The pattern of structural change: testing the Product Space framework

Nicola D. Coniglio, Raffaele Lagravinese, Davide Vurchio, Massimo Armenise

Utrecht University
Urban & Regional research centre Utrecht

The pattern of structural change:
testing the Product Space framework

Nicola D. Coniglio[1], Raffaele Lagravinese, Davide Vurchio, Massimo Armenise

Department of Economics and Mathematical Methods, University of Bari "A.Moro"

Abstract

The set of available local 'capabilities' determines what an economy produces today (its static comparative advantage) and, at the same time, defines the trajectories that the process of structural change may take in the future. The Product Space (PS) framework developed in recent seminal works by economists and physicists suggests that path dependence characterizes the evolution of the production basket (Hausmann and Klinger, 2007; Hidalgo *et al.* 2007). These authors represent economies as sets of productive capabilities that can be combined in different ways to produce different products. Countries progressively change their production baskets and move towards goods that require capabilities that are already available; on the contrary radical structural change rarely happens. In this paper, we analyse the evolution over time of the production baskets in 107 Italian provinces (NUTS 3) and perform the first test on the PS hypothesis of path dependence. We investigate whether new products entering the provincial production baskets are non-randomly related to initial production baskets. We confirm the general tendency of path dependence, but highlight at the same time that a sizable share of 'new products' are an exception to this general pattern. These 'random entries' over the PS are particularly interesting for industrial policy since they represent radical deviations from the initial comparative advantage. In the final part of the paper, we investigate using parametric analysis the product and provincial characteristics that determine these deviations from the PS pattern.

Keywords: product space*;* structural change; trade specialisation; path-dependence; Italy

JEL Classifications: F1, R3, R11, R13

---

[1] Corresponding Author: nicoladaniele.coniglio@uniba.it.
Address: Largo Abbazia Santa Scolastica, 70124 Bari, Italy

1. Introduction

Economies evolve over time in a dynamic process in which available resources are combined in order to produce a bundle of products (production basket) which reflects the comparative advantages of those economies. The process of structural change may take different paths according to whether marginal or radical changes in the composition of production baskets occur over time.

A new wave of intellectual effort in the analysis of the process of economic development has placed structural change at the core of the policy debate (McMillian and Rodrik 2011; Spence 2011, Stiglitz et al. 2013). As in early contributions (Kuznets, 1966), structural change is seen as a pre-condition for sustained economic growth and development since economic wealth strictly depends on the economic structure and sophistication of the production basket.[2] In particular, the Product Space (PS) framework developed in recent seminal works by economists and physicists suggests that the evolution of the production basket is strongly characterized by path dependence (Hausmann and Klinger, 2007; Hidalgo et al. 2007). These authors represent economies as sets of productive capabilities that can be combined in different ways to produce different products. Countries progressively change their production baskets and move towards goods that require capabilities that are already at their disposal or easily obtained; on the contrary, radical structural change rarely happens.[3]

Since capabilities cannot be easily identified, measured and observed, these authors employ an 'agnostic approach' and use an outcome-based measure which relies on the idea that if two goods are 'related' (*i.e.* produced and exported in tandem), they use production factors that are 'common'. Unrelated goods, *i.e.* those goods that are unlikely to be produced and exported by the same country, do not share a similar set of productive factors.

---

[2] An important difference between this wave of 'structural economics' and the early one is rooted in the role of the State and normative implications in general. The first wave of structural economics was based on a firm belief that structural differences were essentially the result of market failures which required pervasive and often highly distortionary Government interventions. This '*dirigiste dogma*' led to the widespread adoption of quantitative restrictions to international trade flows and the heavy use of currency manipulations which caused several crises that paved the way to another extreme, the 'market dogma'. The new wave of structural economics can be seen as a 'market-State' blend that is perfectly represented by the words of one of its main exponents, Justin Yifu Lin "*the market should be the basic mechanism for resource allocation, but that government must play an active role in coordinating investments for industrial upgrading and diversification and in compensating for externalities generated by first movers in the dynamic growth process*" (Lin, 2012).

[3] Boschma et al (2015) using a similar approach introduce the concept of *technological space* and show that the acquisition of new technological capabilities in 366 US cities is more likely if related technologies have already been acquired. This result suggests that a strong path-dependence not only affects the introduction of new products but also characterizes the development of local productive capabilities.

The 'product space' (PS) was first presented in Hausmann and Klinger (2007) and Hidalgo *et al* (2007) as a network of relatedness between 774 globally produced and exported products. The Product Space has been represented effectively using a map (reported in *Figure 1*) of global production in which each node represents a product and connections between nodes represent the degree of proximity between them.[4] The authors assert that goods entering a country's export basket are those highly connected with the set of products that were previously exported. In addition, Hidalgo *et al* (2007) argue that where a country's export basket is 'located' in the product space matters for economic development. As new industries/specialization develop from existing ones, countries that produce goods that are better connected are more likely to develop more sophisticated goods. On the contrary, countries specialized in goods that are located in the periphery of the product space are more likely to be trapped in development 'dead corners' and face higher difficulties in kick-starting new more complex and sophisticated industries.[5]

Figure 1: Hidalgo et al (2007) - representation of the network of relatedness between goods

---
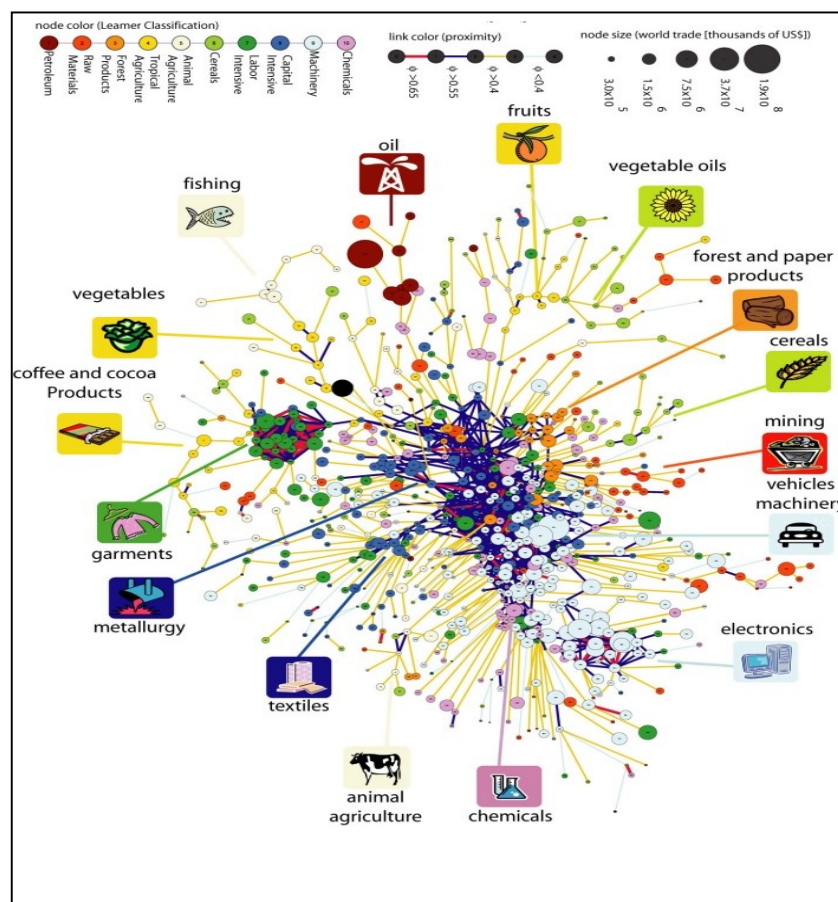
[4] The proximity between each couple of goods is given by the minimum of the pairwise conditional probability of being co-exported. In other words, products are connected or related if they tend to be exported by the same economies.
[5] Several contributions, starting from the work of Hausmann et al (2007), have shown that 'what you produce matters!', the complexity and sophisticatedness of what an economy produces enhances its future growth.

According to many observers, these recent contributions add new 'weapons' to the arsenal of industrial policies since the network of relatedness provides a guide for policymakers in terms of which products/sectors are likely to be successfully developed in a country or region (latent comparative advantage). In fact, most industrial policies that aim to implement ambitious projects have failed because of the existence of capability constraints to 'big leaps'. In this light, the product space poses limits to overoptimistic and 'comparative advantage defying' policies and suggests a step-by-step approach featuring 'small leaps' toward these products where countries may have a latent comparative advantage.

Although the product space framework has spurred considerable interest among the academia[6] and policymakers[7], to date, to the best of our knowledge, there is no systematic empirical test which shows whether the pattern of specialization of countries or regions follows its predictions.

The aim of the analysis performed in this work is to fill this gap by providing a new methodological approach for testing the validity of one of the key hypotheses of the product space: specialization in new products does not follow a random process, but is likely to occur in products that are strongly related (or connected) to the ones that are already produced.

We develop a 'dart-board' approach which allows us to compare the actual short-term evolution of the export baskets in 107 Italian provinces (NUTS 3 classification) with randomly generated counterfactuals. After presenting the methodology, which can be easily applied to other countries/regions in the world, we show 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 do often occur. In order to assess the impact of the recent crisis, we identify two periods: i) pre-crisis, 2002-2006; ii) crisis 2007-2011. Interestingly, we find evidence in both periods of a large heterogeneity in terms of frequency of these 'big leaps' over the PS both across provinces (NUTS 3 areas) and across sectors (HS 6 trade classification).[8]

From a policy perspective, these deviations from the hypothesis of path-dependence are the most interesting ones in our opinion. In fact, the development of products that are unrelated to the pre-

---

[6] Cfr. next section for details.

[7] Industrial policy is back in the agenda of many countries around the world. The framework developed by Hausmann et al. 2007 has received a great deal of attention from several countries which are seeking the support of experts – for instance, the Centre for International Development, CID, based at Harvard University and led by Ricardo Hausmann – to design their industrial strategies. The list of countries inspired by this approach is expanding and includes Albania, Colombia and Mexico, among others.

[8] We find that the share of new goods that is statistically unrelated to the initial export basket ranges from a minimum of 17% in the province of Isernia in the crisis period to a maximum of 75% in the province of Siracusa in the pre-crisis period. Such heterogeneity is also confirmed between sectors: only 27% of new entries belonging to the textile sector are found to be unrelated against 83% for the mineral sector.

existing export basket signals the ability of the economic system to combine old and new capabilities in a way that allows production to be diversified away from the static comparative advantage. As argued in Castaldi et al (2015) in the context of regional innovation, technological breakthroughs are the result of the combination of knowledge from 'unrelated' technological capabilities and allows economies to follow new technological trajectories (Dosi 1982).

Some successful and rather emblematic 'jumps' over the PS network have been hotly discussed in the development literature. The rise of the aircraft industry in Brazil and the automotive industry in Korea are notable examples.[9]The rise of the 64 Kbit DRAM sector in Korea is another emblematic case which defeats the gravity of the product space (Lin and Chang, 2009).[10] In both cases, the role played by public actors in supporting industrial competitiveness has been determinant. The product space framework is not able to explain why these jumps occur; quite the contrary, the framework predicts small-distance and gradual jumps toward related goods.

In the last part of our work, we show that there is a positive relationship between provincial economic performance and the share of new products that are unrelated to the initial composition of the export basket (in particular before the crisis period). We take the analysis a step further and investigate – using probit models – which provincial features are associated with the likelihood of observing these more radical structural changes. Our results show that the diversification of provincial productions away from the initial comparative advantage is more likely the more sophisticated the initial production basket is, the higher the mix of unrelated varieties produced and the more open and skilled intensive provincial economies are.[11]

The remainder of the paper is structured as follows. In Section 2 we discuss recent contributions to the economic literature on Product Space. Then, in Section 3 we describe the data and the methodology used for computing the econometric strategy to test the product space theory on Italian provinces. In Section 4, we present the main evidence of analysis and investigate the determinants of path-dependency in the evolution of Product Space in Italy distinguishing the pre-crisis period from crisis one. Finally, we conclude with some policy remarks.

