The time and frequency of unrelated diversification

Flávio L. Pinheiro[1,*], Dominik Hartmann[2,3], Ron Boschma[4,5], and César A. Hidalgo[6,7,8]

[1] Nova Information Management School (NOVA IMS), Universidade Nova de Lisboa, Portugal
[2] Federal University of Santa Catarina - Department of Economics and International Relations, Brazil
[3] Department of Innovation Economics, University of Hohenheim, Germany
[4] Department of Human Geography and Planning, Utrecht University, the Netherlands
[5] UiS Business School, Stavanger Centre for Innovation Research, University of Stavanger, Norway
[6] Center for Collective Learning, ANITI, University of Toulouse, Toulouse, France
[7] Alliance Business School, University of Manchester, Manchester, UK
[8] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA

This is the accepted author *manuscript of the following article published by Elsevier:*

Pinheiro, F. L., Hartmann, D., Boschma, R., & Hidalgo, C. A. (2021). The time and frequency of unrelated diversification. Research Policy, [104323]. Advanced online publication on 21 July, 2021. https://doi.org/10.1016/j.respol.2021.104323

Funding Information:

The authors acknowledge support from the **Artificial and Natural Intelligence Institute (ANR-19-PI3A-0004), the MIT Media Lab Consortia** and the fund by the **Cooperative Agreement between the Masdar Institute of Science and Technology (Masdar Institute)**, Abu Dhabi, UAE and the Massachusetts Institute of Technology (MIT), Cambridge, This work was also supported by **the Center for Complex Engineering Systems (CCES) at King Abdulaziz City for Science and Technology (KACST), the Massachusetts Institute of Technology (MIT), the São Paulo Research Foundation (PROCESSO 2017/19842-2)**, and **FCT Portugal under the project UIDB/04152/2020 - Centro de Investigação em Gestão de Informação (MagIC)**. The authors are also thankful to Aamena Alshamsi, Pierre-Alexandre Balland, Mary Kaltenberg, and Cristian Jara-Figueroa for helpful insights and discussions, for the feedback obtained during the 4th Geography of Innovation Conference in Barcelona (Jan 31-Feb 2, 2018).

The Time and Frequency of Unrelated Diversification

Flávio L. Pinheiro[1,*], Dominik Hartmann[2,3], Ron Boschma[4,5], and César A. Hidalgo[6,7,8]

[1] Nova Information Management School (NOVA IMS), Universidade Nova de Lisboa, Portugal
[2] Federal University of Santa Catarina - Department of Economics and International Relations, Brazil
[3] Department of Innovation Economics, University of Hohenheim, Germany
[4] Department of Human Geography and Planning, Utrecht University, the Netherlands
[5] UiS Business School, Stavanger Centre for Innovation Research, University of Stavanger, Norway
[6] Center for Collective Learning, ANITI, University of Toulouse, Toulouse, France
[7] Alliance Business School, University of Manchester, Manchester, UK
[8] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA

Abstract. Economic diversification—the process by which locations enter new economic activities—is known to be a combination of related and unrelated diversification. Related diversification is—on average—more frequent, but unrelated diversification is nevertheless considered important to avoid economic lock-in. Here, we study the frequency and timing of unrelated diversification using two international trade datasets at the country level. We find that related diversification is more frequent for countries at low levels of development but becomes less frequent as countries climb the complexity ladder. These findings contribute to our understanding of the role of relatedness in the diversification of economies at different levels of complexity.

Keywords. Unrelated Diversification; Economic Complexity; Economic Development, Relatedness, Diversification

* Correspondence to fpinheiro@novaims.unl.pt

Introduction

While path dependencies are known to constrain the economic development of countries, cities, and regions (Arthur, 1994; Boschma & Lambooy, 1999; Dosi, 1984; Frenken & Boschma, 2007; Hidalgo, Klinger, Barabási, & Hausmann, 2007; Neffke, Henning, & Boschma, 2011) leapfrogging—*i.e.* a country or region skipping stages of development and entering unrelated activities—is also known to occur (Brezis, Krugman, & Tsiddon, 1993; Lee & Lim, 2001; Perez & Soete, 1988; Yoguel, 2015).

Changes in a country's economic structure (Pasinetti, 1981) require the accumulation of knowledge, capabilities, and innovative capacities (Abramovitz, 1986; Fagerberg & Srholec, 2008; Freeman, 1982; Lall, 1992; Perez & Soete, 1988; Saviotti & Pyka, 2004; Schumpeter, 1942). Structural change, thus, favors activities that require inputs that are similar or complementary to those that are already present in an economy. This intuition is at the heart of the idea of *relatedness* (Boschma, 2017; Hidalgo et al., 2007; Neffke & Henning, 2013; Neffke et al., 2011) which estimates the affinity or compatibility between geographies and economic activities. Metrics of relatedness have helped formalize classical ideas in economic geography by allowing scholars to separate diversification events between related events—involving affine activities—and unrelated events—involving less affine activities.

By now, it is well known that—on average—related diversification is much more common. A vast body of literature has shown that countries (Hidalgo et al., 2007), regions (Boschma, Minondo, & Navarro, 2013; Cicerone, McCann, & Venhorst, 2020; Coniglio, Lagravinese, Vurchio, & Armenise, 2018; Gao, Jun, Pentland, Zhou, & Hidalgo, 2021; Neffke et al., 2011; Zhu, He, & Zhou, 2017), and cities (Boschma, Balland, & Kogler, 2015) are more likely to enter activities that are related to the ones that are already present in them. This *principle of relatedness* (Hidalgo et al., 2018) also applies to a diverse gamut of activities, such as products (Hidalgo et al., 2007), industries (Neffke et al., 2011), technologies (Boschma et al., 2015; Boschma & Capone, 2015), and research areas (Guevara, Hartmann, Aristarán, Mendoza, & Hidalgo, 2016). Yet, while the relatedness literature has been successful at documenting economic path dependencies in countries and regions, we still have much to learn about situations in which economies enter unrelated activities (Boschma, 2017). Here, we contribute to this literature by studying how the frequency of unrelated diversification varies with a country's level of economic development.

While unrelated diversification is uncommon (Lee & Lim, 2001; Yoguel, 2015), it is considered important because it can help avoid economic lock-in and provide new long-term opportunities for economic development (Saviotti & Frenken, 2008). But unrelated diversification is hard to achieve because becoming competitive in new activities requires accumulating new capabilities, from human capital to institutions, that may be hard to accumulate all at once (Hausmann & Hidalgo, 2011). Thus, unrelated diversification is not only infrequent, but a risky development endeavor.

So, when should we expect unrelated diversification to take place?

This paper contributes to the growing understanding of unrelated diversification by empirically studying when it happens. We use two datasets that summarize product exports to analyze the relatedness of diversification events at different levels of development—proxied by economic complexity. To answer this question, we introduce a normalization of the relatedness density metric (Hidalgo et al., 2007) that allows us to compare the relative relatedness of diversification events for countries at different levels of development. We find that unrelated diversification events grow together with a country's development, until a relatively high level of complexity.

We explain this change in the frequency of unrelated diversification by looking at the relationship between complexity and relatedness. At low levels of complexity, locations are related to low complexity activities, so unrelated diversification is unlikely, but also desirable as a means to climb the sophistication ladder. As countries develop, unrelated diversification becomes more feasible, and according to theory, optimal for accelerating diversification (Alshamsi, Pinheiro, & Hidalgo, 2018). Yet, at high levels of complexity, relatedness is no longer correlated with low complexity activities, reducing the need to engage in unrelated diversification to continue climbing the complexity ladder. This explains why unrelated diversification is uncommon at low levels of complexity (unfeasible), attractive at medium and medium high levels of complexity (a mean to increase sophistication), and less relevant at the highest levels of complexity (related activities are complex).

These findings contribute to our empirical understanding of unrelated diversification and its connection to economic complexity.

Relatedness and Diversification

Innovation and diversification are key to economic development (Freeman, 1982; Pasinetti, 1981; Saviotti & Pyka, 2004). Hence, it is crucial to understand how countries catch up, diversify, and enter new activities (Abramovitz, 1986; Dosi, 1984; Lall, 1992). Entering new activities require accumulating a variety of inputs, and in particular, knowledge (Romer, 1990, 1994; Weitzman, 1998). This means that organizations and regions compete for niches in knowledge landscapes (Breschi, Lissoni, & Malerba, 2003), while limited by bounded rationality (Simon, 1972) and their absorptive capacity (Cohen & Levinthal, 1990). Together, these ideas imply that organizations and locations (e.g. countries, cities) face a cost of diversification that decreases with the level of relatedness of activities (Atkinson & Stiglitz, 1969; Chatterjee & Wernerfelt, 1991), and hence, should be more likely to enter activities that are similar to the ones they have previously engaged in.

In recent years, a vast body of literature has documented these path dependencies. Originally, this literature focused on geographic spillovers, like those observed in patent data (Audretsch & Feldman, 1996; Autant-Bernard, 2001; Jaffe, Trajtenberg, & Henderson, 1993; Moreno, Paci, & Usai, 2005), and, more recently, in exports data (Bahar, Hausmann, & Hidalgo, 2014). But soon scholars realized that path dependencies extended beyond geographical proximity (Boschma, 2005), pushing a new wave of literature focused on path dependencies that hinged on the relatedness of activities (Frenken & Boschma, 2007; Hidalgo et al., 2007; Neffke et al., 2011). This empirical literature validated the intuition developed in theoretical work by showing that new economic activities typically do not emerge randomly, but build on and combine existing local capabilities (Rigby & Essletzbichler, 1997).

A large number of studies have provided strong evidence supporting the notion that diversification in countries and regions is path-dependent (Boschma, 2017; Hidalgo et al., 2018; Hidalgo et al., 2007). For instance, Neffke et al. (2011) used information on product portfolios of manufacturing plants to show that regions tend to diversify into new industries related to existing local industries. Kogler, Rigby, and Tucker (2013), Rigby (2015), Boschma et al. (2015) and Petralia, Balland, and Morrison (2017), among others, used measures of technological relatedness between patent classes to show that countries and cities enter related technologies. Guevara et al. (2016) did the same for research areas, finding that the new publications of countries, universities, and researchers tend to be in related research areas.

Relatedness, however, is not the only factor shaping the path dependencies of economies. A key contribution of this new literature was the introduction of the idea of product sophistication or complexity (Fleming & Sorenson, 2001; Hidalgo & Hausmann, 2009; Kogut & Zander, 1993). This idea helped formalize the notion that not all diversification events are equal, and hence, it added a new strategic dimension to the study of economic diversification. Products with higher levels of complexity are associated with higher levels of income (Hidalgo & Hausmann, 2009) and lower levels of income inequality (Hartmann, Guevara, Jara-Figueroa, Aristarán, & Hidalgo, 2017). Moreover, countries and regions with more sophisticated export baskets grew—on average—faster than similar countries with less sophisticated export baskets (Hausmann et al., 2014; Hidalgo & Hausmann, 2009). So, the process of economic development was expanded from one of diversification *per se* to one of diversifying into sophisticated activities. This helped revive twentieth-century ideas of structural transformation, developed originally for a few aggregate sectors (agriculture, manufacturing, and services), but expanded into thousands of different products.

In principle, it should be beneficial for a country to build comparative advantages in complex technologies. Yet, for many countries, this is difficult to achieve when the productive capabilities needed are unrelated to the current ones (Balland, Boschma, Crespo, & Rigby, 2019). Thus, a key goal for the literature on relatedness is to understand the ability of countries or regions to defy the principle of relatedness and enter sophisticated yet unrelated economic activities.

Unrelated diversification has been a topic of growing interest. Xiao, Boschma, and Andersson (2018) showed that European regions with higher innovation capacity are more inclined to enter less related industries. Boschma and Capone (2016) observe that Western European economies also tend to diversify more into unrelated industries than Eastern European economies. In terms of the agents responsible for unrelated diversification, Neffke, Hartog, Boschma, and Henning (2018) and Elekes, Boschma, and Lengyel (2019) show that entrepreneurs and firms coming from outside a region are more likely to introduce unrelated diversification and to shift a region's specialization pattern. This is especially true for subsidiaries of large firms, which can rely on resources from the parent organization that may be unavailable in their host region (Cortinovis, Crescenzi, & Van Oort, 2020; Crescenzi, Gagliardi, & Iammarino, 2015). Boschma and Capone (2015) explored the role of institutions in unrelated diversification in 23 countries, showing that countries with more liberal and less coordinated forms of capitalism are more likely to diversify into unrelated activities. Montresor and Quatraro (2017) found that regions with a strong presence of key enabling technologies tended to diversify into more unrelated technologies. Petralia et al. (2017) showed that high-income countries have a higher tendency to diversify into unrelated and sophisticated technologies.

Despite this growing literature, we still have little understanding when countries and regions are more likely to enter unrelated activities (Boschma, 2017). Here, we attempt to fill these gaps by analyzing the timing at which countries enter activities with varying degrees of relatedness, and whether countries and regions that succeed in entering more unrelated activities also enter higher complexity activities.

Our work is informed by recent theoretical work by Alshamsi et al. (2018) that helps predict when unrelated diversification should be more beneficial. At lower stages of economic development, countries mostly catch-up by entering simple and related products. Yet, at intermediate stages of economic development, economies have a window of opportunity where unrelated diversification can trigger an acceleration of a country's subsequent structural transformation (Alshamsi et al., 2018). This difficult transition can slow down economic growth momentarily but can plant the seeds for future growth. When countries fail to achieve this transition, they risk falling into the middle-income trap (Eichengreen, Park, & Shin, 2013; Felipe, Abdon, & Kumar, 2012; Gill & Kharas, 2015; Hartmann, Bezerra, Lodolo, & Pinheiro, 2020; Hartmann, Zagato, Gala, & Pinheiro, 2021; Kharas & Kohli, 2011; Lee, 2013).

The remainder of the paper is structured as follows. The next section introduces the data and methods used throughout the paper. Then, we present our results, starting by the introduction and validation of a measure of relative relatedness, and then, by using this metric to explore the connection between economic complexity and unrelated diversification. We then explore this relationship through statistical models and by introducing a measure of relative complexity. The last discussion summarizes the results and discusses some policy implications.

Data and Methods

International Trade Data

We use two datasets on international trade obtained from the *Observatory of Economic Complexity* (http://oec.world), the long times series SITC-4 rev2 dataset, and the HS92 dataset. The SITC-4 datasets merge data compiled by Feenstra, Lipsey, Deng, Ma, and Mo (2005) (between 1962 to 2000) and the UN Comtrade (between 2001 and 2017). It contains information on the exports of 774 products from 1962 to 2017. The HS92 (4-digit) dataset, compiled by BACI (Gaulier & Zignago, 2010), contains detailed information on the exports of 1,241 products between 1995 to 2018. HS92 is a more recent trade classification than SITC (which is a legacy classification), making the HS92 dataset better for more recent periods.

These datasets have been cleaned in order to reduce noise resulting from underreporting, from variations in the size of the economies of countries and products, and from changes in classification (this applies to the SITC dataset). The filters include discarding all countries that have a population of fewer than 1.25 million, a total yearly trade below USD 1 billion in 2010, and economies that are known for which reliable and consistent data is not available (Iraq, Chad, Macau, and Afghanistan). Moreover, we apply sequential filters on a year-per-year basis that discard flows valued at less than 5,000 USD, products whose export value is equal to zero for more than 80% of the countries, products with global exports of less than USD 10 million, and countries whose exports equal to zero for 95% of the products. After these steps, the SITC (HS92) dataset captures the trade of 764 (1211) products between 124 (130) countries, which represents 73.65% (75.7%) of global GDP and 96.57% (92.12%) of global trade in 2010.

Data on country GDP, population size, and human capital were sourced from the Penn World Tables (PWT 9.1). GDP data comes from real GDP National Accounts, measuring GDP at constant 2011 USD national prices (Feenstra, Inklaar, & Timmer, 2015).

Specialization Matrix and Entry Events

We consider a country to be a significant exporter of a product if it has a Revealed Comparative Advantage (RCA) in it. Formally, the RCA of a country c in a product p is equivalent to the location quotient and is defined as

$$R_{cp} = \left(\frac{X_{cp}}{\sum_{p'} X_{cp'}}\right) / \left(\frac{\sum_{c'} X_{c'p}}{\sum_{c'p'} X_{c'p'}}\right) \quad (1)$$

where X_{cp} is a matrix summarizing the exports of country c in product p, and R_{cp} is a matrix of specialization or Revealed Comparative Advantages.

We call the set of products present in a country with an RCA greater one ($R_{cp} \geq 1$) as that countries' product basket. Conversely, products with RCA below a one ($R_{cp} < 1$) represent countries' option set (O_c), *i.e.*, they are not present yet in a country product basket but could be, thus, present opportunities of development.

We say that a country (c) enters a new product (p), between years y and y', when there is an increase in RCA from $R_{cp} < 1$ in year y to $R_{cp} \geq 1$ in year y'. However, since RCA time series can have a significant amount of noise (e.g., due to exchange rate and commodity price fluctuations), we consider two additional conditions. First, we consider a backward condition requiring a location to have $R_{cp} < 1$ for activity p for the Δ years preceding y, and a forward condition requiring that a location c maintains $R_{cp} \geq 1$ for activity p during Δ years. In the main manuscript we will consider a standard $\Delta = 4$, in Appendix B we show the robustness of our findings for variations of this parameter and RCA thresholds. These conditions help reduce the number of false-positive observations. We only consider entry events as those events that satisfy these conditions.

Economic Complexity

The Economic Complexity Index (ECI) is a measure of the factors that explain the geography of economic activities that can be estimated directly from the spatial distribution of economic data (Hidalgo, 2021). Here, we compute the ECI and the Product Complexity Index (PCI) following (Hidalgo & Hausmann, 2009). The ECI of a country is computed as the average complexity of its activities (*i.e.* the average PCI of a country's exports). Likewise, the PCI of an activity is the average complexity of the countries where it is present. This circular argument gives rise to the following iterative mapping:

$$\mathrm{ECI}_c = \frac{1}{k_c} \sum\nolimits_p M_{cp} \mathrm{PCI}_p \tag{2a}$$

$$\mathrm{PCI}_p = \frac{1}{k_p} \sum\nolimits_c M_{cp} \mathrm{ECI}_p \tag{2b}$$

Replacing (2b) into (2a) leads to an eigenvalue equation whose solution is a location's ECI:

$$\mathrm{ECI}_c = \sum\nolimits_p \frac{M_{cp}}{k_p k_c} \sum\nolimits_c M_{cp} \mathrm{ECI}_c \tag{3}$$

ECI offers a measure of the combination of factors that drive the specialization pattern of a location, whereas PCI measure the combination of factors required by a product or industry (Hidalgo, 2021). Like ECI, PCI can be computed by solving the following eigenvalue equation:

$$\mathrm{PCI}_c = \sum\nolimits_c \frac{M_{cp}}{k_p k_c} \sum\nolimits_p M_{cp} \mathrm{PCI}_c \tag{4}$$

Results and Discussion

Principle of Relatedness

We estimate relatedness using the density approach introduced by Hidalgo et al. (2007). We start by estimating the proximity between a pair of products. To that end, we define M_{cp}, a matrix that is equal to one if a location has RCA in a product ($R_{cp} \geq 1$), and zero otherwise. The number of activities present in a country, or its diversity, can be computed from M_{cp}, as $M_c = \sum_p M_{cp}$. Conversely, the ubiquity of a product is equal to the number of countries where the product is present with $R_{cp} \geq 1$ ($M_p = \sum_c M_{cp}$). The proximity between a pair of products p and p' ($\phi_{pp'}$) is the minimum of the conditional probability that a country has RCA in both of them:

$$\phi_{\mathrm{pp'}} = \frac{\sum_c M_{cp} M_{cp'}}{\max(M_p, M_{p'})} \tag{5}$$

For instance, a pair of products with a proximity of 0.4 means that there is at least a 40% chance that a country with RCA ($R_{cp} \geq 1$) in one product has RCA in both of them (Hidalgo et al., 2007). Measuring proximity through co-location is a method that takes into consideration outcomes rather than the multiple factors that may mediate relatedness (Hidalgo et al., 2007). Yet, more recently, there have been efforts to unpack proximities into multiple channels, such as labour mobility, input-output linkages, research collaborations, et cetera (Bahar, Rosenow, Stein, & Wagner, 2019; Balland & Boschma, 2021; Delgado, Porter, & Stern, 2015; Diodato, Neffke, & O'Clery, 2018; Farinha, Balland, Morrison, & Boschma, 2019; Hidalgo, 2021; Jara-Figueroa, Jun, Glaeser, & Hidalgo, 2018).

We can estimate the relatedness between products and regions by dividing the proximities from a product to all of the products present in a country's basket by the sum of the proximities from that product to all products. The relatedness density, ω_{cp}, of product p in country c is defined as:

$$\omega_{\mathrm{cp}} = \frac{\sum_{p'} M_{cp'} \phi_{pp'}}{\sum_{p'} \phi_{pp'}} \tag{6}$$

Figure 1 shows the probability that a country enters a product as a function of their relatedness. While this figure verifies the basic finding of the principle of relatedness (Hidalgo et al., 2018), it also shows that relatedness (ω_{cp}) cannot be readily used for cross-country comparison. More complex economies tend to have higher relatedness values across all activities; so high-relatedness entry events are skewed to those of high complexity economies (green segments in **Figure 1**). Similarly, low relatedness entry events are skewed towards those of low complexity economies (red segments). Nevertheless, a relatedness value of 0.3 may be considered low for a high complexity economy (e.g., an unrelated entry event) and high for a low complexity economy (related entry event). This motivates the introduction of a metric of relative relatedness ($\widetilde{\omega}_{cp}$) that makes diversification events comparable across countries.

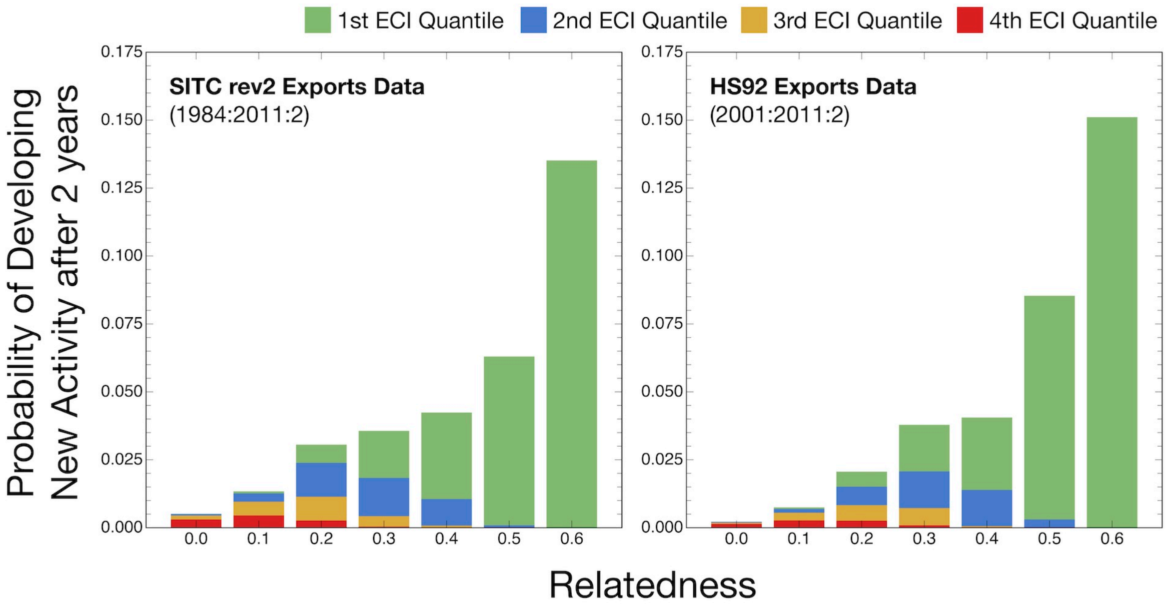

Figure 1 – Probability of entering a new activity in two years as a function of relatedness density, estimated for the SITC (left) and HS92 (right) international trade product classifications. More details on the procedure for the identification of new products can be found in Appendix A.

Relative Relatedness

Unlike relatedness (ω_{cp}), which is an absolute measure, relative relatedness ($\widetilde{\omega}_{cp}$) is a measure that compares the relatedness of a countries' new activities with the average relatedness of the activities in a countries' option set[1] (activities with $R_{cp} < 1$). The relative relatedness of country c in product p, in year y is computed as a simple z-transform:

$$\widetilde{\omega}_{cp} = \frac{\omega_{cp} - \sum_{p'} {}^{\omega_{cp'}}/_{N_{O_c}}}{\sigma_{\omega_{cp'}}} \tag{7}$$

[1] Alternatively, we could have made the measure relative by ranking products according to their level of relatedness, but we chose not to use rankings since these are uniformly distributed and carry no clear benefit to our analysis.

where $\sum_{p'} \omega_{cp'}/N_{O_c}$ is the simple average relatedness of all products (p') in O_c, $N_{O_c}$is the number of products in the set of opportunities of country c, and $\sigma_{\omega_{cp'}}$ is the standard deviation of the relatedness of the products (p') in O_c. Relative relatedness ($\widetilde{\omega}_{cp}$) is comparable across countries because it indicates whether a location enters a product that was more or less related than the average product in its option set. Note that we are omitting the year index from all the formulas, but these quantities are measured independently on a year-by-year basis.

Figure 2 shows the probability that a country enters a new product as a function of the relative relatedness. While results are consistent with the principle of relatedness, in that countries are more successful in entering activities for which they have relatively higher relatedness, relative relatedness overcomes the limitations shown in **Figure 1**. In particular, we now see that for the entire range of relative relatedness's we have a more evenly distributed representation of countries with different levels of complexity. This helps mitigate the bias introduced by an unnormalized metric of relatedness.

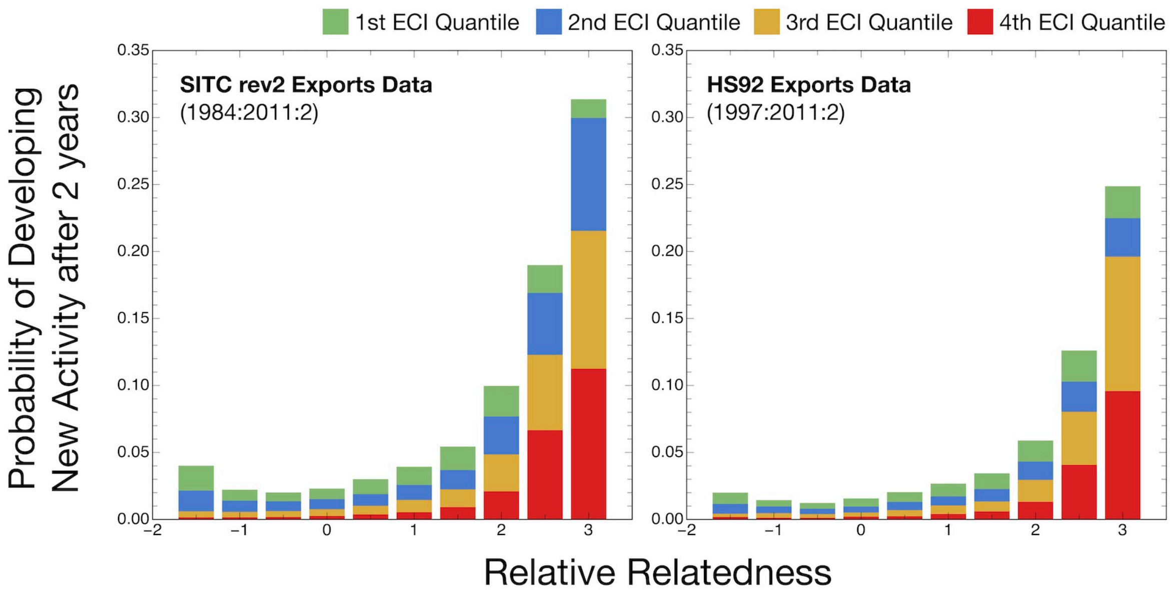

Figure 2 – Probability of entering a new activity in two years as a function of the relative relatedness density, estimated for the SITC (left) and HS92 (right) international trade product classifications.

Unrelated Diversification and Economic Complexity

Equipped with our normalized measure of relatedness, we now explore the timing when locations enter relatively more unrelated activities. To that end, we define $\Omega_{c,y\rightarrow y+2}$ as the average relative relatedness $\widetilde{\omega}_{cp}$ of the products entered by country c between years y and $y+2$.

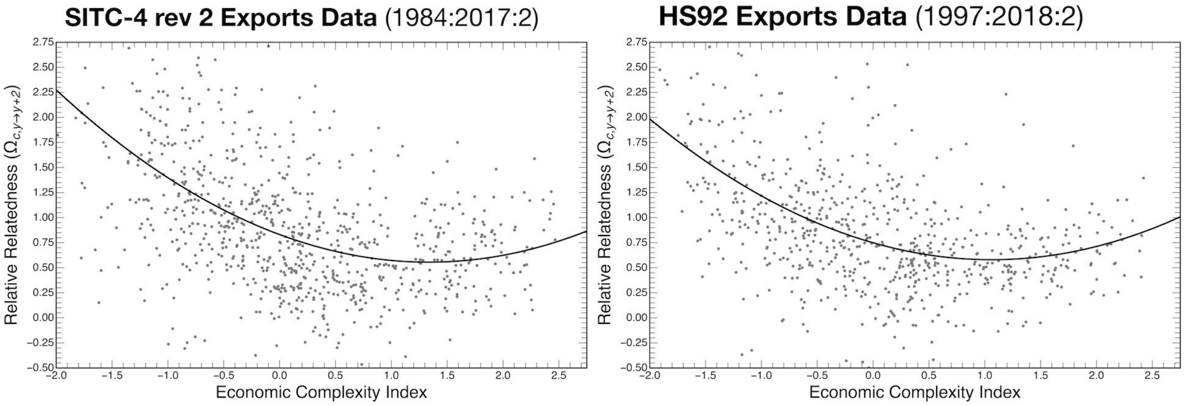

Figure 3 – Average deviation in the relatedness of new activities ($\Omega_{c,y\rightarrow y+2}$) as a function of the Economic Complexity Index (ECI) of countries using the SITC-4 (left) and HS92 (right) datasets. Solid lines show the best quadratic polynomial fit.

Figure 3 shows the average relative relatedness of newly developed products and industries in a two-year interval as a function of a location's economic complexity index (ECI). In all four scenarios, the relatedness of newly developed products and industries exhibits a U-shaped relationship, with relative relatedness decreasing with increasing ECI. That is, at low levels of economic complexity, new entries tend to be mostly at high levels of relative relatedness (>1.5), whereas at high levels of economic complexity, the average relative relatedness is much lower (~0.75).

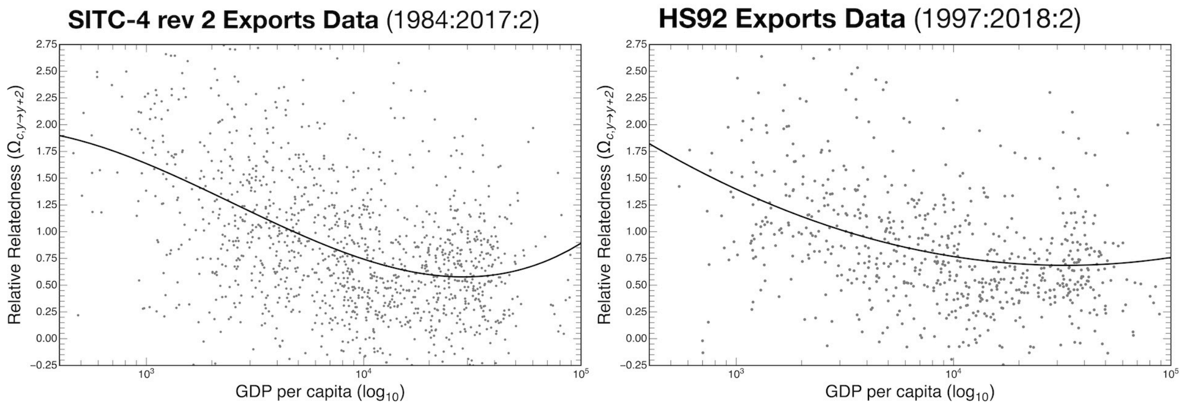

Figure 4 – Average deviation in the relatedness of newly developed products ($\Omega_{c,y\rightarrow y+2}$) as a function of the GDP per capita of countries using the SITC-4 (left) and HS92 (right) datasets. Solid lines show the best third-order polynomial fit.

Figure 4 extends the analysis of **Figure 3** but uses GDP per capita as a metric of development (for the datasets for which this data was available). Although a similar behavior is captured, the effect is not as pronounced as with the Economic Complexity Index.

Next, we explore the relationship between relative relatedness and development using a model that regresses the relative relatedness of the new products developed as function of linear and quadratic terms for a Human Capital Index, ECI, and GDP per capita; while controlling for year fixed effects. Formally, the models take the form:

$$\begin{aligned}\Omega_{c,y\rightarrow y+y'} &= \alpha_1 \times \mathrm{ECI}_{cy} + \alpha_2 \times \mathrm{ECI}_{cy}^2 + \alpha_3 \times Log\big(GDP_{cy}\big) + \alpha_4 \times Log^2\big(GDP_{cy}\big) \\ &+\alpha_5 \times Log\big(\mathrm{HC_{cy}}\big) + \alpha_6 \times Log^2\big(\mathrm{HC}_{cy}\big) + \mu_y + \epsilon_{cy}\end{aligned} \tag{8}$$

where the dependent variable $\Omega_{c,y\rightarrow y+y'}$ represents the average relative relatedness of the newly products developed by country c between year y and y'. The remaining variables indicate the Economic Complexity Index (ECI_{cy}); GDP per capita (GDP_{cy}); and Human Capital (HC_{cy}) of country c in year y. Finally, μ_y represents a dummy variable that controls for year fixed effects and ϵ_{cy} is the error term.

Table 1 – Summarizes the several models that regress the relative relatedness ($\Omega_{c,y\rightarrow y+2}$) of newly developed products as a function of linear and quadratic terms of the Economic Complexity Index (ECI), GDP per capita (GDP), and human capital (HC). Models S1 to S4 focus on the SITC dataset. Models H1 to H4 focus on the HS92 dataset. Values of independent variables are measured at year y. We consider the time interval between 1984 and 2011 (SITC) and 2001 to 2011 (HS92). We use a validation interval of $\Delta = 4$ years. In order to avoid overlapping in the data we consider only every other year, which means that only even years are considered in the SITC and odd years in the HS92.

	Relative relatedness of newly developed products ($\Omega_{c,y\rightarrow y+2}$)							
	(S1)	(S2)	(S3)	(S4)	(H1)	(H2)	(H3)	(H4)
ECI	-0.404***			-0.343***	-0.302***			-0.267***
	(0.018)			(0.031)	(0.021)			(0.042)
ECI^2	0.184***			0.138***	0.158***			0.148***
	(0.014)			(0.016)	(0.018)			(0.021)
Log $\mathrm{GDP_{pc}}$		-1.662***		-0.383		-0.838***		0.275
		(0.205)		(0.235)		(0.272)		(0.319)
$\mathrm{Log^2\ GDP_{pc}}$		0.081***		0.019		0.037**		-0.015
		(0.012)		(0.013)		(0.015)		(0.017)
HC			-1.968***	-0.869***			-1.131***	-0.433
			(0.182)	(0.218)			(0.229)	(0.293)
HC^2			0.332***	0.181***			0.163***	0.072
			(0.037)	(0.044)			(0.046)	(0.058)
Constant	0.916***	9.290***	3.548***	3.768***	0.737***	5.370***	2.589***	0.080
	(0.060)	(0.904)	(0.214)	(0.959)	(0.051)	(1.206)	(0.275)	(1.302)
Observations	1,182	1,182	1,182	1,182	607	607	607	607
R^2	0.313	0.226	0.218	0.335	0.272	0.155	0.174	0.276
Adjusted R^2	0.304	0.216	0.208	0.325	0.263	0.145	0.164	0.263
F Statistic	35.456***	22.676***	21.640***	30.876***	31.926***	15.726***	18.011***	20.637***
Notes:						*p<0.1; **p<0.05; ***p<0.01		
						All models control for year fixed effects		

Table 1 summarizes the results obtained for four different specifications of Equation 8. Models S1 to S4 report results stemming from the SITC dataset, while models H1 to H4 show results for the HS92 product classification dataset. Results support the behavior characterized

in **Figures 3** and **4**. The combination of linear and quadratic terms reveals that economies at lower levels of development (lower complexity, GDP per capita, or human capital) enter mostly related activities. But as economies progress, the role of relatedness decreases, becoming relatively unimportant. For instance, considering the coefficients of equation S1, we expect the relative relatedness of the new activities of an economy with $\text{ECI} = -1.0$ to be $\widetilde{\omega}_{cp}(\text{ECI} = -1.0) = 1.504$, whereas an economy with $\text{ECI} = 1.0$ is expected to enter—on average—products with a relative relatedness of $\widetilde{\omega}_{cp}(\text{ECI} = -1.0) = 0.696$. These are relatively large effects, since one unit in $\widetilde{\omega}_{cp}$ equals one standard deviation. Similar calculations can be done to estimate the expected level of relative relatedness as a function of GDP per capita and human capital.

When comparing the different predictors (ECI, GDPpc, and HC), we find that ECI is the predictor that explains most of the variance in the relationship ($R^2 = 0.304$ for SITC and $R^2 = 0.263$ for HS), and that the addition of GDPpc and HC as predictors does not substantially increase the explanatory power of the model. This suggests that the information about a country's development stage that explains the observed variation in the relatedness of entries is better captured by ECI than by GDPpc or HC.

In sum, we show evidence that related diversification has a stronger pull at lower levels of development, and that this pull eases as countries climb the complexity ladder.

The Complexity of Unrelated Diversification Events

Similar to the definition of relative relatedness, we can introduce a measure of relative product complexity or just relative complexity for simplicity. We recall that the Product Complexity Index (PCI) is a measure of the combination of factors needed to engage in an activity, which is also indicative of knowledge intensity. Activities with higher (lower) PCI contribute more (less) to the knowledge intensity of a region (which is reflected in the ECI).

The relative complexity of a product ($\widetilde{\text{PCI}}_{cp}$) compares its complexity to the average complexity (PCI) of all products in a countries' option set. Formally, the relative complexity of product p in country c is defined as:

$$\widetilde{\text{PCI}}_{cp} = \frac{\text{PCI}_{cp} - \sum_{p'} {\text{PCI}_{cp'}}/{N_{O_c}}}{\sigma_{\text{PCI}_{cp'}}} \tag{9}$$

where $\sum_{p'} \text{PCI}_{cp'}/N_{O_c}$ is the simple average PCI of all products (p') in O_c, and $\sigma_{\text{PCI}_{cp'}}$ is the standard deviation of the PCI of the same set of products. This measure of relative complexity indicates whether a location entered an activity that was more or less complex than the average activity that was not yet present in that location. The rationale for relative complexity is that the same product (*e.g.,* Grated Cheese, $\text{PCI} = -0.05$) can represent an increase in sophistication for a low complexity economy (*e.g.,* Paraguay, $\text{ECI} = -0.45$) and a decrease in complexity for a high complexity economy (*e.g.,* Finland, $\text{ECI} = 1.56$). We note that while PCI is constant for each product across all countries, that is, it does not depend on a country's product basket, the relative $\widetilde{\text{PCI}}_{cp}$ of a product varies from country to country.

Next, we characterize the development paths of different economies combining both relative relatedness and relative complexity. To that end, and as we did above with relative relatedness, we define $\Pi_{c,y\to y+2}$ as the average $\widetilde{\text{PCI}}_{cp}$ of the products developed by country c between years y and $y+2$. **Figure 5** shows the average relative relatedness and complexity of countries.

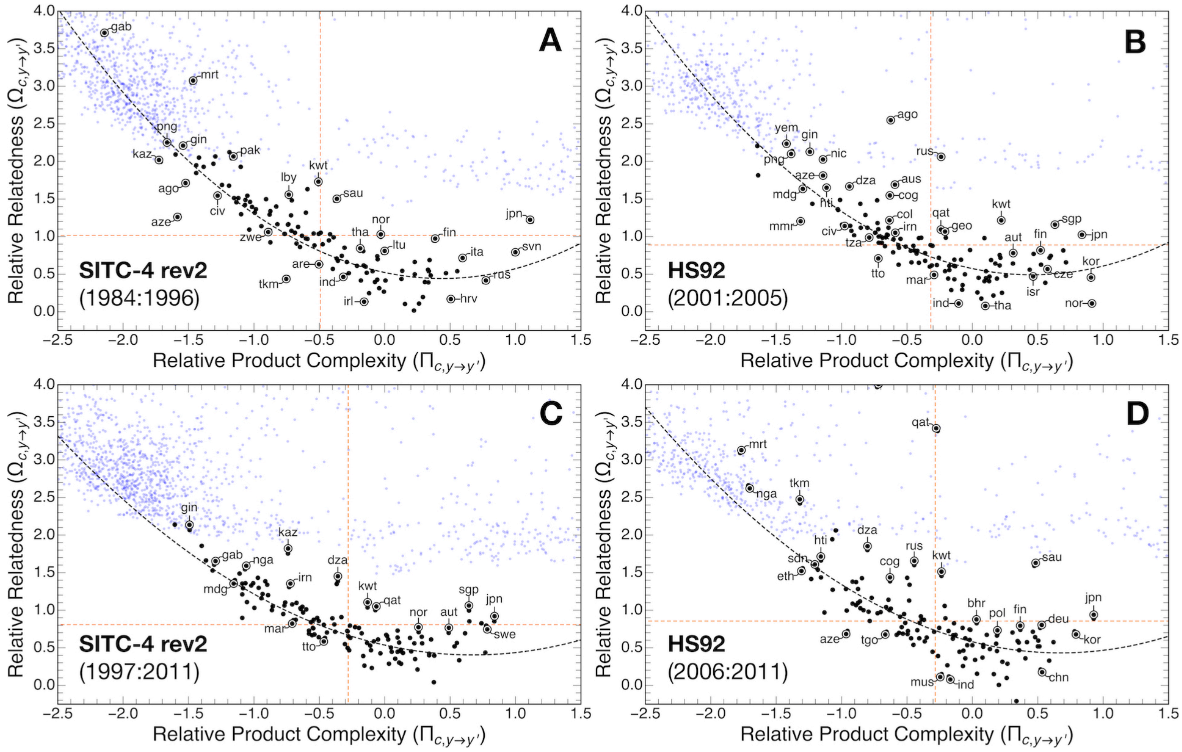

Figure 5 – Average Development Directions. Panels show the relationship between the average relative *relatedness* and relative complexity of the newly developed products aggregated by country. Results represent results of 2-year interval jumps in a year-by-year basis during two different time intervals for each dataset. Blue dots provide a control based on a simulation were the entry events of a location are replaced by a synthetic set composed of the most related products in the option set. Orange dashed lines show the average observed values on each dimension, and the point they cross indicate the centroid of all countries. Black dashed line shows the best quadratic polynomial for eye reference.

Across both datasets we find a similar relationship between relative relatedness and relative complexity. Countries that enter more unrelated activities also enter relatively complex activities. This is a non-trivial pattern. If countries would enter activities at random, they would all cluster around (0,0) (because of the definitions of $\widetilde{\omega}_{cp}$ and $\widetilde{\mathrm{PCI}}_{cp}$). If countries would enter the most related products only (blue dots simulation), they would cluster on the top of the chart. In this simulation, we estimate the relative relatedness and relative complexity of products of each country if they would have entered only in the most related products. For this purpose, we consider the same number of new products as observed in the real datasets.

Stages of Economic Diversification

To explain the patterns identified above, we study the correlation between the relative relatedness and relative complexity of a location's option set. This is defined using the Pearson correlation coefficient as:

$$\rho_c = corr\left(\widetilde{\mathrm{PCI}}_{pc}, \widetilde{\omega}_{pc}\right) = \frac{\sum_{O_{c,y}}\left(\widetilde{\mathrm{PCI}}_{pc} - \langle\widetilde{\mathrm{PCI}}_{pc}\rangle\right)\left(\widetilde{\omega}_{pc} - \langle\widetilde{\omega}_{pc}\rangle\right)}{\sum_{O_{c,y}}\left(\widetilde{\mathrm{PCI}}_{pc} - \langle\widetilde{\mathrm{PCI}}_{pc}\rangle\right)^2\left(\widetilde{\omega}_{pc} - \langle\widetilde{\omega}_{pc}\rangle\right)^2} \tag{10}$$

In this equation, a negative correlation ($\rho_c < 0$) implies that a location is related to simple activities, while a positive correlation ($\rho_c > 0$) implies that location is related to complex activities. **Figure 6A**, **B**, and **C** illustrate how such correlations identify different relationships between the relative relatedness and relative complexity of the development opportunities of South Korea for different years.

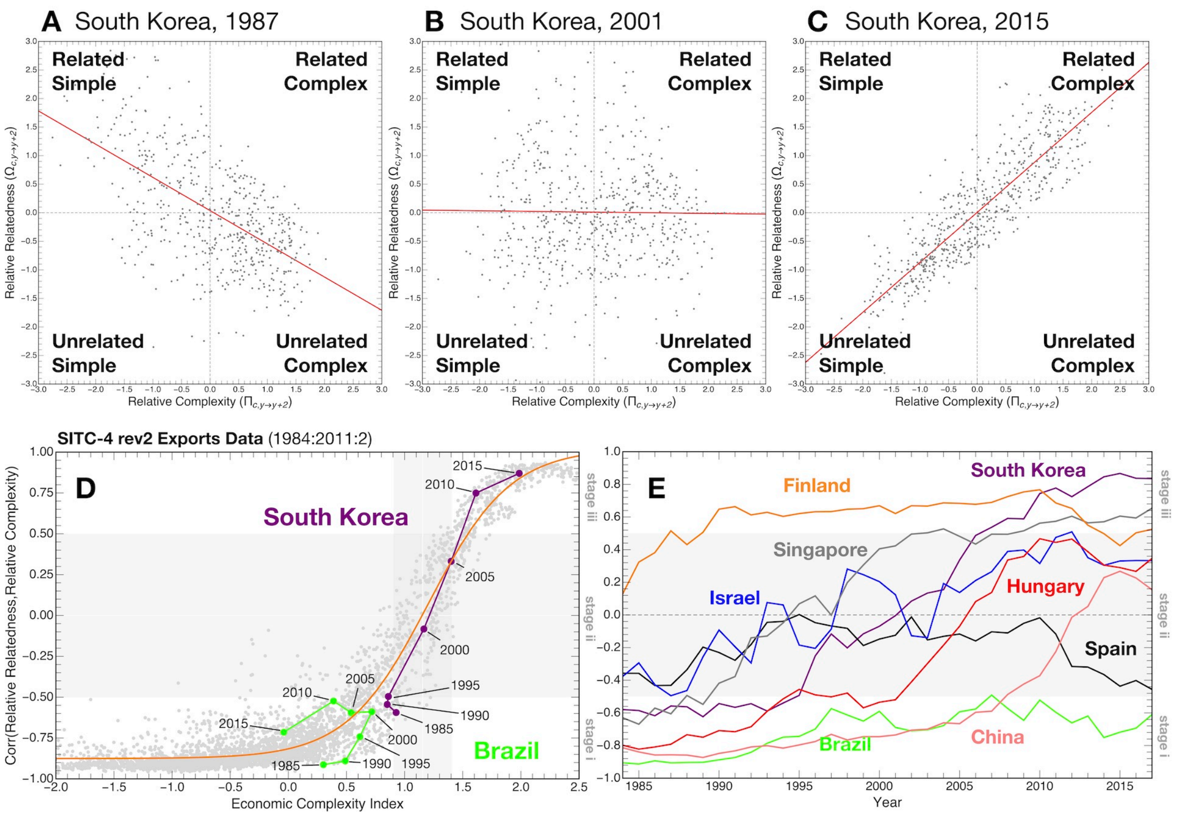

Figure 6 – Panel A, B, and C illustrate scenarios that result in the possible correlations between the relative *relatedness* and relative *complexity* of products of the option set of South Korea in the years 1987 (A), 2001 (B) and 2015 (C). South Korea provides a representative example of the behavior that can be found across countries with similar correlations and ECI levels. The option set corresponds to all products over which a location has an RCA lower than 1.0; in other words, the development opportunities of a location. Panel D shows the relationship between the correlation and the ECI of each location, results for HS92 can be found in the appendix. Panel E shows the evolution of the correlation for some particularly successful countries (Singapore, Israel, South Korea, Finland, China, and Hungary) and representative examples of countries that got stuck (Spain and Brazil).

Interestingly, we find that the correlation exhibits an S-Shaped behavior (**Figure 6D**) and that the correlation is nil precisely at the stage of development (ECI level) at which locations become more related to complex products (ECI greater than approximately 1.1). More importantly, the shape of **Figure 6D** clarifies that most complex activities entries can be explained by high relative relatedness from high complexity economies. **Figure 6E** shows the time evolution in the correlation of several countries that successfully managed to transitioned between stages (South Korea, Finland, Israel, Hungary, China, and Singapore) and two representative examples of countries that did not (Spain and Brazil). In contrast, low complexity economies are related to low complexity products and are therefore more likely to enter them. But why would countries develop unrelated activities at the point in which the correlation between relative complexity and relative relatedness vanishes?

Recently, Alshamsi et al. (2018) showed that accelerating diversification requires countries to enter unrelated activities while at an intermediate level of development. For countries at an intermediate level of development, this optimal strategy means that entering unrelated products may be more beneficial in the long run because of the future diversification opportunities they provide. So even though countries at an intermediate level of development have related and unrelated opportunities, entering the more unrelated products may be more strategic because of the subsequent diversification opportunities these may provide.

The strategic use of unrelated diversification at medium and medium high levels of development is also in line with the experience of some Asian countries, like South Korea and Singapore, where the state has actively invested in promoting the development of entirely new sectors (Fagerberg & Srholec, 2008). This happened in combination with programs to attract foreign direct investment that may have helped countries move into more complex and unrelated products.

Conclusions

The literature on economic diversification has successfully explained why related diversification is more common, while also providing ample evidence supporting this claim (Hidalgo et al., 2018). Yet, countries sometimes deviate from the principle of relatedness, albeit infrequently, by entering products that are on average less related than the average product in their option sets. When does this happen? And why? As a research community, we still have a limited understanding of unrelated diversification (Boschma, 2017). To explore this question, we introduced relative measures of relatedness and complexity, and conducted a cross-location comparison of empirical diversification paths followed by countries.

Our study shows that related diversification is more frequent at lower levels of complexity, and saturates at medium and high stages of development. This coincides with a period of time when economies experience a structural transformation: from being more related to simple products to being more related to complex ones.

These findings introduce an important distinction. For the most part, research on relatedness has failed to ask whether relatedness plays a different role, or even if it occurs at different times, in economies at high or low levels of complexity. Our work shows it definitely does. This means that diversification policy advice for economies at different levels of complexity should take this into consideration. Low complexity economies, suffering a strong pull from relatedness, have more limited diversification options than higher complexity economies, where the pull of relatedness is weaker. This is good news for relatively high complexity economies, since it tells them that they are not limited to their current pattern of specialization in the product space. What is probably more challenging, is finding policy options for low complexity economies. This requires new creative efforts to identify and promote avenues for unrelated diversification that could help open new paths for relatedness. At the same time, it is key for these economies to avoid the temptation of engaging in huge capital investments projects, which have been the norm in several low complexity and resource rich economies.

While, unrelated diversification might be beneficial under certain circumstances, how to accomplish it effectively is still an open question. Investments in research and education (Xiao et al., 2018), or in key enabling technologies (Montresor & Quatraro, 2017), might be a recommendable means to support unrelated diversification. But external factors can also induce unrelated diversification in countries, as other studies have shown. These can be, for instance, entrepreneurs and (multi-national) firms that come from other countries (Crescenzi et al., 2015; Elekes et al., 2019; Neffke et al., 2018), high-skilled migrants that bring new ideas and new experiences to a country (Caviggioli, Jensen, & Scellato, 2020; Fassio, Montobbio, & Venturini, 2019), or international linkages (trade relationships, research networks) that can provide access to capabilities that are missing in a particular country (Balland & Boschma, 2021).

Our study also shows that the policy implications of relatedness go beyond efforts to identify activities and should instead think about timing. Several efforts to apply relatedness in an international or regional development context have focused on which sectors to target (Balland et al., 2019; Hausmann et al., 2014). But when to time unrelated activities is critical. This is a different policy perspective, since it does not focus on sectors (*e.g.*, clusters of related activities), but rather, conceptualizes related and unrelated diversification as a portfolio allocation problem. This focus on timing (Alshamsi et al., 2018), rather than activities, suggests investing more in related activities at early stages of development, and more in unrelated activities at medium and medium high levels of development.

Acknowledgments

The authors acknowledge support from the Artificial and Natural Intelligence Institute (ANR-19-PI3A-0004), the MIT Media Lab Consortia and the fund by the Cooperative Agreement between the Masdar Institute of Science and Technology (Masdar Institute), Abu Dhabi, UAE and the Massachusetts Institute of Technology (MIT), Cambridge, This work was also supported by the *Center for Complex Engineering Systems* (CCES) at *King Abdulaziz City for Science and Technology* (KACST), the *Massachusetts Institute of Technology* (MIT), the São

Paulo Research Foundation (PROCESSO 2017/19842-2), and FCT Portugal under the project UIDB/04152/2020 - *Centro de Investigação em Gestão de Informação* (MagIC). The authors are also thankful to Aamena Alshamsi, Pierre-Alexandre Balland, Mary Kaltenberg, and Cristian Jara-Figueroa for helpful insights and discussions, for the feedback obtained during the 4th Geography of Innovation Conference in Barcelona (Jan 31-Feb 2, 2018).

Appendix A – Additional Results

One of the key findings in the Economic Complexity literature is the relationship between Economic Complexity Index (ECI) and GDP per capita. Namely, more complex economies exhibit higher values of GDP per capita. **Figure A1** summarizes the main findings for the two datasets.

Figure A2 shows, the frequency distribution of the relative relatedness (A and B) and relative complexity (C and D) of newly developed products for all countries in each dataset: SITC (A and C); HS92 (B and D).

Figure A3 extends the results from Figure 6 of the main manuscript for the HS92 dataset. Showing the S-curve (panel A), time evolution of the correlation (panel B), and three examples of the correlations (panels C, D, and E) for the same examples discussed in the main manuscript.

Figure A4 shows the fraction of unrelated jumps (with relative relatedness below zero) both averaged by country (black) and disaggregated (orange). Results are in accordance with the findings discussed in the manuscript. In particular, that countries tend to enter more unrelated activities at a intermediate stage of economic development.

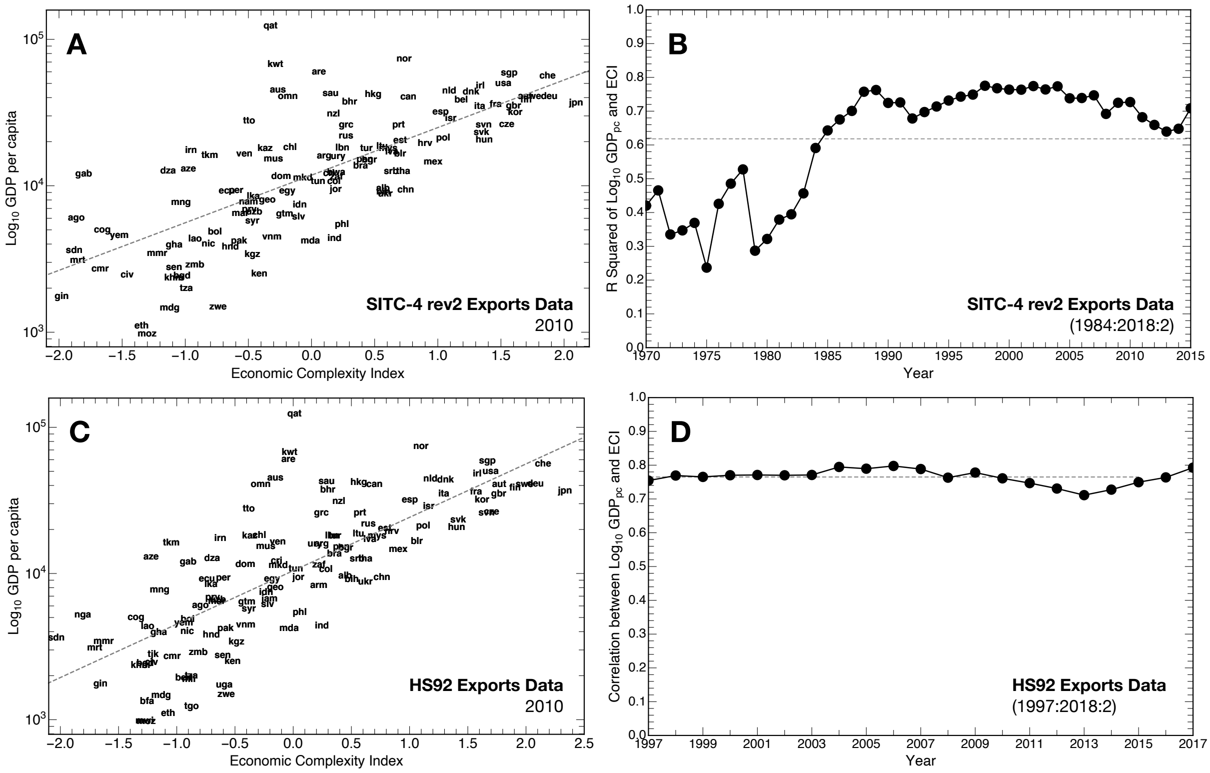

Figure A1 – Panel A and C show the correlation between location Economic Complexity Index (ECI) and Log of GDP per capita for a year (shown in the panel) as an illustrative example. Panel B and D show the correlation between ECI and Log of GDP per capita per year of observation. In the latter dashed horizontal line corresponds to the average correlation.

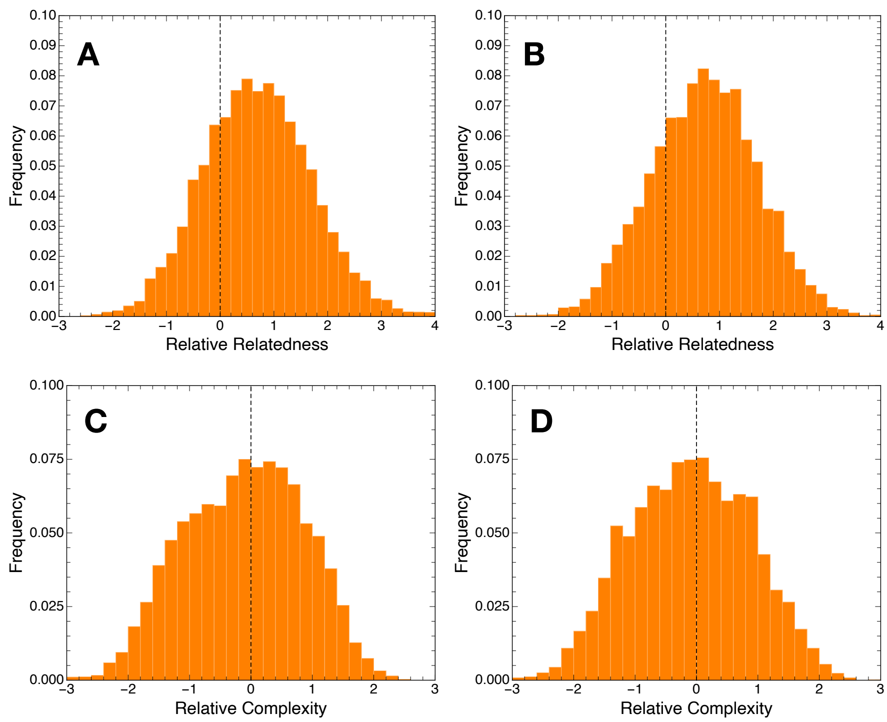

Figure A2 – Panel A and B show the distribution of the relative relatedness of identified jumps, and Panel C and D show the distribution of relative complexity of identified jumps. Panels A and C report results for the SITC dataset, while Panel B and D results for the HS92 dataset.

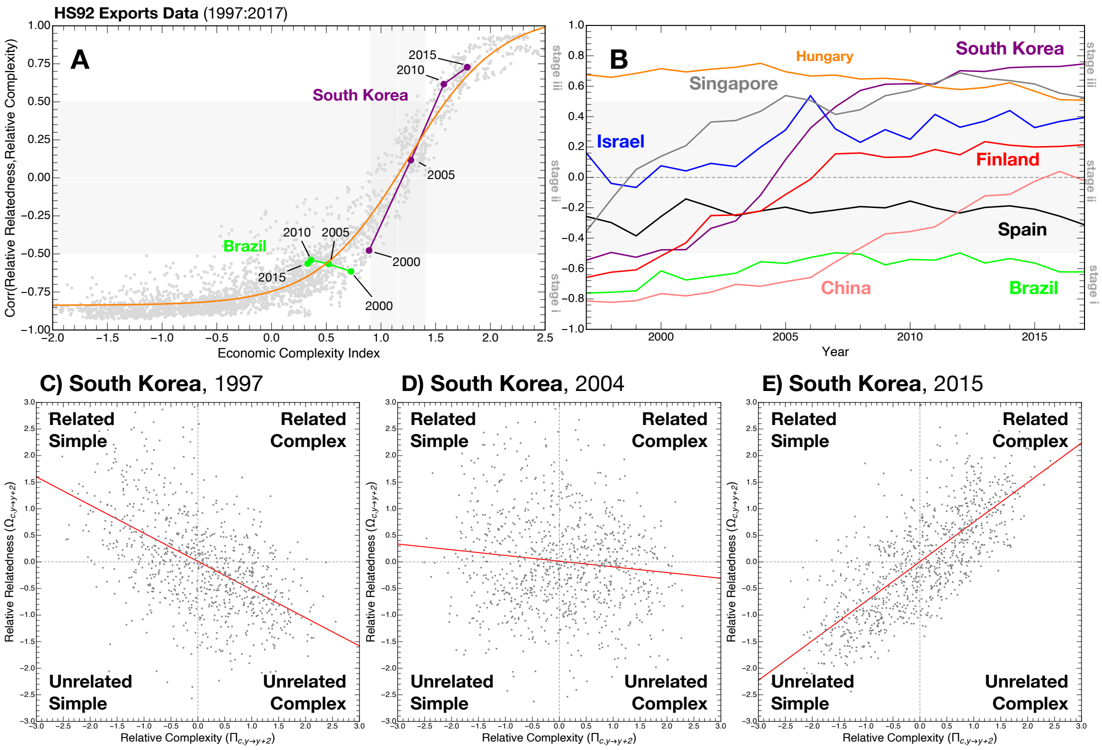

Figure A3 – Panel A shows the relationship between the correlation of the relative relatedness with the relative complexity of the option set of each country and the Economic Complexity Index. Results reveal the S-shaped relationship, with the trajectories of two countries (South Korea and Brazil) highlighted. Panel B shows the temporal evolution of the correlation for eight examples. Panel C, D, and E show representative examples of the relationships between relative relatedness (Y-axis) and relative complexity (X-Axis) of the option set of South Korea exhibiting different correlation levels characteristic of each stage.

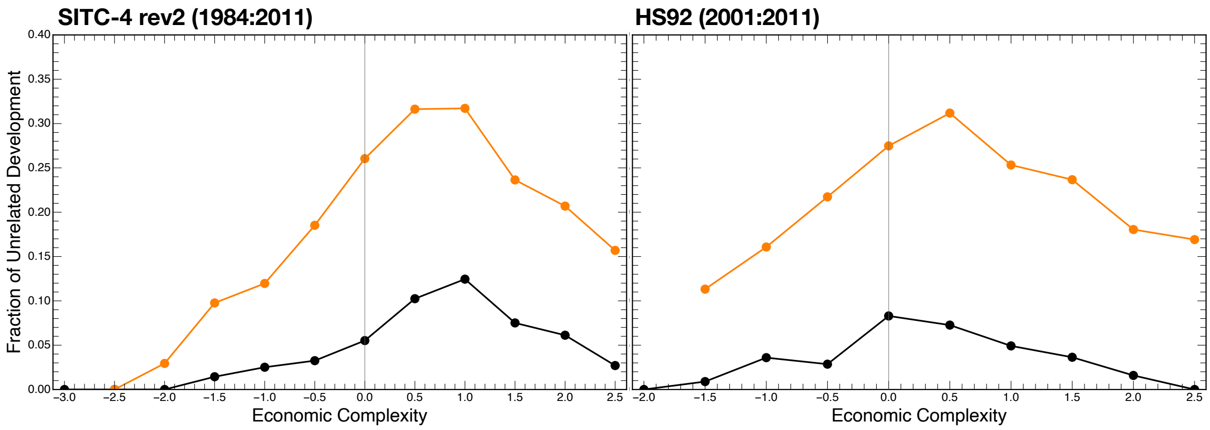

Figure A4 –Fraction of unrelated development entries for each dataset. Results are shown both when aggregated by country (Black line), in which we consider the average relative relatedness of all new entries of a country in a year, and desegregated (Orange), where we look at each entry event at product-level independently.

Appendix B – Robustness Checks

We start by showing how results could are impacted by different choices of the time interval (y to y') used to estimate the jumps. We show how these impacts the results in Table 1 for model S4 and H4. These models regress the average relative relatedness $\Omega_{c,y\rightarrow y+2}$ of jumps against a set of linear and quadratic terms of macro-economic indicators.

Table B1 – Results for SITC model S4 and HS92 model H4 for different values of the time interval in the estimation of the Jumps.

	Relative relatedness of newly developed products ($\Omega_{c,y\rightarrow y+2}$)									
	$y' - y = 2$		$y' - y = 3$		$y' - y = 4$		$y' - y = 5$		$y' - y = 10$	
	(S4a)	(H4a)	(S4b)	(H4b)	(S4c)	(H4c)	(S4d)	(H4d)	(S4e)	(H4e)
ECI	-0.343***	-0.267***	-0.413***	-0.277***	-0.469***	-0.340***	-0.400***	-0.254***	-0.365***	-0.386***
	(0.031)	(0.042)	(0.040)	(0.053)	(0.047)	(0.061)	(0.058)	(0.073)	(0.086)	(0.116)
ECI^2	0.138***	0.148***	0.173***	0.166***	0.186***	0.143***	0.161***	0.136***	0.176***	0.115*
	(0.016)	(0.021)	(0.021)	(0.026)	(0.025)	(0.030)	(0.031)	(0.036)	(0.051)	(0.059)
Log GDP_{pc}	-0.383	0.275	-0.096	0.230	-0.390	0.046	-0.178	-1.211**	-0.588	0.546
	(0.235)	(0.319)	(0.297)	(0.408)	(0.360)	(0.487)	(0.421)	(0.597)	(0.647)	(0.937)
Log^2 GDP_{pc}	0.019	-0.015	0.001	-0.014	0.020	0.001	0.006	0.063*	0.025	-0.034
	(0.013)	(0.017)	(0.016)	(0.022)	(0.020)	(0.026)	(0.023)	(0.032)	(0.036)	(0.052)
HC	-0.869***	-0.433	-0.890***	-0.677*	-0.873***	-0.672	-1.471***	-0.302	-0.768	-1.265
	(0.218)	(0.293)	(0.275)	(0.363)	(0.337)	(0.453)	(0.407)	(0.546)	(0.660)	(0.802)
HC^2	0.181***	0.072	0.195***	0.127*	0.180***	0.110	0.309***	0.060	0.151	0.298*
	(0.044)	(0.058)	(0.056)	(0.072)	(0.068)	(0.089)	(0.084)	(0.108)	(0.141)	(0.162)
Constant	3.768***	0.080	2.672**	0.679	3.784**	1.195	3.719**	6.831***	5.164**	-0.130
	(0.959)	(1.302)	(1.221)	(1.672)	(1.483)	(2.004)	(1.727)	(2.472)	(2.611)	(3.843)
Observations	1,182	607	734	384	546	295	375	190	145	90
R^2	0.335	0.276	0.398	0.321	0.392	0.313	0.408	0.332	0.454	0.256
Adjusted R^2	0.325	0.263	0.386	0.304	0.378	0.293	0.391	0.306	0.426	0.202
F Statistic	30.876***	20.637***	33.901***	19.620***	28.648***	16.258***	25.035***	12.897***	16.297***	4.752***

Notes: *p<0.1; **p<0.05; ***p<0.01

All models control for year fixed effects

Next, we look at how the choice of the validation time interval (Δ) impacts the results in Table 1 of the main manuscript. To that end, Table B3 shows the results for $\Delta = 2$, 4 (which has been used in the main manuscript), and 6 for $y' - y = 2$.

Table B2 – Results for SITC model S4and HS92 model H4 for three different validation intervals ($\Delta = 2$, 4, and 6) used in the validation of Jumps.

	Relative relatedness of newly developed products ($\Omega_{c,y\rightarrow y+2}$)					
	$\Delta = 2$		$\Delta = 4$		$\Delta = 6$	
	(S4a)	(H4a)	(S4b)	(H4b)	(S4c)	(H4c)
ECI	-0.331***	-0.197***	-0.343***	-0.267***	-0.371***	-0.191***
	(0.024)	(0.034)	(0.031)	(0.042)	(0.038)	(0.059)
ECI^2	**0.146*****	**0.085*****	**0.138*****	**0.148*****	**0.176*****	**0.131*****
	(0.013)	(0.017)	(0.016)	(0.021)	(0.020)	(0.029)
Log GDP_{pc}	-0.144	-0.423*	-0.383	0.275	-0.159	-0.468
	(0.188)	(0.254)	(0.235)	(0.319)	(0.285)	(0.432)
Log^2 GDP_{pc}	0.008	0.027*	0.019	-0.015	0.007	0.027
	(0.010)	(0.014)	(0.013)	(0.017)	(0.016)	(0.023)
HC	-0.665***	-0.406*	-0.869***	-0.433	-0.809***	-0.705*
	(0.176)	(0.242)	(0.218)	(0.293)	(0.265)	(0.413)
HC^2	0.134***	0.066	0.181***	0.072	0.154***	0.104

	(0.036)	(0.048)	(0.044)	(0.058)	(0.054)	(0.080)
Constant	2.345***	2.912***	3.768***	0.080	2.815**	3.796**
	(0.764)	(1.034)	(0.959)	(1.302)	(1.176)	(1.800)
Observations	1,316	650	1,182	607	962	372
R^2	0.367	0.200	0.335	0.276	0.336	0.270
Adjusted R^2	0.358	0.187	0.325	0.263	0.323	0.252
F Statistic	39.513***	14.537***	30.876***	20.637***	26.468***	14.859***
Notes:						*p<0.1; **p<0.05; ***p<0.01
						All models control for year fixed effects

Next we replicate the results of the main text for different thresholds of RCA, that is, when considering that the presence of a product in a country happens at different RCA thresholds. We start by showing the results for RCA = 0.9 (Figures B1 to B4 and Table B3) and for RCA = 1.1 (Figures B5 to B8 and Table B4).

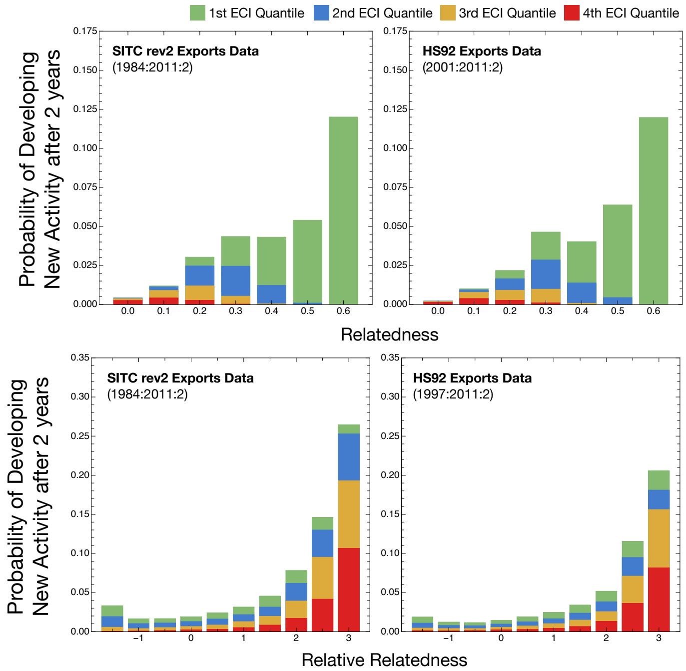

Figure B1 – Probability of entering a new activity in two years as a function of *relatedness density* (top panels) and relative relatedness (bottom panels), estimated for the SITC (left) and HS92 (right) international trade product classifications.

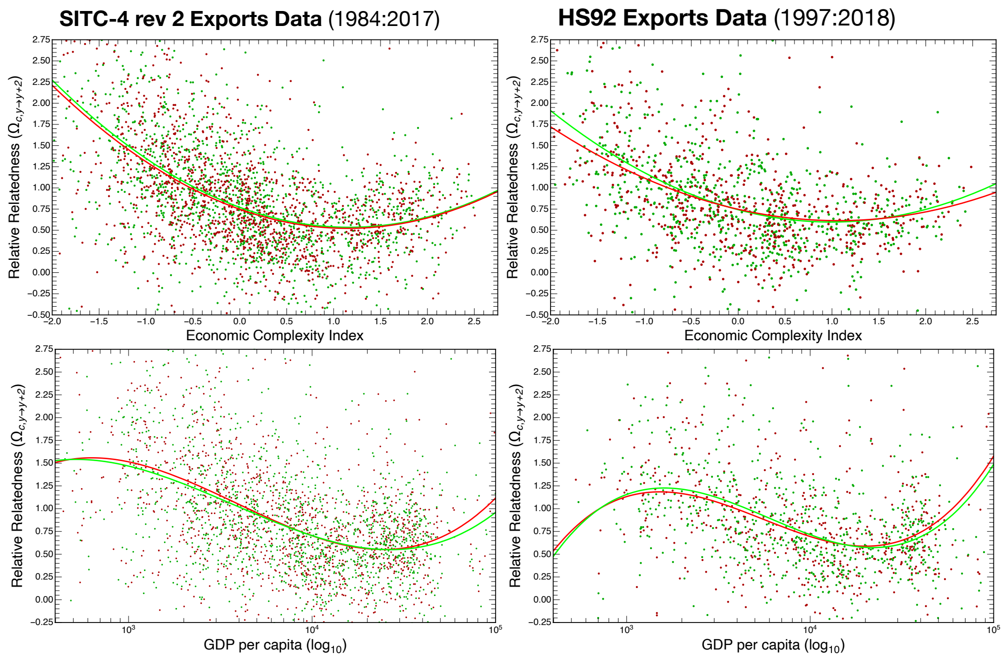

Figure B2 – Top panels show the Average deviation in the relatedness of newly developed products ($\Omega_{c,y\to y+2}$) as a function of the ECI of countries using the SITC-4 (left) and HS92 (right) datasets. Solid lines show the best second-order polynomial fit. Bottom panel shows similar results for GDP per capita. Red/Green points and curves concern the results on odd/even years.

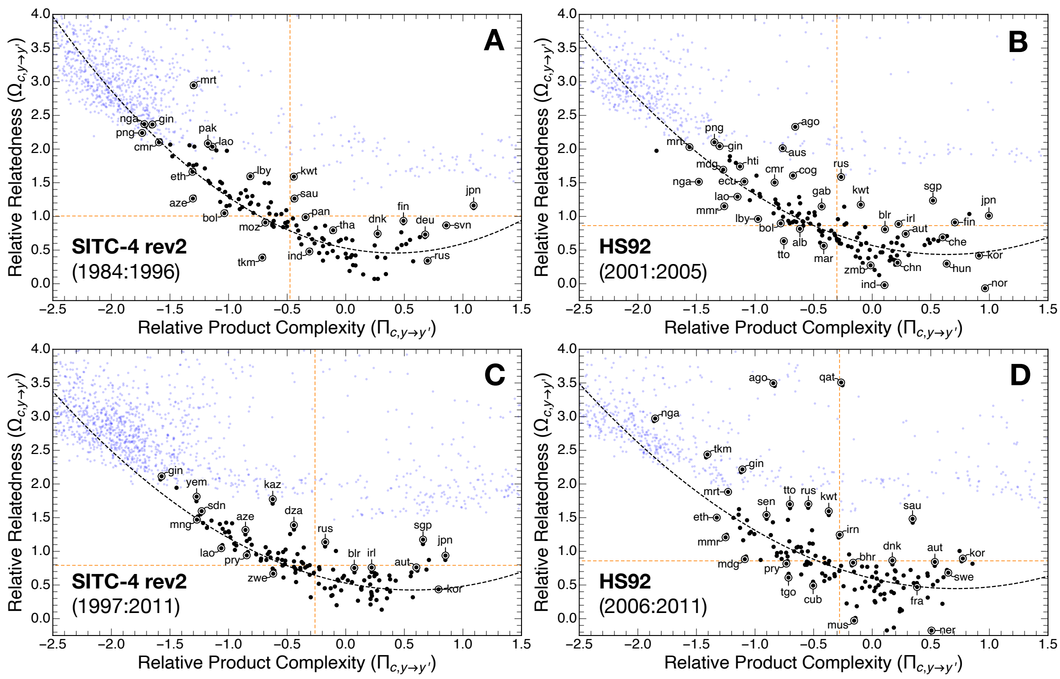

Figure B3 – Average Development Directions. Panels show the relationship between the average relative relatedness and relative complexity of the newly developed products aggregated by country. Results represent results of 2-year interval jumps in a year-by-year basis during two different time intervals for each dataset. Blue dots provide a control based on a simulation where the entry events of a location are replaced by a synthetic set composed of the most related products in the option set. Orange dashed lines show the average observed values on each dimension, and the point they cross indicate the centroid of all countries. Black dashed line shows the best quadratic polynomial for eye reference.

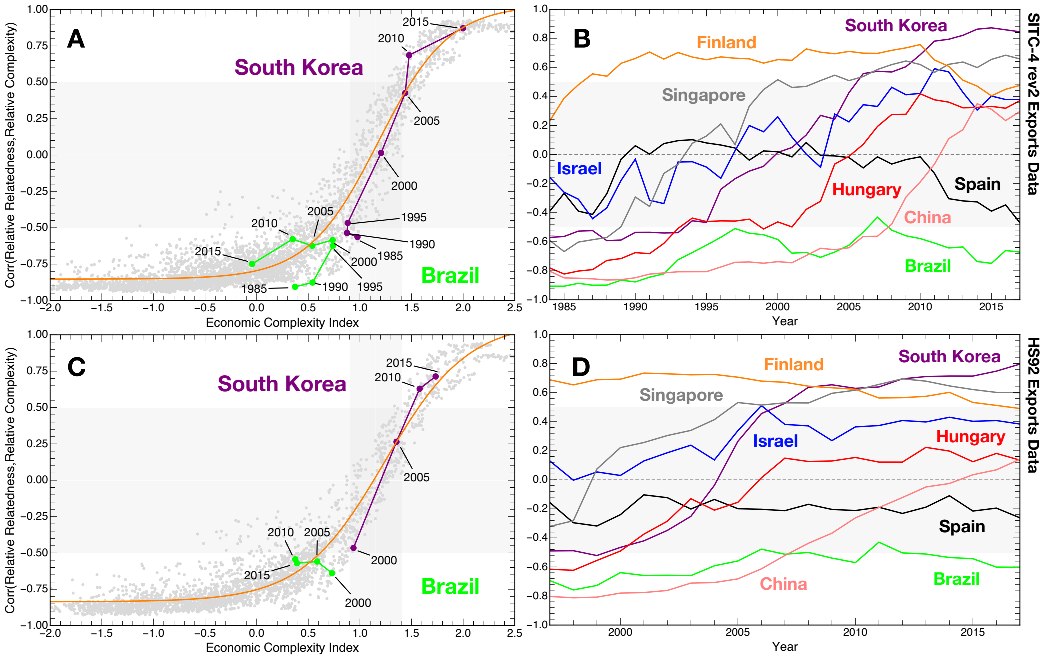

Figure B4 – Left panels (A and C) show the relationship between the correlation (of relative relatedness and complexity of the option set) and the ECI of each country, results for SITC (top) and HS92 (bottom). Panels on the right side show the evolution of the correlation for some particularly successful countries (Singapore, Israel, South Korea, Finland, China, and Hungary) and representative examples of countries that got stuck (Spain and Brazil).

Table B3 – Summarizes the several models that regress the relative relatedness ($\Omega_{c,y\to y+2}$) of newly developed products as a function of linear and quadratic terms of the Economic Complexity Index (ECI), GDP per capita (GDP), and human capital (HC). Models S1 to S4 focus on the SITC dataset. Model H1 to H4 focus on the HS92 dataset. Values of independent variables are measured at year y. We consider the time interval between 1984 and 2011 (SITC) and 2001 to 2011 (HS92) using a validation interval of $\Delta = 4$ years. In order to avoid overlapping in the data we consider only every other year, which means that only even years are considered in the SITC and odd years in the HS92.

	Relative relatedness of newly developed products ($\Omega_{c,y\to y+2}$)							
	(S1)	(S2)	(S3)	(S4)	(H1)	(H2)	(H3)	(H4)
ECI	-0.381***			-0.280***	-0.260***			-0.187***
	(0.017)			(0.030)	(0.027)			(0.056)
ECI^2	0.164***			0.118***	0.129***			0.089***
	(0.014)			(0.016)	(0.023)			(0.028)
Log GDP$_{pc}$		-1.682***		-0.567**		-1.336***		-0.173
		(0.191)		(0.227)		(0.352)		(0.419)
Log^2 GDP$_{pc}$		0.082***		0.028**		0.065***		0.011
		(0.011)		(0.013)		(0.019)		(0.023)
HC			-1.808***	-0.668***			-1.557***	-0.834**
			(0.173)	(0.210)			(0.302)	(0.398)
HC^2			0.296***	0.132***			0.247***	0.138*
			(0.036)	(0.042)			(0.060)	(0.077)
Constant	0.970***	9.425***	3.413***	4.527***	0.697***	7.476***	3.029***	2.578
	(0.058)	(0.842)	(0.205)	(0.927)	(0.057)	(1.571)	(0.365)	(1.752)
Observations	1,198	1,198	1,198	1,198	373	373	373	373
R^2	0.315	0.260	0.236	0.340	0.211	0.153	0.190	0.231
Adjusted R^2	0.306	0.251	0.226	0.329	0.200	0.142	0.179	0.212

F Statistic	35.456***	27.742***	24.351***	31.882***	19.650***	13.272***	17.168***	12.113***
Notes:								*p<0.1; **p<0.05; ***p<0.01
								All models control for year fixed effects

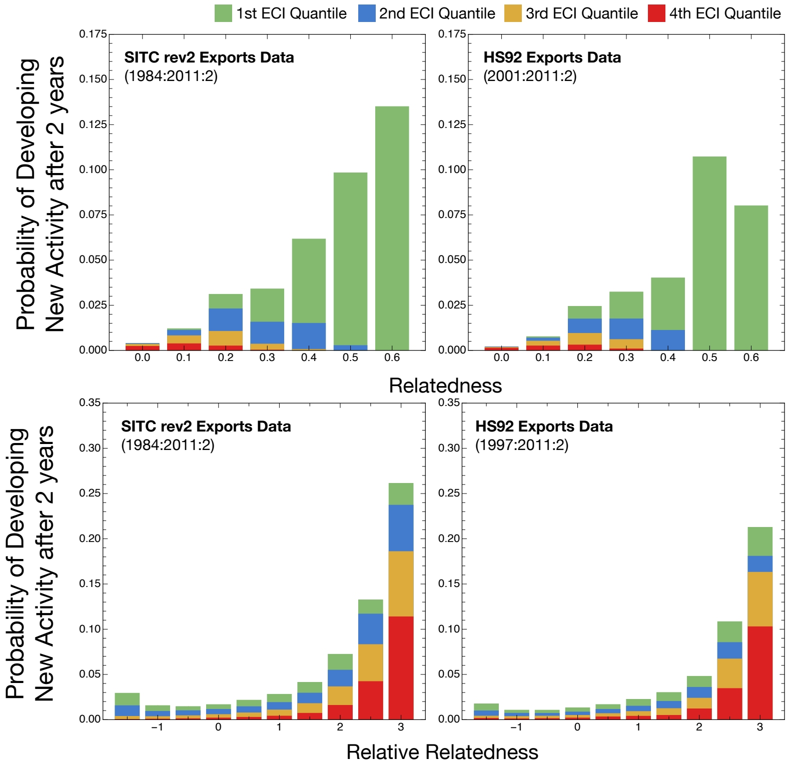

Figure B5 – Probability of entering a new activity in two years as a function of relatedness density (top panels) and relative relatedness (bottom panels), estimated for the SITC (left) and HS92 (right) international trade product classifications.

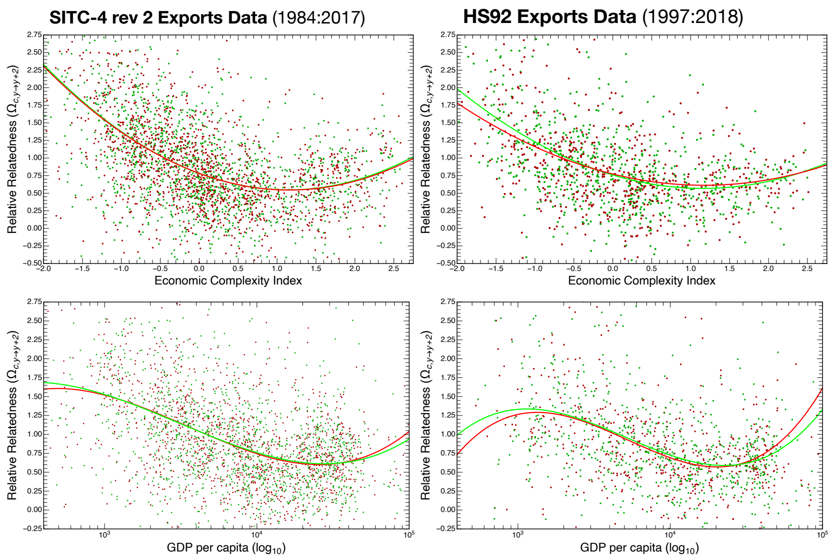

Figure B6 – Top panels show the Average deviation in the relatedness of newly developed products ($\Omega_{c,y\to y+2}$) as a function of the ECI of countries using the SITC-4 (left) and HS92 (right) datasets. Solid lines show the best second-order polynomial fit. Bottom panel shows similar results for GDP per capita. Red/Green points and curves concern the results on odd/even years.

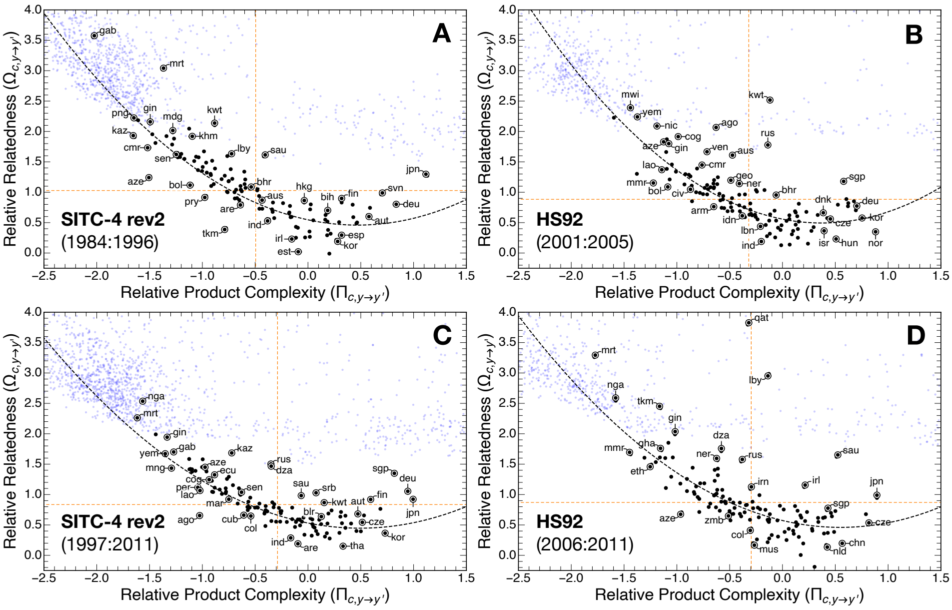

Figure B7 – Average development directions. Panels show the relationship between the average relative *relatedness* and relative complexity of the newly developed products aggregated by country. Results represent results of 2-year interval jumps in a year-by-year basis during two different time intervals for each dataset. Blue dots provide a control based on a simulation where the entry events of a location are replaced by a synthetic set composed of the most related products in the option set. Orange dashed lines show the average observed values on each dimension, and the point they cross indicate the centroid of all countries. Black dashed line shows the best quadratic polynomial for eye reference.

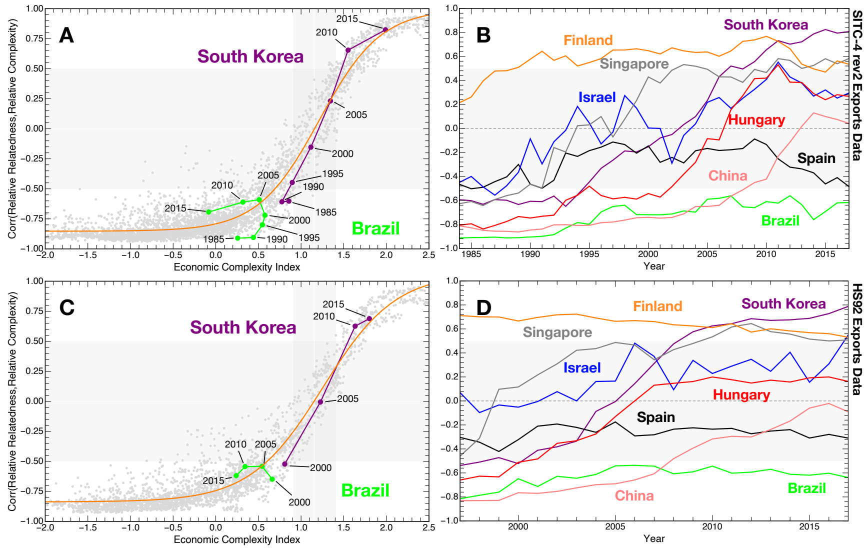

Figure B8 – Left panels (A and C) show the relationship between the correlation (of relative relatedness and complexity of the option set) and the ECI of each country, results for SITC (top) and HS92 (bottom). Panels on the right side show the evolution of the correlation for some particularly successful countries (Singapore, Israel, South Korea, Finland, China, and Hungary) and representative examples of countries that got stuck (Spain and Brazil).

Table B4 – Summarizes the several models that regress the relative relatedness ($\Omega_{c,y\to y+2}$) of newly developed products as a function of linear and quadratic terms of the Economic Complexity Index (ECI), GDP per capita (GDP), and human capital (HC). Models S1 to S4 focus on the SITC dataset. Model H1 to H4 focus on the HS92 dataset. Values of independent variables are measured at year y. We consider the time interval between 1984 and 2011 (SITC) and 2001 to 2011 (HS92) using a validation interval of $\Delta = 4$ years. In order to avoid overlapping in the data we consider only every other year, which means that only even years are considered in the SITC and odd years in the HS92.

	Relative relatedness of newly developed products ($\Omega_{c,y\rightarrow y+2}$)							
	(S1)	(S2)	(S3)	(S4)	(H1)	(H2)	(H3)	(H4)
ECI	-0.404***			-0.318***	-0.297***			-0.251***
	(0.019)			(0.032)	(0.032)			(0.065)
ECI^2	0.180***			0.139***	0.156***			0.088***
	(0.014)			(0.016)	(0.026)			(0.032)
Log GDP$_{pc}$		-1.723***		-0.690***		-1.842***		-0.391
		(0.206)		(0.240)		(0.388)		(0.459)
Log^2 GDP$_{pc}$		0.084***		0.036***		0.093***		0.025
		(0.012)		(0.013)		(0.021)		(0.025)
HC			-1.701***	-0.510**			-2.012***	-1.364***
			(0.185)	(0.222)			(0.322)	(0.424)
HC^2			0.276***	0.097**			0.338***	0.244***
			(0.038)	(0.045)			(0.064)	(0.083)
Constant	0.892***	9.481***	3.219***	4.728***	0.660***	9.674***	3.551***	3.945**
	(0.062)	(0.909)	(0.218)	(0.978)	(0.061)	(1.732)	(0.391)	(1.931)
Observations	1,165	1,165	1,165	1,165	373	373	373	373
R^2	0.291	0.219	0.202	0.310	0.199	0.148	0.198	0.238
Adjusted R^2	0.282	0.209	0.192	0.299	0.188	0.137	0.187	0.219
F Statistic	31.416***	21.534***	19.441***	27.127***	18.235***	12.776***	18.065***	12.577***
Notes:								*p<0.1; **p<0.05; ***p<0.01
								All models control for year fixed effects

Appendix C – Country List

In this appendix we provide the list of countries included in the analysis of each dataset after the pre-processing and cleaning steps.

The countries present in the SITC rev2 dataset include: Angola (ago); Albania (alb); United Arab Emirates (are); Argentina (arg); Australia (aus); Austria (aut); Azerbaijan (aze); Belgium (bel); Bangladesh (bgd); Bulgaria (bgr); Bahrain (bhr); Bosnia and Herzegovina (bih); Belarus (blr); Bolivia (bol); Brazil (bra); Botswana (bwa); Canada (can); Switzerland (che); Chile (chl); China (chn); Cote d'Ivoire (civ); Cameroon (cmr); Republic of the Congo (cog); Colombia (col); Costa Rica (cri); Cuba (cub); Czech Republic (cze); Germany (deu); Denmark (dnk); Dominican Republic (dom); Algeria (dza); Ecuador (ecu); Egypt (egy); Spain (esp); Estonia (est); Ethiopia (eth); Finland (fin); France (fra); Gabon (gab); United Kingdom (gbr); Georgia (geo); Ghana (gha); Guinea (gin); Greece (grc); Guatemala (gtm); Hong Kong (hkg); Honduras (hnd); Croatia (hrv); Hungary (hun); Indonesia (idn); India (ind); Ireland (irl); Iran (irn); Israel (isr); Italy (ita); Jordan (jor); Japan (jpn); Kazakhstan (kaz); Kenya (ken); Kyrgyzstan (kgz); Cambodia (khm); South Korea (kor); Kuwait (kwt); Laos (lao); Lebanon (lbn); Libya (lby); Sri Lanka (lka); Lithuania (ltu); Latvia (lva); Morocco (mar); Moldova (mda); Madagascar (mdg); Mexico (mex); Macedonia (mkd); Burma (mmr); Mongolia (mng); Mozambique (moz);

Mauritania (mrt); Mauritius (mus); Malaysia (mys); Namibia (nam); Nigeria (nga); