. It is computed following Coniglio et al (2017) as an Herfindal index of export diversification and represents the weighted sum of the entropy indicator at the 6-digit level within each 4-digit sectors. (log of)	Own calculation based on CEPII-BACI database	2902	1.51	1.21	-5.13	2.99
Unrelated varieties (log)	A measure of diversification within sectors. We employ an Herfindal index measuring the entropy of the 2-digit distribution of the HS trade classification. (Coniglio et al 2017) (log of)	Own calculation based on CEPII-BACI database	2902	3.53	1.08	-1.77	4.83
Natural resources (% of GDP)	Natural resources rents (% of GDP)	World Bank (WDI)	2868	7.66	12.22	0.00	89.17
FDI inflows (net, % of GDP)	Net inflows of FDI as a share of the GDP	World Bank (WDI)	2828	5.26	16.20	-82.89	451.72
Scientific & tech. publications (log)	Number of scientific and technical journal articles published in the following fields: physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences (log of).	World Bank (WDI)	1936	5.30	3.12	-2.30	12.89
Educational expenditure (% of GDP)	Public expensed in education as a share of GDP.	World Bank (WDI)	1517	14.83	4.91	2.92	44.80
Total new entries (last 5 years; log)	Number of new products which enter the country export basket with a Balassa RCA index higher than 1 from an initial Balassa RCA index lower than 0.5 in a 5-years interval. (log of)	Own calculation based on CEPII-BACI database	2885	3.10	0.84	0.00	5.37
Share of new entries over total export value	Share of exports of total new entries over the value of total exports. (log of)	Own calculation based on CEPII-BACI database	2928	0.10	0.17	0.00	0.99
East Asia and Pacific	Dummy equals 1 if the country belongs to East Asian and Pacific	World Bank (WDI)	2928	0.17	0.38	0.00	1.00
Latin America and the Caribbean	Dummy equals 1 if the country belongs to Latin America and Caribbean Islands.	World Bank (WDI)	2928	0.17	0.38	0.00	1.00
North America	Dummy equals 1 if the country belongs to North America.	World Bank (WDI)	2928	0.02	0.13	0.00	1.00
Middle East and North Africa (MENA)	Dummy equals 1 if the country belongs to Middle East and North Africa.	World Bank (WDI)	2928	0.11	0.32	0.00	1.00
South Asia	Dummy equals 1 if the country belongs to South Asia.	World Bank (WDI)	2928	0.04	0.20	0.00	1.00
SSA	Dummy equals 1 if the country belongs to Sub-Saharan Africa.	World Bank (WDI)	2928	0.22	0.42	0.00	1.00
Europe	Dummy equals 1 if the country belongs to Europe.	World Bank (WDI)	2928	0.26	0.44	0.00	1.00
Low income country	Dummy equals 1 if the country belongs to the income categories defined by the World Bank..	World Bank (WDI)	2895	0.15	0.36	0.00	1.00
Lower-middle income country	Dummy equals 1 if the country belongs to the income categories defined by the World Bank..	World Bank (WDI)	2895	0.26	0.44	0.00	1.00

Upper-middle income country	Dummy equals 1 if the country belongs to the income categories defined by the World Bank..	World Bank (WDI)	2895	0.28	0.45	0.00	1.00
High income country	Dummy equals 1 if the country belongs to the income categories defined by the World Bank..	World Bank (WDI)	2895	0.30	0.46	0.00	1.00
Economic Freedom Index - EFI	Composite Index that measures the degree of economic freedom present in five major areas: [1] Size of Government; [2] Legal System and Security of Property Rights; [3] Sound Money; [4] Freedom to Trade Internationally; [5] Regulation. The index ranges from 10 (max freedom) to 0 (min freedom).	Fraser Institute (2017) Economic freedom of the world (online database)	1540	6.64	1.04	2.93	9.19
Size of Government	Sub component of the EFI. It measure the degree of Government intervention in the economy. The sub-index is composed of several elements. Countries with low levels of government spending as a share of the total, a smaller government enterprise sector, and lower marginal tax rates earn the highest ratings in this area. The index ranges from 10 (max freedom) to 0 (min freedom).	Fraser Institute (2017) Economic freedom of the world (online database)	1538	6.40	1.38	1.75	9.90
Legal system and property rights	Sub component of the EFI. It measures the degree of protection of citizens and property rights and the efficiency of the judicial system. The index ranges from 10 (max freedom) to 0 (min freedom).	Fraser Institute (2017) Economic freedom of the world (online database)	1550	5.31	1.82	0.99	9.28
Sound money	Sub component of the EFI. It measure the degree of consistency of monetary policy (or institutions) with the goal of long-term price stability as well as the freedom to use foreign currencies. The index ranges from 10 (max freedom) to 0 (min freedom).	Fraser Institute (2017) Economic freedom of the world (online database)	1540	7.76	1.75	0.00	9.89
Freedom to trade internationally	Sub component of the EFI. It measure the degree of freedom to trade internationally (for instance tariff and non-tariff barriers to trade) and to move capital across borders (FDI and speculative investments). The index ranges from 10 (max freedom) to 0 (min freedom).	Fraser Institute (2017) Economic freedom of the world (online database)	1532	7.02	1.40	0.00	9.72
Regulation	Sub component of the EFI. It measures the burden associated to bureaucratic procedures and regulations. The index ranges from 10 (max freedom) to 0 (min freedom).	Fraser Institute (2017) Economic freedom of the world (online database)	1540	6.70	1.13	2.70	9.38
Government consumption	Sub component of the 'Size of Government' index. The measure is based on government consumption as a share of total consumption.	Fraser Institute (2017) Economic freedom of the world (online database)	1558	6.03	2.15	0.00	10.00
Transfers and subsidies	Sub component of the 'Size of Government' index. The measure is based on transfers and subsidies as a share of GDP.	Fraser Institute (2017) Economic freedom of the world (online database)	1404	7.65	2.13	0.84	10.00
SOEs and investments	Sub component of the 'Size of Government' index. The measure is based on the extent to which countries use private investment and enterprises rather than government investment and firms (State Owned Enterprises) to direct resources.	Fraser Institute (2017) Economic freedom of the world (online database)	1578	6.00	3.23	0.00	10.00
Top marginal tax rate	Sub component of the 'Size of Government' index. The measure is based on the top marginal income and payroll tax rate and the income threshold at which these rates begin to apply.	Fraser Institute (2017) Economic freedom of the world (online database)	1341	5.97	2.44	0.00	10.00
Non tariff trade barriers	Sub component of the 'Freedom to trade internationally' index. The measure is based on the extent of non-tariff barriers to international trade.	Fraser Institute (2017) Economic freedom of the world (online database)	1217	6.02	1.44	1.83	9.69

Credit market regulations	Sub component of the 'Regulation' index. The measure is based on a set of conditions reflecting the extent of Government intervention in the domestic credit market (bank ownership, interest rate controls and other restrictions to private credit.	Fraser Institute (2017) Economic freedom of the world (online database)	1544	8.07	1.68	0.00	10.00
Ownership of banks	Sub component of the 'Regulation' index. The measure is based on the extent to which the banking industry is privately owned.	Fraser Institute (2017) Economic freedom of the world (online database)	1482	6.96	3.19	0.00	10.00
Business regulations	Sub component of the 'Regulation' index. The measure is based on a set of regulations and constraints that increase the cost of doing business for private firms (bureaucratic costs, cost of starting a business, bribes and corruption, licensing restrictions and tax compliance).	Fraser Institute (2017) Economic freedom of the world (online database)	1327	6.03	1.26	2.18	9.50

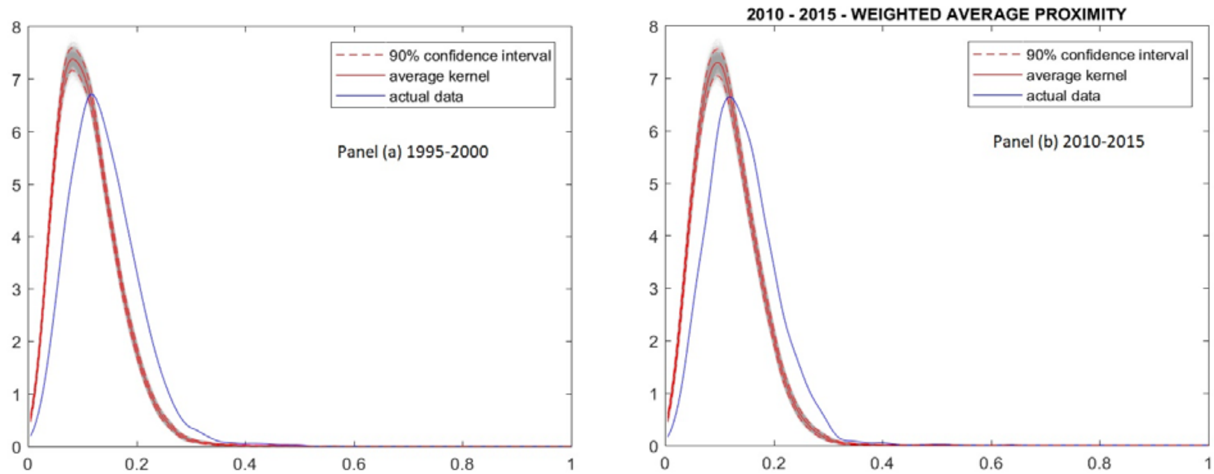
Path-dependence, radical changes and growth: variable descriptions and summary statistics

Variables	Description	Source	Obs.	Mean	St. Dev	Min	Max
Dependent variable: GDP per capita growth, yearly average (5-year panel; % change)	5-year average growth rate of GDP per capita (constant 2010 US$)	World Bank (WDI)	694	2.54	3.08	-8.21	17.12
Non path-dependent new entries (log)	Share of new entries (i.e. products that are exported with RCA > at time *t* that where not exported or exported with a RCA<0.5 at time t - 5) that are non 'path-dependent' (see Section 3) over total new entries. (i.e. radical new entries)	Own calculation based on CEPII-BACI database	714	3.66	0.96	-2.05	4.61
Initial GDP per capita (log)	GDP per capita measured in constant US$ (year 2010; in log)	World Bank (WDI)	699	8.32	1.53	5.14	11.39
Initial Export sophisticatedness, ExpY (log)	Measure of the sophisticatedness level associated to the country export basket. *See details in Hausmann & al (2007).* (in log)	Own calculation based on CEPII-BACI database	710	9.27	0.33	8.17	10.53
Initial Economic Freedom Index	Composite Index that measures the degree of economic freedom present in five major areas: [1] Size of Government; [2] Legal System and Security of Property Rights; [3] Sound Money; [4] Freedom to Trade Internationally; [5] Regulation. The index ranges from 10 (max freedom) to 0 (min freedom).	Fraser Institute (2017) Economic freedom of the world (online database)	513	6.54	1.13	2.93	9.11
Initial Scientific & tech. publications (log)	Number of scientific and technical journal articles published in the following fields: physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences (log of).	World Bank (WDI)	515	5.37	3.13	-1.61	12.89

Appendix 1 – Kernel estimate with alternative measures of relatedness

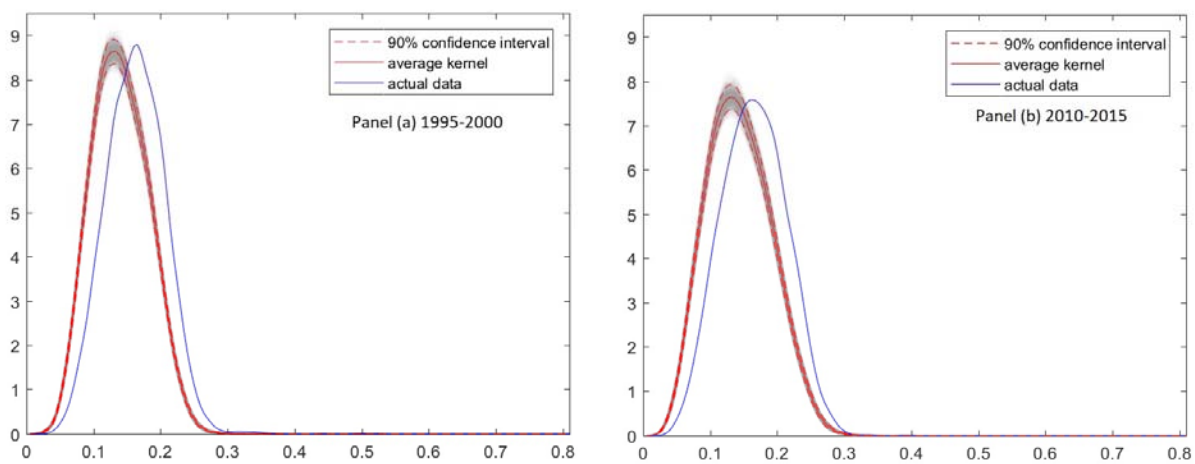
WEIGHTED AVERAGE PROXIMITY

Distribution of proximities between new export products and pre-existing export baskets: actual data versus random counterfactual (5-years intervals 1995-2000 and 2010-2015)



AVERAGE PROXIMITY

Distribution of proximities between new export products and pre-existing export baskets: actual data versus random counterfactual (5-years intervals 1995-2000 and 2010-2015)



Appendix 2 – List of countries included in the growth regressions (Tables 5 and 6)

Low and middle income countries		High income countries
Algeria	Madagascar	Australia
Angola	Malawi	Austria
Argentina	Malaysia	Bahamas
Armenia	Mali	Bahrain
Bangladesh	Mauritania	Belgium-Luxembourg
Belize	Mauritius	Brunei Darussalam
Benin	Mexico	Canada
Bolivia	Mongolia	Chile
Brazil	Montenegro	China Hong Kong SAR
Bulgaria	Morocco	Cyprus
Burkina Faso	Mozambique	Czech Rep.
Burundi	Myanmar	Denmark
Cabo Verde	Nepal	Estonia
Cambodia	Nicaragua	Finland
Cameroon	Niger	France
Central African Rep.	Nigeria	Germany
Chad	Pakistan	Greece
China	Panama	Hungary
Colombia	Papua New Guinea	Iceland
Congo	Paraguay	Ireland
Costa Rica	Peru	Israel
Croatia	Philippines	Italy
Côte dIvoire	Rep. of Moldova	Japan
Dem. Rep. of the Congo	Romania	Kuwait
Dominican Rep.	Russian Federation	Latvia
Ecuador	Rwanda	Lithuania
Egypt	Senegal	Malta
El Salvador	Serbia	Netherlands
Ethiopia	Sierra Leone	New Zealand
Fiji	South Africa	Norway
Gabon	Sri Lanka	Oman
Gambia	Suriname	Poland
Georgia	TFYR of Macedonia	Portugal
Ghana	Tajikistan	Qatar
Guatemala	Thailand	Rep. of Korea
Guinea-Bissau	Timor-Leste	Saudi Arabia
Guyana	Togo	Singapore
Haiti	Tunisia	Slovakia
Honduras	Turkey	Slovenia
India	Uganda	Spain
Indonesia	Ukraine	Sweden
Iran	United Rep. of Tanzania	Switzerland
Jamaica	Venezuela	Trinidad and Tobago
Jordan	Viet Nam	USA
Kazakhstan	Yemen	United Arab Emirates
Kenya	Zambia	United Kingdom
Kyrgyzstan	Zimbabwe	Uruguay
Lebanon		

Appendix 3. One-sided Kolmogorov-Smirnov Test for First Order Stochastic Dominance: comparing New Entries' relatedness distributions with randomly generated counterfactual distributions (all countries; five-year time intervals from 1995 to 2015)

New entry identification strategy: RCAt0<0.5 and RCAt1>=1.						
Time interval	**Maximum proximity**		**Weighted average proximity**		**Average proximity**	
	Dn	**p-value**	**Dn**	**p-value**	**Dn**	**p-value**
1995 - 2000	0.106	0.000	0.177	0.000	0.172	0.000
1996 - 2001	0.106	0.000	0.176	0.000	0.161	0.000
1997 - 2002	0.110	0.000	0.175	0.000	0.163	0.000
1998 - 2003	0.108	0.000	0.192	0.000	0.175	0.000
1999 - 2004	0.122	0.000	0.189	0.000	0.175	0.000
2000 - 2005	0.116	0.000	0.181	0.000	0.167	0.000
2001 - 2006	0.118	0.000	0.163	0.000	0.159	0.000
2002 - 2007	0.105	0.000	0.167	0.000	0.159	0.000
2003 - 2008	0.118	0.000	0.163	0.000	0.169	0.000
2004 - 2009	0.120	0.000	0.142	0.000	0.151	0.000
2005 - 2010	0.118	0.000	0.164	0.000	0.166	0.000
2006 - 2011	0.123	0.000	0.168	0.000	0.170	0.000
2007 - 2012	0.129	0.000	0.167	0.000	0.162	0.000
2008 - 2013	0.111	0.000	0.158	0.000	0.152	0.000
2009 - 2014	0.109	0.000	0.166	0.000	0.147	0.000
2010 - 2015	0.109	0.000	0.168	0.000	0.152	0.000

Note: for all three definitions of relatedness we cannot reject the null hypothesis and we conclude that cumulative distribution of randomly generated data stochastically dominates that of actual data (i.e. higher and statistically significant concentration of the latter at higher proximities.