---

[9] Cfr. Altenburg (2011)

[10] According to H.-J. Chang, such a successful production specialization decision supported by active industrial policy in Korea is the proof that defying a country's comparative advantage (in that period the economy was mainly specialized in the production of labour intensive goods) allows 'learnable-by-doing' competences that then made Korea one of the major producers of electronic components to be developed. On the other hand, J.Y. Lin asserts that the kind of electronic components produced at that time in Korea did not require very high skills since 64 Kbit DRAM was no longer at the technology frontier (Lin and Chang, 2009). Transposing these two views to the context of the network of relatedness between goods implies either that an economy is able – under certain conditions – to specialize in products that are not very proximal to the pre-existing export basket or that the product space is dynamic and that links connecting nodes change over time.

[11] Using US patent citation data, Castaldi et al (2015) find that technological breakthrough – i.e. radical innovations – are more likely to happen in US states endowed with a large set of unrelated varieties.

2. Specialization and path-dependence: a brief review of the Product Space framework

The Product Space framework briefly outlined above provides a powerful prediction of path-dependence in the evolution of countries or regional specialization over time. In fact, the inclusion of new products in the export basket of an economy is likely to be strictly related to the pre-existing specialization. The economic intuition is the following: products that are closely connected in the PS (*i.e.* high degree of proximity) 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 also develop a comparative advantage in products requiring the same set of capabilities.

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 specialization – measured by revealed comparative advantage - as a proxy of production specialization and analyse the pattern of specialization across the PS over time.

An important contribution made by this approach is the evidence that countries at a different level of development tend to be positioned differently in the PS. While industrialized countries are mainly specialized in the production of 'central goods' *i.e.* goods with higher average connections to others and higher sophisticatedness[12], low income countries have most of their export baskets located in the periphery of the Product Space. 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) respectively mostly consist of peripheral products and highlight a rather strong persistence of position on the PS over time. Felipe et al. (2013a; 2013b) perform single country analyses on a long-term perspective (from the 1960s to the 2000s) for two important emerging economies, China and India. Their works suggests that the process of development in these two countries is accompanied by a gradual and continuous increase in export sophisticatedness.

Further studies based on the product space approach have focused attention 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 actual diversification level. 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

---

[12] The degree of sophisticatedness of a product is generally proxied by the ProdY index originally presented in Hausmann *et al* (2007). The ProdY index represents the productivity level associated with the production of a certain product and is defined as in Appendix A2.

developed revealed comparative advantage at time *t* in products related to those in which they were already specialized at time *t-3 / t-5*.[13]

So far only a few contributions have analysed the pattern of trade diversification at the sub-national level which is likely to be the most significant since capabilities are lumpy across space and have a strong local dimension. Using US States data in the period 2002-2012, Donoso and Martin (2016) show that only the local capabilities have a role in the path-dependence process of industrial structure dynamics whereas the industrial structure at the national level has a negative effect on States' export diversification. The authors also find that the higher internal migration, firm cluster strength and R&D spending over GDP are, the stronger the effect of current structure on the probability of diversifying a State's production. The importance of looking at sub-national areas is confirmed by the contribution of Boschma *et al* (2013). The authors show that during the period 1988-2008, Spanish regions diversified into those new sectors that were related to the existing set of industries. Moreover, Boschma *et al* (2013) find strong evidence that capabilities available at the regional level played a larger role than capabilities available at the country level in the emergence and development of new industries

A small but growing number of works have investigated the path dependence of structural transformation using firm level data (Neffke *et al* 2011; Cirera *et al* 2012; Lo Turco and Maggioni 2016). Using plant-level data for 70 Swedish regions in the period 1962-2002, Neffke *et al* (2011) find evidence of path-dependence in the evolution of long-term production diversification since industries that are technologically related to pre-existing ones have a higher probability of entering the region's production portfolio whereas unrelated ones have a higher probability of exiting. Analogous results are found by Cirera *et al* (2012) in Brazil for the period 2000-2009. The authors document that trade diversification mostly stems from related sectors. Diversification in sectors that are unrelated to the pre-existing production basket is limited and mainly concerns vertically integrated firms which specialize in one or few stages in a specific value chain.

Lo Turco and Maggioni (2016) using Turkish firm-level data show that the introduction of new products by manufacturing firms is significantly higher if related products are produced by the same firm or by other firms in the affected province. In this study, relatedness is also measured using

---

[13] The authors use the 'density' measure developed by Hidalgo et al (2007) in their parametric analysis of the probability that (new) goods enter the export basket of a country, computed as the average proximity of a new potential product to a country's current productive capability.

'density' variables *à la* Hidalgo *et al* (2007). The local set of available capabilities is important – although less than internal (firm-specific) resources – in explaining what firms produce.[14]

All these studies confirm the importance of the set of available local capabilities in guiding the evolution of the comparative advantage of countries and/or regions and in shaping the introduction of new goods and products by firms. It is easier to produce new goods requiring capabilities that are already present in the local economy rather than those requiring different sets of capabilities. In other words, according to the capability approach, jumps over the Product Space are unlikely and the process of structural change – measured as the evolution of the basket exported with revealed comparative advantage – is path-dependent. These studies also suggest that it is important to consider the geographical dimension of the changes in the production basket since capabilities are not uniformly distributed within a country.

Although the contributions described above document a large extent of path-dependence in the evolution of the production basket, these studies cannot be considered as a formal test of the PS framework since they do not discriminate between the relatedness due to shared production capabilities (as the framework suggests) and spurious relatedness which is the result of a random process.

Our analysis – using different definitions of 'relatedness' and different 'new entry' identification methods – aims at testing whether new products in Italian provinces are non-randomly related to those previously exported with revealed comparative advantage. While other studies employ measures of 'density'[15] or 'open-forest' indexes *à la* Hausmann and Klinger (2007) for assessing path-dependence, we develop a test for inferring whether new entries in the export basket are related in a statistically significant way compared with randomly generated ones. The methodology developed in the paper allows us to measure the extent to which structural change deviates from the hypothesis of path dependence and, in turn, to shed light on the provincial characteristics that are significantly associated with 'big leaps' over the PS (*i.e.* rather radical changes in the composition of the production basket).

---

[14] Using Chinese firm-level data, Poncet and Starosta de Waldemar (2013) show that 'domestic capabilities' matter not only for explaining what firms produce, but also for the growth enhancement effects of new products and new technologies.

[15] Previous studies (e.g. Boschma *et al*, 2013; Boschma and Capone, 2016; Lo Turco and Maggioni, 2016; Donoso and Martin, 2016) have used measures of 'density' as a predictor of the entry of a given product that was not previously exported.

3. Data and Methodology

3.1. A general test of path-dependence: a dart-board approach

What countries/regions produce and export changes over time as new products enter the production baskets. In this section our aim is to test - using Italian NUTS 3 data (provinces) - whether new products that enter the export baskets at time t_1 are related *à la* Hidalgo et al (2007) with the pre-existing comparative advantage at time t_0. As in the seminal contribution of these authors, the relatedness between any two products is measured using their proximity in the PS, *i.e.* the minimum of the pairwise conditional probability of being co-exported.

We develop a 'dart-board approach' in order to test the non-randomness of the development of provincial production space over time during the period 2002-2011[16]. Given the important role that the crisis played in re-shaping the provincial pattern of trade (see Coniglio *et al* 2016), we consider separately the pre-crisis period (2002-2006) from the crisis one (2007-2011)[17]. For this purpose, we select two base years as t_0 (2002 and 2007) and two as t_1 (2006 and 2011). We allow for a lag of 4 years between t_0 and t_1 in order to investigate short-term changes in the structure of provincial production.[18] Data on Italian provinces' exports are provided by the A.D.ELE. Laboratory at 6-digit whereas data on country exports used to create proximity matrixes are obtained from the UNCOMTRADE dataset.

In the first step of our analysis we need to define 'new entries' as those goods that are not part of the production basket at time t_0 and enter the provincial 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; *i.e.* the ratio between the provincial export share and the world export share for each good is higher than unity. More precisely, in our study a new entry is a product with a RCA lower than 0.5 at t_0 and higher than unity at t_1.[19] For each province $k \in K$ we identify the set of new entries $n \in N_k$ in both sub-periods.

---

[16] Our approach has some similarity with the one employed by Duranton and Overman (2005) to measure the non-randomness of the geographical concentration of industrial plants in the UK.

[17] Between the two sub-periods, four new provinces have been formed (in 2005), hence the total number of provinces used in the analysis is 103 and 107, respectively. We do not have data in both periods for three provinces which are excluded from the analysis (Barletta-Andria-Trani, Fermo and Monza-Brianza).

[18] For robustness, different base and term years have been used and are available upon request from the authors. Note that the split of the two sub-periods reported in the paper is also preferred because it allows us to use the same nomenclature for international and national trade statistics (Harmonized System revisions H2 and H3 have been issued in 2002 and 2007 respectively) between t_0 and t_1 hence avoid the use of correspondence tables that may result in a less precise conversion of the data.

[19] Since this choice of RCA thresholds is arbitrary, for robustness we identify a new entry using three additional alternative thresholds. We use one definition of a new entry that is less restrictive that the one presented in the paper ($RCA_{t_0} < 1$ and $RCA_{t_1} \geq 1$) and two definitions that are more restrictive, respectively RCA_{t_0} lower than 0.1 and lower than 0.2 and

In the second step, we compute - as in Hidalgo et al (2007) - a MxM matrix containing the relatedness measures between any pair of goods ij exported in the world, $(i,j \in W_{t_1}$ where W is the set of goods exported in year t_1, 2006 and 2011, respectively).[20] More precisely the matrix is built as follows. For each country in the World, c, and for each of the two years, we denote x_{ic} as 1 if country c has a revealed comparative advantage in the production of good i and 0 otherwise:

$$x_{ic} = \begin{cases} 1 \ if \ RCA_{ic} > 1 \\ 0 \ otherwise \end{cases} \tag{1}$$

where RCA_{ic} 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 distances between each couple of goods i and j as the minimum of the pairwise conditional probability of being co-exported:

$$\varphi_{ij} = \min\{P(x_i|x_j), P(x_j|x_i)\} \tag{2}$$

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

In the third step, we denote with $B_{k,t}$ the set of goods exported with RCA by province k at time t. We then define $D_{i,k}$, a MxK matrix of relatedness measures between the new N_k products (entering the export basket between t_0 and t_1) and the pre-existing export basket, for each province $k \in K$, as follows:

$$D_{i,k} = \begin{cases} d_{ik}(\varphi_{i,j}) \ when \ j \in B_{k,t_0}, i \in N_k \\ no \ value \end{cases} \tag{3}$$

where $d_{ik}(\varphi_{i,j})$ is a measure of proximity of the new product i with the pre-existing export basket in province k. Given that the export basket at time t_0 typically consists of several goods belonging to a variety of sectors (and hence positioned in different branches of the PS), the concept of relatedness can be specified either in an absolute term (*i.e.* the distance in the product space of each new product i with each of all the products already in the export basket) or in a relative term (*i.e.* the distance of

$RCA_{t_1} \geq 1$. These range from 8,568 'new entries' in the pre-crisis period for the most restrictive definition to 18,656 'new entries' in the crisis period for the less restrictive definition. The results are qualitatively similar and are available upon request from the authors.

[20] We obtain a 5,222-by-5,222 and a 5,050-by-5,050 matrix for 2006 and 2011 respectively. Note that we use a more detailed network of relatedness than the original version (Hidalgo *et al* 2007, 774 goods in the SITC rev.4 Nomenclature) which, in our opinion, allows us to obtain a more precise representation of the evolution of export baskets.

new products relative to the overall pre-existing basket). For this reason, we employ three alternative measures of relatedness:

Maximum proximity: $\qquad d_{ik,m}(\varphi_{i,j}) = max(\varphi_{i,j})$ (4)

Average proximity: $\qquad d_{ik,a}(\varphi_{i,j}) = \frac{\Sigma_j \varphi_{i,j}}{J_k}$ (5)

Weighted Average proximity: $\qquad d_{ik,wa}(\varphi_{i,j}) = \frac{\Sigma_j \frac{exp_{jk,t_0}}{\Sigma_j exp_{jk,t_0}} \varphi_{i,j}}{J_k}$ (6)

with J_k being the number of goods in the export basket of province k at t_0. Equation (4) represents relatedness of a new entrant product i with the set of products J_k exported with revealed comparative advantage at t_0 as the maximum value among the proximities between i and all $j \in J_k$; in other words, distance is measured with respect to the closest product in the provincial product space that is already exported. As an alternative, the measure computed using equation (5) identifies the distance as the average proximity between good n and all the goods $j \in J_k$. Finally, equation (6) computes distance as the weighted average proximity with weights represented by the export share of goods in J_k at time t_0.

The relatedness (or unrelatedness) of new products can be easily appreciated with a simple graphical example. *Figure 2* represents a subset of the export basket of a fictitious province at t_0 using a simplified representation of the PS in which node **A** identifies a new product that enters the export basket at t_1. In *Panel A*, the maximum proximity measure (eq. 4) is employed, thus the relevant pre-existing product is the closest one. This measure represents an absolute dimension of proximity which probably better captures the role of available capabilities in shaping the path-dependence of product diversification. In *Panel B*, the relatedness takes into consideration all products exported with revealed comparative advantage in time t_0 (respectively eq. 5 if no weights are employed and 6 otherwise). These two measures can provide information on the relative degree of proximity of new entries.

Figure 1: Export basket subset of a fictitious province
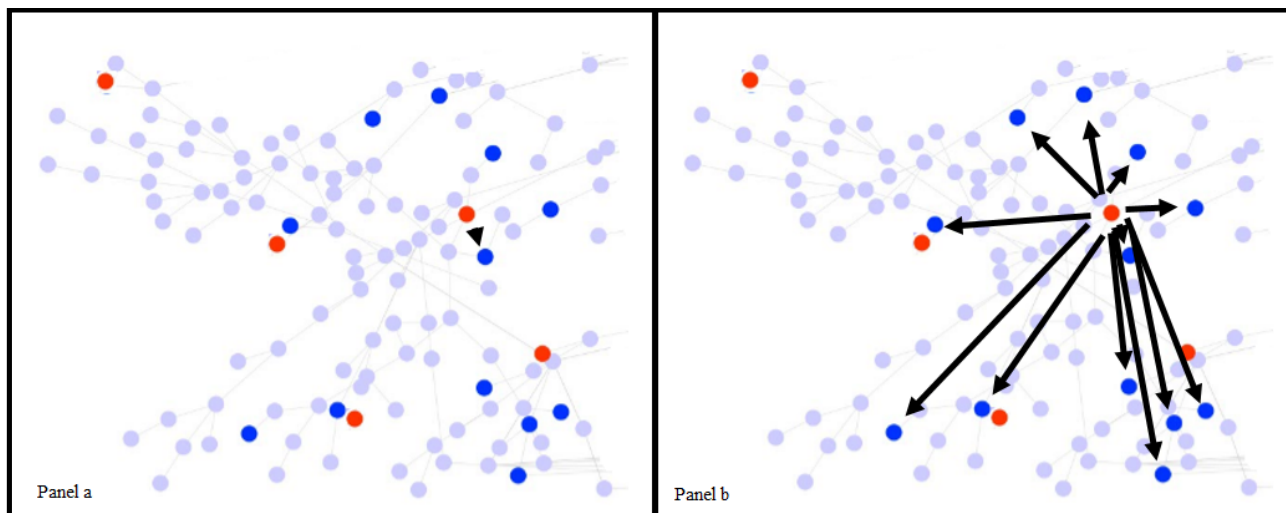
Note: Blue dots represent products exported with RCA at time t0 while red ones represent new entrants at time t1.

Once new entries and proximities have been defined, the subsequent step is to perform a test that allows us to reject the hypothesis that the new entries in province k are randomly related to the initial export basket of that province. Our idea is that if new entries are driven by path-dependence– as the product space framework asserts – we should observe that the distribution of relatedness based on the observed new entries (N_k) significantly differs from that of randomly generated new entries of identical size. By drawing a parallel between the product space and a dartboard, each new entry is equivalent to a dart and will be localized in a given place on the board. The actual data will tell us where the $n_k \in N_k$ darts are localized. Our test is based on the random draw – for 1,000 times and for each k province– of a number of darts equal to N_k and compares the resulting patterns with that derived from the actual data. We then reject the null hypothesis of random localization on the board when the actual draws produce a pattern that is different in a statistically significant way from the random counterfactual.

In order to run our test, we implement a Kernel smoothed density estimation of relatedness of new entries in provincial export baskets.[21]Like Duranton and Overman (2005), we estimate a smoothed Kernel density function of relatedness for any level of proximity, d, defined as:

---

[21] A vector of distances for each of the four definitions of new entries and for each alternative measure of relatedness is created in order to ensure the robustness of our results to the definition of these two key elements. In the paper, we only present, for the sake of brevity, the results for one definition of a new entry ($RCA_{t_0} < 0.5 \; and \; RCA_{t_1} \geq 1$).

$$\widehat{K}(d) \equiv \frac{1}{\left(\sum_{i=1}^{M} \sum_{k=1}^{K} I_{i,k}\right)h} \sum_{i=1}^{M} \sum_{k=1}^{K} f\left(\frac{d-d_{i,k}}{h}\right) \tag{7}$$

with densities calculated non-parametrically using a Gaussian Kernel function with bandwidth h set according to Silverman's optimal rule of thumb (Silverman, 1986), where $d_{i,k}$ is measured using one of the three alternative definitions of relatedness reported in eqs. (4)-(6) while $I_{i,k}$ is a product by province matrix of size MxK which has values of 1 for each new entrant product for each province and 0 otherwise. Therefore, $\sum_{i=1}^{M} \sum_{k=1}^{K} I_{i,k}$ is equal to the total number of new entries across all provinces from t_0 to t_1.

Finally, we build a counterfactual distribution of relatedness and compare it with the actual one obtained from eq. (7). The counterfactual density function is based on simulated relatedness computed from 1,000 random draws of size $\sum_{i=1}^{M} \sum_{k=1}^{K} I_{i,k}$ (total number of entries).[22] More specifically, for any value of proximity $d \in [0,1]$, we rank the simulated kernel density values and obtain the 95th percentiles in order to identify a 5% confidence threshold. If the kernel density estimation of our actual data lies above the counterfactual's threshold, for those proximities d we can reject the hypothesis of random relatedness. On the contrary, when the kernel density values of actual data fall under the counterfactual's threshold, we cannot reject the null hypothesis of random relatedness. Following our discussion in Section 2, the prediction of the PS framework is a higher concentration of actual kernel densities at higher values of proximities vis-à-vis the simulated ones (non-random relatedness).

3.2. *New entry in the export basket: a single product's test of path-dependence*

The methodology explained above allows us to test the 'aggregate' (or general) compliance of the evolution of provincial specialization with the predictions of the product space framework. In this section we describe a Monte-Carlo methodology which allows us to shift the analysis to each of the new products that enters provinces' export baskets at time t_1. For each Italian province $k \in K$, we randomly draw, from the set of products that were not present in the export basket at time t_0[23], a

---

[22] In every simulation, for each province we randomly draw a number of new entries from the products not in the basket at time t_0 which is identical to the number of effective ones. In other words, our counterfactual exercise takes explicit account of the province-specific distribution of new entries.
[23] In other words, we draw random samples from all goods $i \in \left[W_{t_1} - B_{k,t_0}\right]$, where W is the set of all goods exported in the world at time t_1 with $t_1 = 2006, 2011$.

number of products equal to the actual number of new goods that enter in the export basket at time t_1, J_k and compute proximities using eqs. (4) to (6) and generate an average value per draw. The random draw is carried out 2,000 times in order to compute a distribution of random average proximities which represent our province-specific statistical counterfactual. From these counterfactual distributions of proximities, we identify the 95th percentile values. In this way, for each new product entering the Italian provinces' export basket between t0 and t1, we can test its compliance to the product space (statistically non-random relatedness).

4. Results

4.1. Testing non-random concentration in the Product Space

Our non-parametric analysis aims to test whether the short-term evolution of export specialization for Italian provinces follows the Product Space's prediction of relatedness of new export goods to those already exported with revealed comparative advantage. As described in the methodology section, our null hypothesis is that new products are randomly located in the product space; a rejection of the null hypothesis implies that the distribution of actual data is – for high levels of relatedness – above our counterfactual threshold for a 95% confidence interval.

Table 1: descriptive statistics on new entrants' proximity to preexisting export baskets.

	Nr. of jumps	Mean	Stdv	Min	Max
2002-2006					
Maximum proximity	13,024	0.5085	0.0906	0.125	1
Average proximity	13,024	0.1876	0.0402	0.0607	0.396
Weighted average proximity	13,024	0.1858	0.0569	0.0042	0.4841
2007-2011					
Maximum proximity	14,340	0.5119	0.0942	0.125	1
Average proximity	14,340	0.1936	0.0416	0.0503	0.3714
Weighted average proximity	14,340	0.1945	0.0592	0.0058	0.4576

Note: new entries include all products having an RCA lower than 0.5 at t_0 and higher than unity at t_1.

In *Table 1* we report the main descriptive statistics regarding the new entries in the provincial export baskets for the two sub-periods considered in our analysis, respectively 2002-2006 (*pre-crisis* period) and 2007-2011 (*crisis* period). The number of new products that enter the Italian provinces' export baskets in the two periods is slightly increased, 13,024 and 14,340 respectively.[24] As expected, the mean values of proximities differ according to the employed definition. When using the maximum

---

[24] The increase is not due to the slightly higher number of provinces in the crisis period. In fact, the new Italian provinces (Carbonia-Iglesias, Medio-Campidano, Ogliastra and Olbia-Tempio) are all located in Sardinia and present a low number of new products in the export basket.

proximity, the 'distance' between the new products and the initial specialization is lower. When using the relative measured of proximity (eqs. 5-6), these distances are higher since the proximity is computed with respect to all goods exported with a RCA larger than unity at time t_0. Interestingly, from *Table 1* it is clear that the mean values of proximities – regardless of the definition of new products or the measure of proximity adopted – increase during the second sub-period. It is important to note that we cannot infer evidence of increasing relatedness from this change in average levels of proximities since the network structure of the provincial baskets is different in the two sub-periods. For this reason, a methodology that explicitly takes into account how provincial export baskets evolve is fundamental for testing the dynamics of specialization.

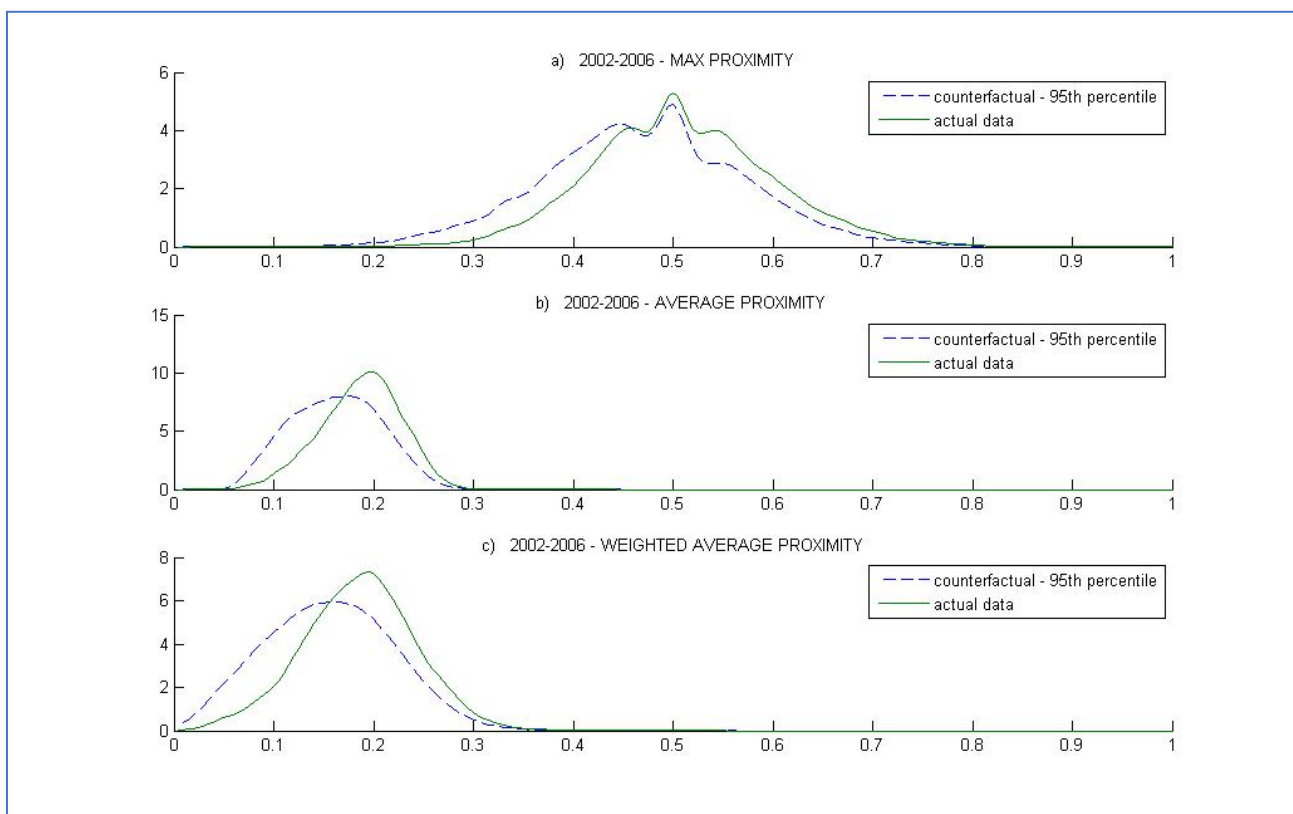*Figure 3: Kernel Density estimates for actual data and counterfactuals in the pre-crisis period.*

Figure 3 represents the Kernel smoothed density estimates for our three definitions of proximity in the sub-period 2002-2006.[25] The horizontal axis measures the proximity between new products at t_1 and the bundle of goods exported at time t_0. The higher the value of our measure of proximity is, the closer the new entrant product to the export basket at time t_0. In all these estimates, we can see that

---

[25] For both sub-periods, we report the figures representing the Kernel density distributions for the alternative identification strategies of new entries in Appendix A1. The results confirm the findings reported in this paragraph.

for high levels of relatedness, the distribution of proximities in the actual data is above that of the counterfactual.

When we employ the absolute measure of proximity, eq. (4), the hypothesis of random relatedness is rejected for proximities ranging from 0.46 to 0.84; in this range, where 70.8% of new entries fall, the actual data's distribution lies above our counterfactual 95th percentile's distribution (see *Figure 3, panel a*).

Using relative measures of relatedness, eqs. (5) and (6), produce highly similar results. In Panel (b) of *Figure 3,* we report the Kernel densities when we use the average proximity of new products with all the products exported at time t_0. The Kernel distribution of actual data is above the counterfactual 95th percentile distribution (our randomness threshold) for values of proximity between 0.17 and 0.4. The 70.35% of new products that enter the Italian export basket in 2006 fall in this range. Similar results are obtained when using the weighted average relatedness specified in eq. (6). In this case, the range of non-random proximities is between 0.16 and 0.49 which represents 71.34% of new entries between 2002 and 2006.[26]

Our results thus confirm the non-random overall evolution of the Italian product space in the pre-crisis period and are robust to alternative specifications of the definition of both 'new products' and 'proximities'. More than 70% of the new products are related to the pre-existing Italian production space.

Is this result of a relatively strong path-dependence confirmed for new products that enter the export basket during the crisis period? The Kernel density estimations for the period 2007-2011 are reported in the three panels of *Figure 4*. For both absolute and relative measures of proximities, the strong path-dependence in the evolution of the export basket is confirmed. We note a slight shift of the threshold values to the right, i.e. the null hypothesis of randomness is rejected for slightly higher values of proximity. In line with this result, we find that on average a smaller percentage of new entries is significantly related to the pre-existing export basket (respectively 69.65%, 68.9% and 72.5% for the three definition of proximities employed).

---

[26] Note that for proximity values that are in the upper tail of the distribution (above the threshold values), we cannot statistically reject the null hypothesis since few observations both in actual and simulated data fall in this area. It is important to underline that the percentage of actual proximities falling in the upper tails of the three distributions in Figure 1 are higher than the simulated ones; we interpret this as an indication of non-random relatedness (although statistically not significant).

In other words, although the crisis has not halted the development of new products in quantitative terms, the number of 'new apples that fall closer to the tree' is slightly lower compared to the pre-crisis period. This result seems to go in the same direction as the Schumpeterian process of creative destruction during crisis confirming the role of such shocks as engines of structural change (Schumpeter, 1942), at least in the very short term.

Figure 4: Kernel Density estimates for actual data and counterfactuals in the crisis period.
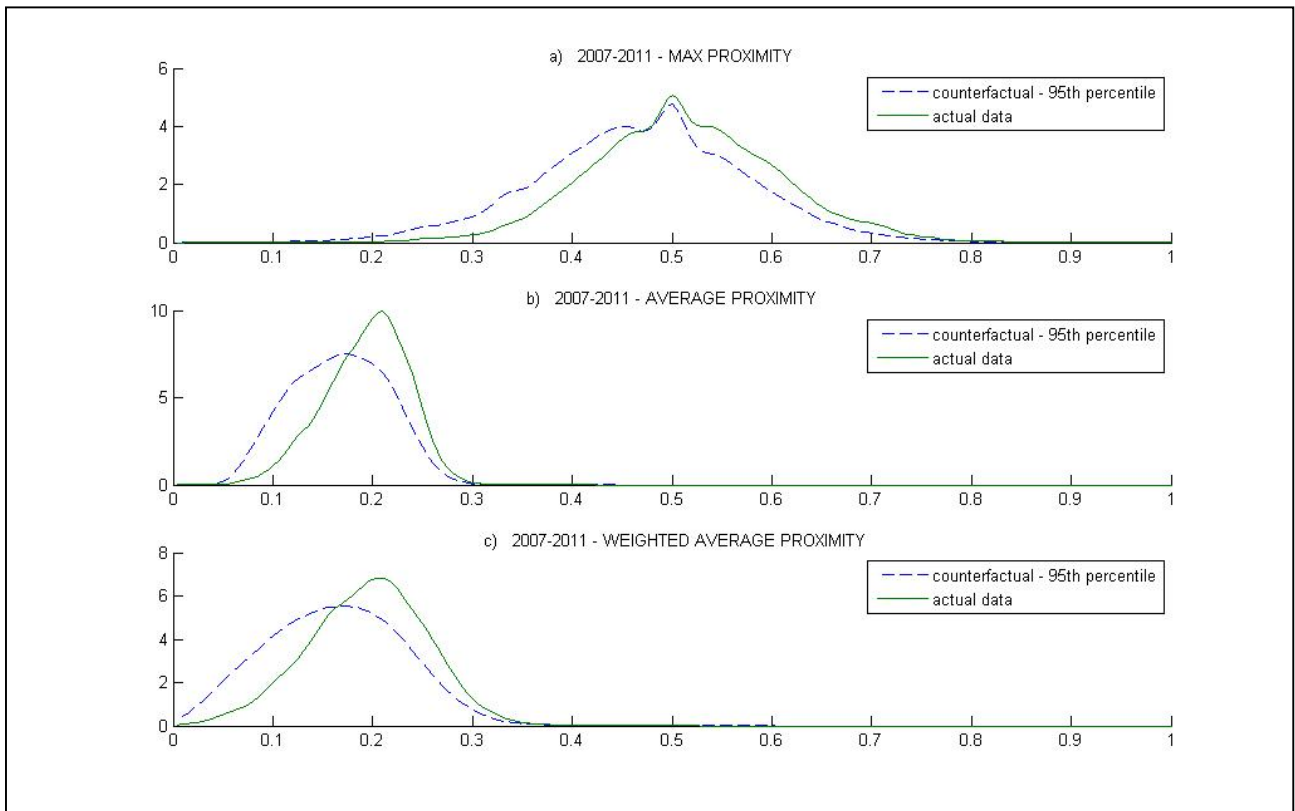
The analysis of the Kernel smoothed densities shows us that, regardless of the definition used to compute our measure of proximity and the definition of new products (see Appendix A1), our general test of non-randomness highlights a clear pattern: new products entering the export baskets of Italian provinces, both in the pre-crisis and the crisis periods, are not randomly distributed over the network of relatedness between goods developed by Hidalgo et al (2007). Indeed, a large share of new entries is related in a non-random manner to the initial export basket.

4.2. Test on single products

In the previous paragraph, we have shown that the evolution of Italian provincial export baskets follows a non-random pattern that confirms the existence of strong path dependence for most of the (new) products in which provinces develop a revealed comparative advantage.

Our global test of non-random relatedness also reveals that a significant share of new products - over 30% - defeats the 'static' comparative advantage contrary to the product space framework's dictates.[27] In our opinion, these 'apples that fall far from trees' are probably the most interesting ones for informing the current debate on structural change and industrial policy. It is thus interesting to analyse whether there are any province- or product-specific characteristics that are systematically associated with these 'long distance' jumps over the product space of Italian provinces.

In *Table 2*, we report the main results from our test of non-random relatedness performed on each single product that entered the Italian export basket in the two periods considered (see previous section). In particular, we show the percentage of new goods by macro-sector and macro geographical areas for which we confirm the hypothesis of non-random relatedness (*i.e.* path-dependence).

Table 2: test on single products. Relatedness measured as 'maximum proximity'

Precrisis	North-West		North-East		Centre		South and Isles		Italy	
	nr. of new entries	max prox	nr. of new entries	max prox	nr. of new entries	max prox	nr. of new entries	max prox	nr. of new entries	max prox
Animal & Animal Products	97	53%	93	45%	41	49%	134	42%	365	46%
Vegetable Products	88	35%	104	33%	84	32%	202	50%	478	40%
Foodstuffs	100	33%	103	52%	80	59%	207	61%	490	53%
Mineral Products	56	16%	47	21%	32	22%	48	13%	183	17%
Chemicals & Allied Industries	340	53%	239	50%	154	49%	260	50%	993	51%
Plastics / Rubbers	136	71%	132	64%	90	63%	144	68%	502	67%
Raw Hides, Skins, Leather, & Furs	60	58%	85	39%	85	47%	92	50%	322	48%
Wood & Wood Products	145	39%	143	57%	99	62%	99	73%	486	56%
Textiles	980	69%	871	73%	713	74%	871	75%	3,435	73%
Footwear / Headgear	31	52%	43	60%	50	52%	46	52%	170	54%
Stone / Glass	84	58%	105	57%	93	62%	143	59%	425	59%
Metals	432	57%	398	60%	220	53%	341	58%	1,391	57%
Machinery / Electrical	759	65%	699	65%	551	63%	738	62%	2,747	64%
Transportation	63	43%	69	43%	51	43%	94	48%	277	45%
Miscellaneous	204	54%	212	51%	171	50%	173	61%	760	54%
Crisis										
Animal & Animal Products	94	43%	102	55%	43	40%	179	42%	418	45%
Vegetable Products	109	31%	150	34%	95	35%	287	46%	641	39%
Foodstuffs	128	50%	104	43%	100	48%	276	60%	608	53%
Mineral Products	64	25%	45	22%	48	25%	62	35%	219	27%
Chemicals & Allied Industries	398	54%	235	46%	164	57%	245	59%	1,042	54%
Plastics / Rubbers	182	68%	104	70%	110	67%	174	69%	570	69%
Raw Hides, Skins, Leather, & Furs	62	42%	102	43%	90	58%	84	46%	338	48%
Wood & Wood Products	182	52%	196	61%	130	62%	171	59%	679	58%
Textiles	808	67%	789	74%	683	74%	677	74%	2,957	72%
Footwear / Headgear	42	43%	53	77%	41	63%	48	60%	184	62%
Stone / Glass	171	63%	139	68%	123	63%	195	69%	628	66%
Metals	544	60%	443	63%	273	66%	432	64%	1,692	63%
Machinery / Electrical	872	70%	813	71%	676	69%	756	63%	3,117	69%
Transportation	97	57%	75	44%	61	52%	134	57%	367	53%
Miscellaneous	269	58%	205	68%	195	59%	211	54%	880	60%

---

[27] In other words, a significant number of new products enter the export baskets at t_1 in areas of the product space that are ideally 'far away' from those where the export basket at t_0 lies.

The results in *Table 2* highlight a heterogeneous pattern through HS sections and Italian macro-regions. The percentages reported represent the number of products for which the null hypothesis of random relatedness is rejected over the total number of new entries for each region and each product section. The higher the share, the more a provincial export basket follows a path-dependent pattern. Textile products are, on average, those with the highest number of new entries and percentages of randomness rejection, irrespective of the region that is taken into account (overall ratio of 73 and 72% in the pre-crisis and crisis periods, respectively). High frequencies of new entries are recorded also for machinery and electrical products with a total number of new entries increasing in the second sub-period from 2,747 to 3,117. In this sector, the degree of path-dependence is also quite high and the percentage of relatedness ranges from 62% in the pre-crisis period for southern provinces to 71% in the crisis period for north-eastern provinces. A low degree of path-dependence is found, as expected, given the weak role of local capabilities in resource extractive industries, for mineral products (17 and 27% in pre-crisis and crisis periods, respectively). Heterogeneity within sectors is evident in food industries during the 2002-2006 period; results show path dependence for 33% of products entering the export basket of north-western provinces with 61% for southern regions. A similar outcome is shown – for the crisis period – in the Footwear section for which we find 43% of related entries for the provinces in the north-west and 77% for north-eastern regions.

The high degree of heterogeneity highlighted in *Table* 2 suggests that the characteristics of local economies – as well as product specificities – play an important role in influencing the pattern of structural change.

4.3. Which provinces 'defeat' the static comparative advantage? A probit analysis

What economies produce matters for growth as emphasized by previous studies.[28] The production basket, but also its evolution, crucially depend on local capabilities. In this section, we focus on a particularly important research question: which factors drive more radical (unrelated) changes in the composition of production baskets? *Figure 5* shows the positive relationship between provincial per capita growth rate and the share of random entries in the analysed period. Provinces whose production baskets have diversified into unrelated areas of the PS experienced on average better economic performance, in particular before the (exogenous) shock of the global crisis.

---

[28] We refer the reader to studies cited in Section 2. See also Coniglio *et al.* (2016) for a detailed analysis of provincial growth and sophistication in Italy.

Figure 5: Unrelated entries and economic growth
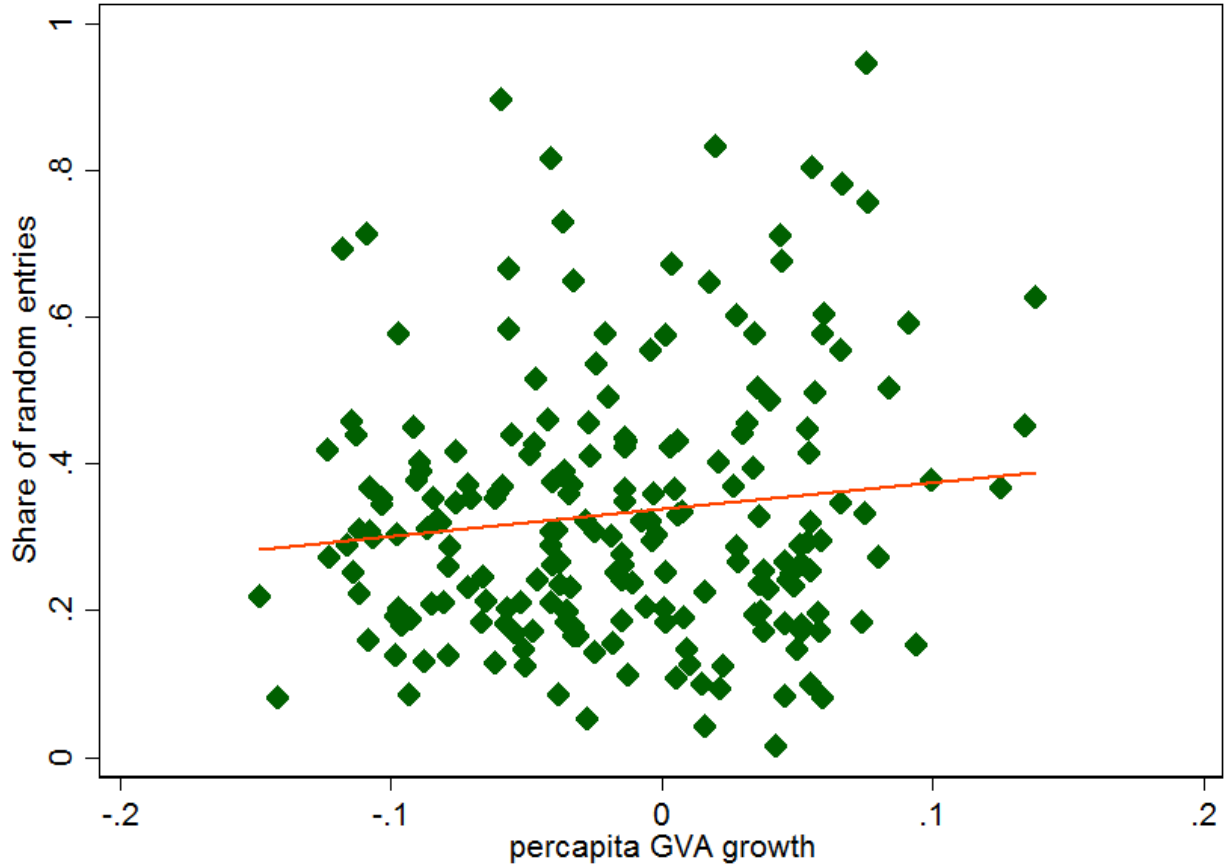
Note: The scatter plot representing the nexus between the percapita GVA growth and the share of random entries measured as the export value of random entries over the export value of the products exported with revealed comparative advantage.

We implement a parametric analysis, by means of a probit model, with the aim of investigating which factors are associated with a higher ability of provinces to diversify away from the initial comparative advantage in the pre-crisis and crisis periods.[29]

Our dependent variable, *random_entry*, is a dummy which is equal to 1 if the new product *i* entering the production space of province *k* at time t1 is statistically unrelated to pre-existing economic as defined in the previous section, and 0 otherwise. We define new entries using our absolute measure of distance as defined in eq. (4), *i.e.* maximum proximity. Our estimated model is the following:

$$\text{Prob(random\_entry}_{ik}) = \alpha_{ik} + \beta_X X_k + \beta_Z Z_i + \beta_{IND} IND_i + \varepsilon_{ik} \qquad (8)$$

---

[29] The test will be on data with distances measured as the maximum among the proximities between new entrant goods and those present in the export basket at time t_0. Among the three methods, this is the one that has an interpretation that is closer to geographical distance.

where X_k includes our main province-level covariates, Z_i includes product-level controls and IND_i is a set of (macro)sector fixed effects.[30]

We employ the measure developed by Hausmann et al (2007), ExpY, as a proxy for the level of sophisticatedness of the export basket [31].We expect that the more complex the degree of sophisticatedness of the production basket is, the higher the probability there will be more radical changes. In fact, since more complex goods require a broader set of capabilities, it would be relatively easier in these economies to redeploy these capabilities in order to develop new unrelated products.

Following Boschma and Iammarino (2009), we include a measure of export diversification *variety* in our specification as defined in Frenken (2007).[32] We expect more differentiated economies producing a large number of varieties to be endowed with a relatively larger and broader set of production capabilities which allows the provincial economy to diversify away from the initial production basket. This positive association is expected to be stronger the higher the diversity in terms of unrelated varieties that presumably use different local capabilities (as in Castaldi et al (2015) in the context of the emergence of technological breakthroughs). Conversely, economies that are diversified within a specific sector may experience a 'lock-in' effect which hampers diversification into other areas of the PS. To this purpose, we include measures for *related* and *unrelated variety* in some specifications.

Moreover, we include the variables *trade openness* to control for the provincial exposure to international trade and international knowledge transfers which may affect the ability of the provincial economy to diversify its production bundle 'away' from the pre-existing one. Finally, as a proxy of human capital, we also include a variable measuring the number of students enrolled in tertiary education over the total active population.

Our specifications include two product-level controls. First, we include ProdY as a proxy of the sophisticatedness of the new products that enter the provincial production basket at time t1. Second, a measure of the 'centrality' of the new products in the network of relatedness – equals to its average proximity - is included. Both measures, together with the macro-industry fixed effects, capture product-specific characteristics which affect the dependent variable, but are unrelated to provincial features (our main variables of interest).

---

[30] We include industrial sectors' fixed effects using the 21 Sections of HS nomenclature in order to consider heterogeneity across macro industries. All province-specific variables refer to the year t_0 whereas all product-specific variables refer to year t_1.

[31] Index of export basket sophistication by Hausmann *et al* (2007). It is computed as the weighted sum of ProdYs of the products exported by a province with weights represented by the export shares. For ProdY's definition, see note 25. For details, see Appendix A2.

[32] In order to simplify the interpretation of results, we include all the 'diversification' indexes in logarithms. Such measures are computed as reported in Appendix A2.

Descriptive statistics of the dependent variable and covariates for both the pre-crisis and crisis periods are reported in *Table 3*.

Table 1: Descriptive Statistics

Precrisis: 2002-2006					
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
PS jump	13024	0.3997236	0.4898603	0	1
Ln ProdY t1	13024	9.505272	0.7283572	3.85543	11.256
Centrality t1	13024	0.1528922	0.0306515	0.0269282	0.231003
Ln ExpY t0	13024	9.866243	0.1597781	9.294399	10.22813
Ln Variety t0	13024	1.846666	0.2439395	0.6236902	2.238506
Ln Rel.Variety t0	13024	1.876836	0.7684997	-2.957147	3.2948
Ln Unrel.Variety t0	13024	5.333217	0.8126831	1.619947	6.468346
Op.toTrade t0	13024	0.3756517	0.2102692	0.0178385	1.284592
UER	12848	0.038368	0.037405	2.19E-05	0.1934
Ln GVA t0	13024	22.97884	0.8267497	21.14331	25.55584
Ln Population t0	13024	13.0173	0.756211	11.40497	15.1265
Crisis: 2007-2011					
Variable	Obs	Mean	Std. Dev.	Min	Max
PS_jump	14340	0.3760112	0.4843998	0	1
Ln ProdY t1	14340	9.539061	0.7151328	5.580651	11.24809
Centrality t1	14340	0.1536002	0.031584	0.0077109	0.229469
Ln ExpY t0	14340	10.05196	0.1404983	9.602266	10.51017
Ln Variety t0	14340	1.828332	0.2639739	-0.015982	2.267477
Ln Rel.Variety t0	14340	1.734126	0.90491	-4.60517	3.160912
Ln Unrel.Variety t0	14340	5.287507	0.9041724	0.3585037	6.450047
Op.toTrade t0	14340	0.4682068	0.2732955	0.0012226	2.652769
Ln GVA t0	14340	23.00455	0.8240444	20.54209	25.59303
UER	14340	0.036236	0.033688	0	0.180167
Ln Population t0	14340	12.99877	0.7556906	10.9626	15.1629

In *Table 4* the marginal effects of the probit regressions are reported separately for the two sub-periods, pre-crisis and crisis. In Models 1 and 5, we employ a parsimonious model of the probability of unrelated new entries where we include our measure of overall provincial export basket sophisticatedness, *ExpY* (in log)*,* and a measure of production diversification,*Variety*(in log), controlling for industry and product fixed effects. As expected, provinces characterized by a higher degree of sophisticatedness are found to be positively associated with the probability of experiencing more radical structural changes in the composition of the export basket. We interpret this result as further evidence of the important role of the 'complexity' of what economies produce and export;

complexity has a direct effect on growth performance (Haussmann et al 2007), but also an effect on the ability of an economy to diversify away from the initial comparative advantage.

It is interesting to note that the magnitude of the effect is rather stable in the considered time span and slightly smaller during the crisis period. Furthermore, we find a weak positive effect of the degree of export diversification (Ln*Variety*) on the probability of unrelated new entries only in the pre-crisis period. In models 2 and 6, we test for heterogeneous effects of related versus unrelated varieties. We find that only the number of unrelated varieties is positively associated with the probability of unrelated new entries. On the contrary, a higher provincial diversification within the same macro-sector (i.e. product diversification within a 4-digit sector) is negatively associated with unrelated entries in the provincial export basket. This result is strong evidence of the lock-in effect that 'dense' sectoral specialisation may represent. For instance, the presence of consolidated industrial districts may hamper the diversification of the provincial economy over the PS. In the ascendant phase of development of a district, this path dependence may reinforce growth, but in the maturity or decline phases, this pattern may represent a less desirable feature of the local economy.

In the specifications reported in columns 3-4 and 7-8, we introduce two additional covariates, *tertiary education* (a proxy for provincial human capital) and provincial *international trade openness*. It is interesting to observe that openness to trade and (although less strongly) a larger endowment of human capital are positively associated with the probability of unrelated new entries only in the period of crisis. Intuitively, a larger market access amplifies the value of producing and the ability to adapt to new products. The effect may also be driven by the presence of high-productivity firms in more open economies which are in turn more able to react to market difficulties by diversifying production.[33] 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.

---

[33] Recent contributions in the international trade literature have emphasized the structural differences between firms exporting and investing abroad and purely domestic ones in terms of productivity, wages, size, mark-ups and other crucial firm-level characteristics (Melitz 2003; Melitz and Ottaviano, 2008; Bernard *et al* 2007).

Relatedness measured as the maximum proximity between new entry and export basket at time t_0.

VARIABLES	2002-2006				2007-2011			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	margin	margin	margin	margin	margin	margin	margin	margin
Ln expY	0.117***	0.0845***	0.108***	0.0810***	0.108***	0.0889***	0.103***	0.0872***
	(0.0246)	(0.0282)	(0.0253)	(0.0286)	(0.0251)	(0.0263)	(0.0257)	(0.0265)
LnVariety	0.0275*		0.0239		0.00574		0.0110	
	(0.0161)		(0.0179)		(0.0133)		(0.0164)	
LnRelVariety		-0.0146*		-0.0131		-0.0224***		-0.0179**
		(0.00844)		(0.00862)		(0.00704)		(0.00754)
LnUnrelVariety		0.0243***		0.0240***		0.0208***		0.0149*
		(0.00847)		(0.00908)		(0.00727)		(0.00786)
Openness			0.00824	-0.00897			0.0389**	0.0436**
			(0.0215)	(0.0249)			(0.0152)	(0.0170)
UER			0.181*	0.156			0.154	0.187*
			(0.106)	(0.110)			(0.109)	(0.112)
Observations	13,024	13,024	12,848	12,848	14,340	14,340	14,026	14,026
Industry FE	yes	yes	yes	yes	yes	yes	yes	yes
Product FE	yes	yes	yes	yes	yes	yes	yes	yes
Pseudo R-squared	0.177	0.177	0.176	0.176	0.204	0.205	0.204	0.205

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5. Robustness check. Dependent variable: probability of experiencing a random new entry. $randomjump = 1$ if the new entry is simultaneously random according to our three measures of relatedness.

VARIABLES	2002-2006				2007-2011			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	margin	margin	margin	margin	margin	margin	margin	margin
Ln expY	0.0404**	0.0287	0.0417**	0.0312	0.0613***	0.0521***	0.0620***	0.0544***
	(0.0165)	(0.0188)	(0.0168)	(0.0190)	(0.0167)	(0.0174)	(0.0171)	(0.0176)
LnVariety	0.0311***		0.0372***		0.00693		0.0158	
	(0.0108)		(0.0120)		(0.00872)		(0.0109)	
LnRelVariety		-0.00595		-0.00404		-0.00952**		-0.00638
		(0.00566)		(0.00579)		(0.00464)		(0.00500)
LnUnrelVariety		0.0130**		0.0155**		0.0105**		0.00826
		(0.00571)		(0.00607)		(0.00481)		(0.00519)
Openness			-0.0146	-0.0232			0.0110	0.0124
			(0.0141)	(0.0163)			(0.0102)	(0.0114)
UER			-0.0192	-0.0316			0.0315	0.0411
			(0.0706)	(0.0734)			(0.0735)	(0.0757)
Observations	13,024	13,024	12,848	12,848	14,340	14,340	14,026	14,026
Industry FE	yes	yes	yes	yes	yes	yes	yes	yes
Product FE	yes	yes	yes	yes	yes	yes	yes	yes
Pseudo R-squared	0.415	0.415	0.415	0.415	0.449	0.449	0.449	0.449

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As a robustness check, we replicate the estimates using alternative and more restrictive definitions of unrelated new entries in the provincial export basket. We consider a new entry as unrelated (*i.e.* our dependent variable equals 1) if the new good enters the provincial export basket in a statistically unrelated way according to all three measures of relatedness (maximum, average and weighted average; eqs. 4 to 6). The results reported in *Table 5* confirm the evidence described above. All variables show lower values of the marginal effects, as expected, with a narrow definition of the dependent variable, but the main covariates, *EXPY* and *Unrelated Variety,* retain a positive and statistically significant effect on the probability of more radical changes in both time spans considered.

5. Concluding remarks

In this work we have analysed the evolution of the export basket of Italian provinces between 2002 and 2011 in order to test its conformity with the prediction of path-dependence which is a cornerstone of the Product Space framework. According to the approach of the network of relatedness between goods developed by the seminal contributions of Hausmann and Klinger (2007) and Hidalgo *et al* (2007), the goods that have higher probabilities of entering the export portfolio are those sharing common local capabilities with those previously produced. Hence, local capabilities determine the direction of structural change and, at the same time, constrain the evolution of the comparative advantage of nations and regions to those products that are strongly related to the ones already produced.

These predictions have important implications for industrial and innovation policies since they suggest the implementation of selective policies targeted to sectors *related* to the current comparative advantage.

Our results show that both in pre- and crisis periods, the goods that Italian provinces started exporting with revealed comparative advantage tend to be highly related to the set of goods exported five years before, thus confirming a general pattern of path-dependence.

This paper contributes to the existing literature by providing a methodological approach for testing the non-randomness of the evolution of structural change along the product space. 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 economies over time. We focus on Italian provinces and on relatively short-run changes in the export baskets (5 year intervals), but our methodology can be easily used to analyse the pattern of structural change in other countries or regions and over different time horizons.

The focus on sub-national areas is particularly interesting since capabilities – *i.e.* technologies, capital, skills and institutions – have a strong local dimension and are unequally distributed over space, in particular in countries such as Italy with highly differentiated and heterogeneous economic areas.

Although we confirm the general tendency of path-dependence, we find that on average approximately 30% of new goods that enter the export basket of Italian provinces are largely unrelated with the pre-existing comparative advantage. These apples that fall far from the tree are the most interesting from a policy perspective since they represent cases of more radical structural change.

These unrelated entries are also found to be associated with a higher growth performance, in particular prior to the crisis period.

The significant deviations from the pattern of path-dependence observed in Italian provinces and its high degree of geographical heterogeneity suggests that caution should be exercised in using the Product Space as a map for identifying the 'latent comparative advantage' of countries and regions. Structural change may take a different, often unpredictable, path.

Interestingly, we find that the provinces that are more likely to 'defeat' the initial static comparative advantage are those characterized by a relatively higher production sophisticatedness, a higher initial level of product diversification in unrelated sectors, with relatively more open economies, and (although the evidence is less robust) better endowment of human capital. Those provinces with higher average production sophisticatedness and more complex and diversified sets of local capabilities are less constrained by path-dependence and have a higher probability of experiencing long jumps over the product space.

In addition, our finding of a relatively lower degree of path-dependence in the evolution of structural change during the crisis seems to go in the same direction as the Schumpeterian process of creative destruction during large and pervasive shocks, at least in the short term (Schumpeter, 1942).

It would be interesting to analyse the evolution of the production bundle once the Italian (and global) economy regains momentum in order to test whether our results are confirmed. Another important question is the role played by spatial spillovers in the process of provincial structural change. These interesting analyses are left to future research.

References

Altenburg, T (2011), Industrial Policy in Developing Countries, *DIE WP 4/2011*

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

Bernard, A.B., Jensen, J.B., Redding, S.J. and Schott, P.K. (2007), Firms in International Trade, *Journal of Economic Perspectives, 21(3): 105–30.*

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.*

Boschma, R. Minondo, A. and M. Navarro (2013), The Emergence of New Industries at the Regional Level in Spain: A Proximity Approach Based on Product Relatedness, *Economic Geography, Clark University, vol. 89(1), pages 29-51, 01.*

Boschma, R. Balland P. and D.F. Kogler (2015), Relatedness and technological change in cities: the rise and fall of technological knowledge in US metropolitan areas from 1981 to 2010, *Industrial and Corporate Change*, 24(1): 223-250.

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.

Cirera, X., Marin, A. and Markwald, R. (2012), *Firm Behaviour and the Introduction of New exports: Evidence from Brazil. IDS Working Papers, 2012: 1–105.*

Coniglio, N.D., Lagravinese, R. and Vurchio, D. (2016), Production 'sophisticatedness' and growth: evidence from Italian Provinces before and during the crisis, 1997-2013, *Cambridge Journal of Regions, Economy and Society, vol.9(2), Oxford University Press.*

Donoso, V. and Martin, V. (2016), Product relatedness and economic diversification in the USA: an analysis at the state level, *The Annals of Regional Science, Volume 56(2), pages 449–471.*

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

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

Felipe, J., Kumar, U. and Abdon, A. (2013), Exports, Capabilities, and Industrial Policy in India, *Journal of Comparative Economics, Volume 41, Issue 3, August 2013, Pages 939-956, ISSN 0147-5967.*

Felipe, J., Kumar, U., Usui, N. and Abdon, A. (2013), Why has China succeeded? And why it will continue to do so, *Cambridge Journal of Economics, Oxford University Press, vol. 37(4), pages 791-818.*

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

Hausmann, R. Hwang, J. and Rodrik, D. (2007), "What you export matters," Journal of Economic Growth, Springer, vol. 12(1), pages 1-25, March

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

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.*

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.*

Kuznets, S. (1966), *Modern Economic Growth*. New Haven, CT: Yale University Press.

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.

Lo Turco, A. and Maggioni, D. (2016), On firms' product space evolution: the role of firm and local product relatedness, *Journal of Economic Geography 2016 16: 975-1006.*

McMillan, M. S. and Rodrik, D. (2011), Globalization, Structural Change and Productivity Growth, *NBER Working Papers 17143, National Bureau of Economic Research, Inc.*

Melitz, M.J. (2003), The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity, *Econometrica, 71: 1695-725.*

Melitz, M.J. and Ottaviano G. (2008), Market Size, Trade, and Productivity, *Review of Economic Studies, 75: 295-316.*

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.*

Neffke, F., Henning, M. and Boschma, M.R. (2011), How do regions diversify over time? Industry relatedness and the development of new growth paths in regions, *Economic geography volume 87, issue 3, pages 237-265, July 2011*

Poncet, S. and Starosta de Waldemar, F. (2013), Export Upgrading and Growth: The Prerequisite of Domestic Embeddedness, *World Development, Elsevier, vol. 51(C), pages 104-118.*

Schumpeter, J. A. (1942), Capitalism, Socialism, and Democracy, *New York: Harper and Brothers.*

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

Spence, M. (2011), The Next Convergence: The Future of Economic Growth in a Multispeed World. *New York: Farrar, Straus and Giroux.*

Stiglitz, J., Lin, J., Monga, C. and Patel, E. (2013), Industrial policy in the African context, *Policy Research Working Paper Series 6633, The World Bank.*
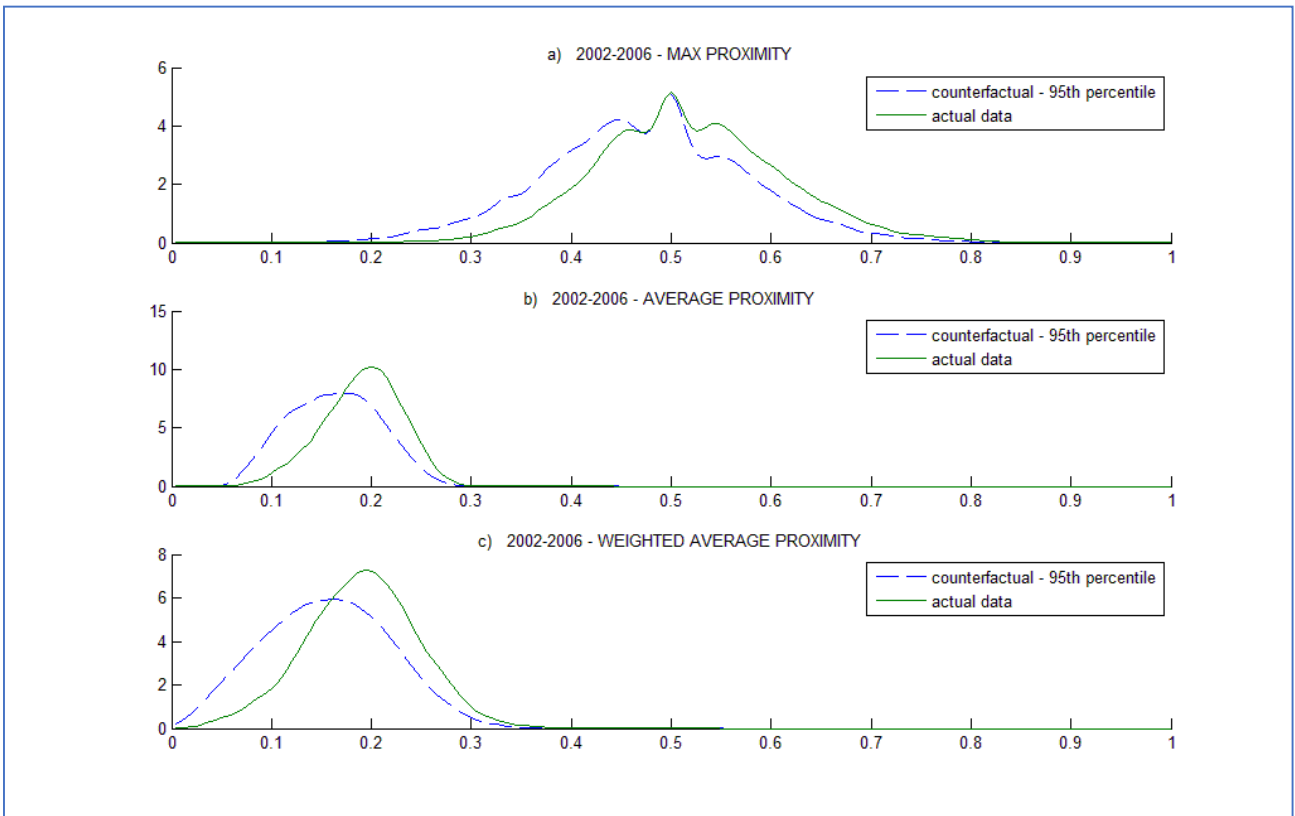
Appendix A1

Figure A1.1: Kernel densities for actual data and counterfactuals in the pre-crisis period. $RCA_{ik,t_0} < 1 \; and \; RCA_{ik,t_1} \geq 1$
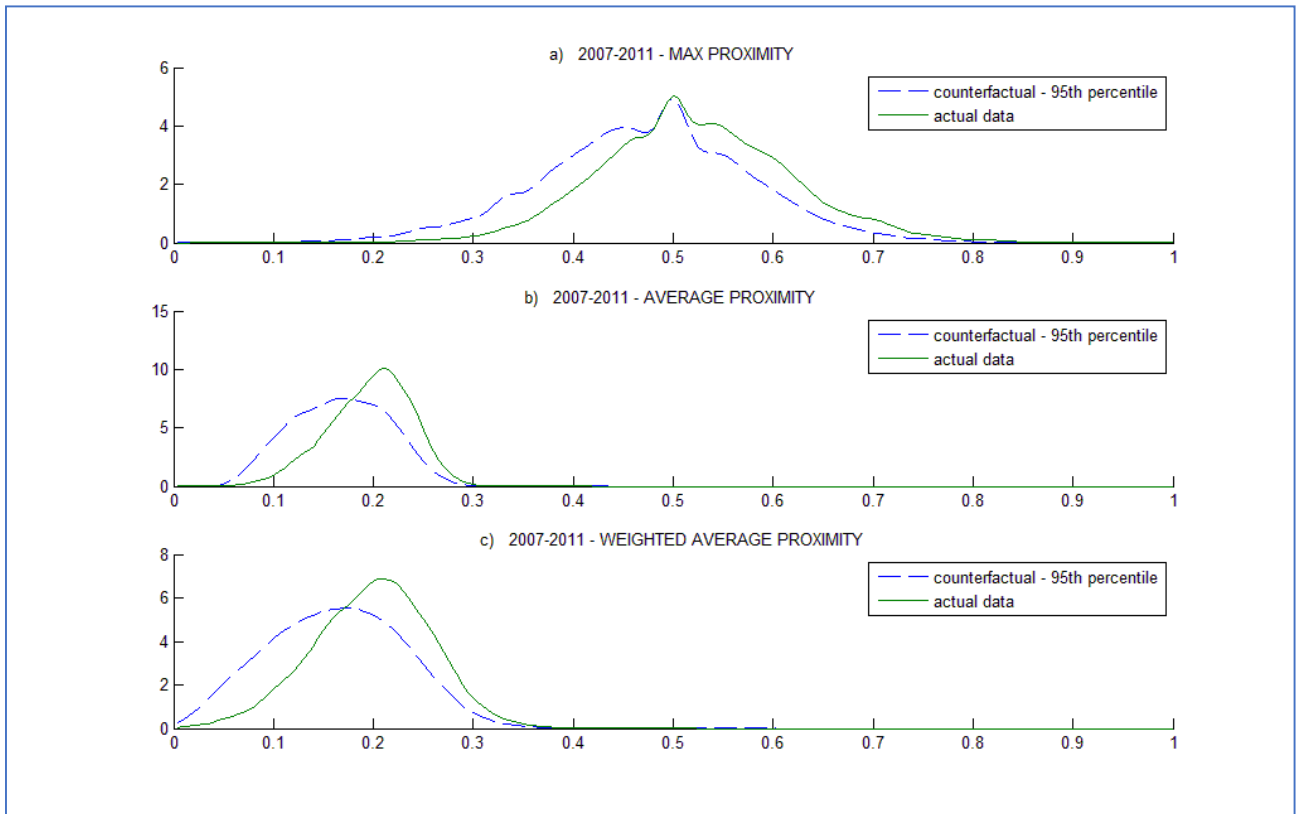
Figure A1.2: Kernel densities for actual data and counterfactuals in the crisis period. $RCA_{ik,t_0} < 1$ and $RCA_{ik,t_1} \geq 1$

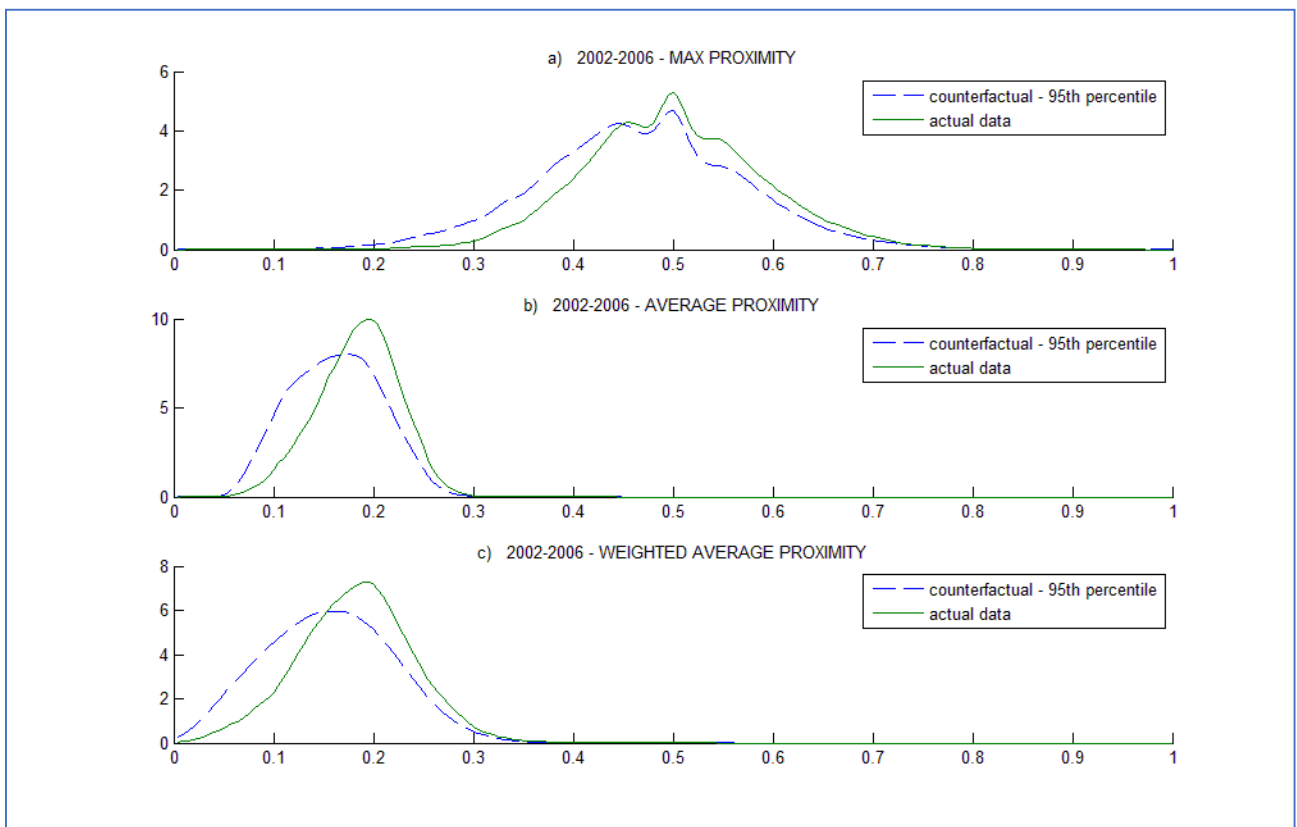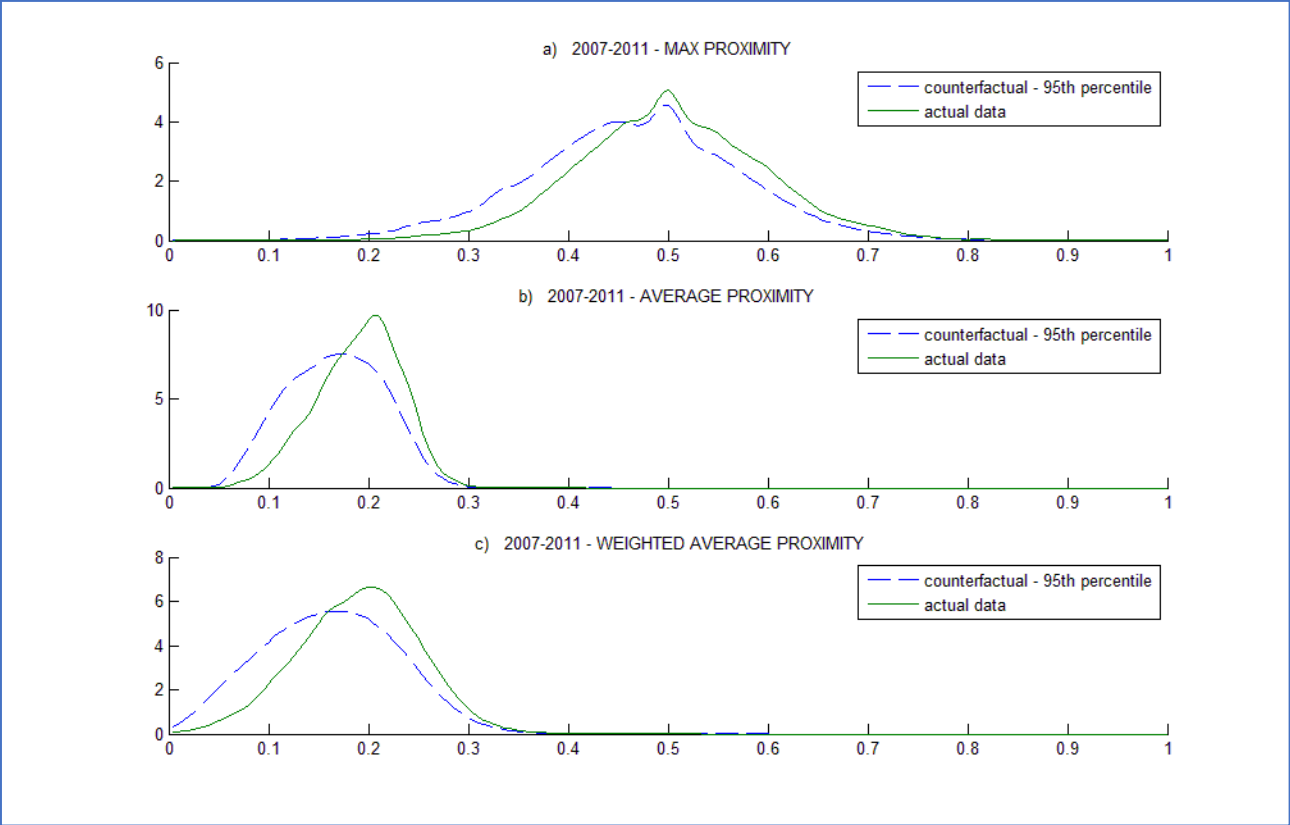**Figure A1.3: Kernel densities for actual data and counterfactuals in the pre-crisis period.** $RCA_{ik,t_0} <$ 0.2 and $RCA_{ik,t_1} \geq 1$

Figure A1.4: Kernel densities for actual data and counterfactuals in the crisis period. $RCA_{ik,t_0} < 0.2 \; and \; RCA_{ik,t_1} \geq 1$

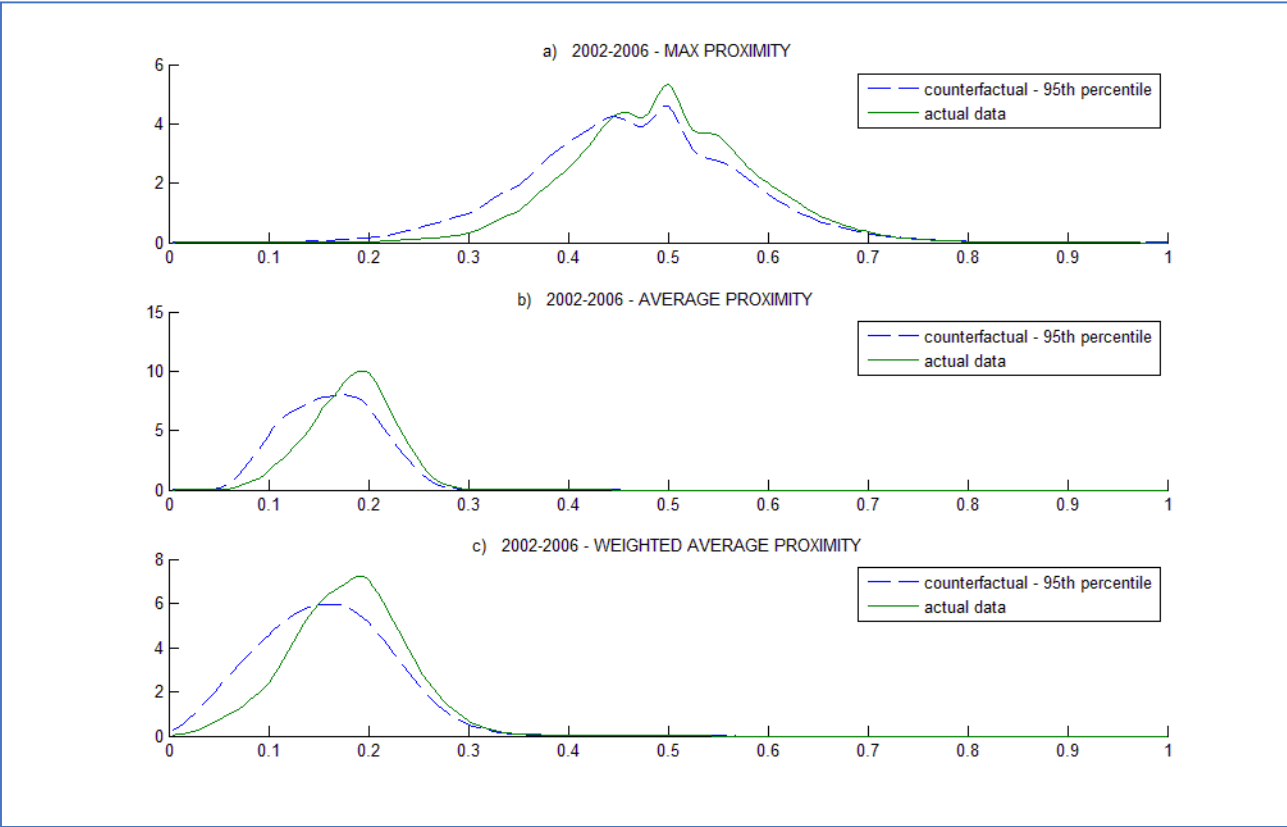**Figure A1.5: Kernel densities for actual data and counterfactuals in the pre-crisis period.** $RCA_{ik,t_0} < 0.1 \ and \ RCA_{ik,t_1} \geq 1$
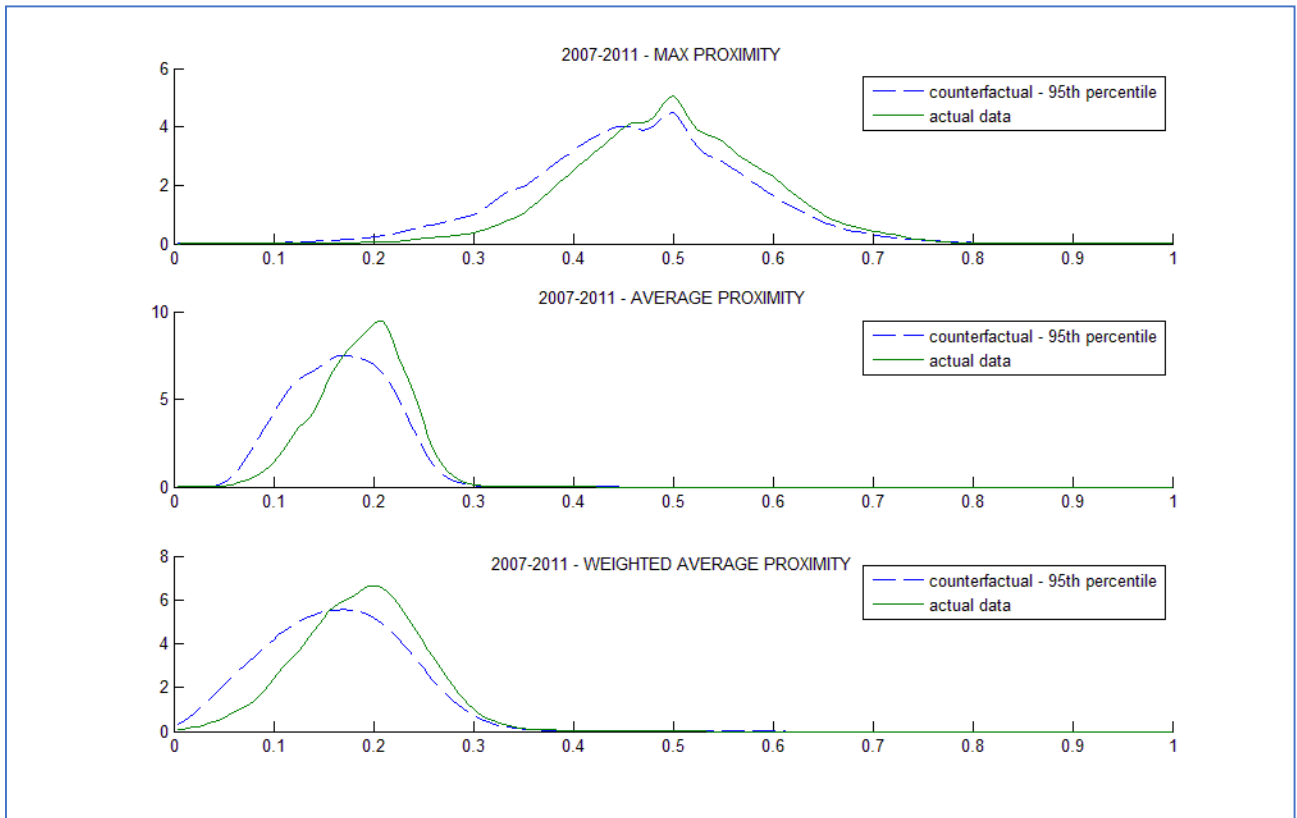
Figure A1.6: Kernel densities for actual data and counterfactuals in the crisis period. $RCA_{ik,t_0} < 0.1 \ and \ RCA_{ik,t_1} \geq 1$

Appendix A2

$ProdY_{i,t_1}$ and $ExpY_{k,t_0}$

Following Hausmann *et al* (2007), we compute two indexes – ProdY and ExpY – proxying average productivity associated to 6-digit products and to provincial export baskets respectively. Using UNComtrade and WorldBank data, we obtain:

$$PRODY_{i,t_1} = \sum_c \frac{x_{cit_1}/X_{ci_1}}{\sum_c x_{cit_1}/X_{ci_1}} Y_{ct_1} \qquad (A2.1)$$

where x_{ict_1} is the export of good i at time t_1 by country c, X_{ic_1} is the total exports for country j at time t_1 and Y_{ct_1} is the per-capita GDP of country c at time t_1. $PRODY_{i,t_1}$ represents the average GDP per capita of the countries which have a comparative advantage in the production of good i and gives us a measure of the sophisticatedness level of each new entrant products at time t_1.

Computing ProdYs also for $t = t_0$, we obtain:

$$EXPY_{k,t_0} = \sum_i \frac{x_{ikt_0}}{X_{kt_0}} PRODY_{i,t_0} \qquad (A2.2)$$

where the subscript k refers to the Italian provinces and x_{ikt_0}/X_{kt_0} is the share of the product i in the total export value of province k at time t_0. $EXPY_{k,t_0}$ represents the average GDP per capita of provincial export baskets and gives us a measure of the sophisticatedness level associated to the provincial production.

Variety, related variety and unrelated variety

Following Boschma and Iammarino (2009), we compute our diversification indexes as follows:

$$1) \quad variety_{k,t_0} = \sum_{i=1}^{M} \frac{x_{ik,t_0}}{X_{k,t_0}} log_2 \left[\frac{1}{\left(\frac{x_{ik,t_0}}{X_{k,t_0}} \right)} \right] \qquad (A2.1)$$

where $x_{ik,t_0}/X_{k,t_0}$ is the export value share of 6-digit product i over province k's export basket at time t_0. *variety* is a Herfindal index measuring the degree of export diversification by means of an entropy measure at the 6-digit level.

$$2) \quad related \; variety_{k,t_0} = \sum_{g=1}^{G} P_{gk,t_0} H_{gk,t_0} \qquad (A2.2)$$

with $P_{gk,t_0} = \sum_{i \in S_g} x_{ik,t_0}/X_{k,t_0}$ and H_{gk,t_0} computed as:

$$H_{gk,t_0} = \sum_{i \in S_g} \frac{\frac{x_{ik,t_0}}{X_{k,t_0}}}{P_{gk,t_0}} \, log_2 \left[\frac{1}{\frac{x_{ik,t_0}}{X_{k,t_0}} / P_{gk,t_0}} \right]$$

where S_g represents 4-digit sectors ($g = 1, \dots, G$) including 6-digit sectors i. *related variety* is a Herfindal index of export diversification and represents the weighted sum of the entropy indicator at the 6-digit level within each 4-digit sector at time t_0.

$$3) \quad unrelated \; variety_{k,t_0} = \sum_{l=1}^{L} P_{lk,t_0} log_2 \left[\frac{1}{P_{lk,t_0}} \right] \quad (A2.3)$$

with $P_{lk,t_0} = \sum_{i \in S_l} x_{ik,t_0} / X_{k,t_0}$ and S_l representing 2-digit sectors ($l = 1, \dots, L$) including 6-digit products i. *unrelated variety* is a Herfindal index measuring the entropy of the 2-digit distribution. It gives a measure of the extent to which a province is characterized by different types of economic sectors at time t_0.

Year t_0	Variable	Obs	Mean	Std. Dev.	Min	Max
	variety	13024	6.511667	1.396737	1.8658	9.379309
2002	*related variety*	13024	8.15183	4.756718	0.051967	26.97201
	unrelated variety	13024	263.3093	146.9643	5.052821	644.4167
	variety	14340	6.418799	1.457188	0.9841455	9.655009
2007	*related variety*	14340	7.430267	4.604068	0	23.5821
	unrelated variety	14340	260.739	150.5167	1.431186	632.7324

Table A2.1: Descriptive statistics of our export diversification indexes. In order to simplify the interpretation of the marginal effect of such variables (wich are not normalized to any value range) on the probability of experiencing random jumps over the product space, we include them in logarithms. Among the 107 provinces taken into consideration during the crisis period, two of them (Medio Campidano and Ogliastra) result to have related variety equal to zero. To avoid the loss of information due to the transformation, for the crisis period we compute the log of such index as $ln(related \; variety + 0.01)$.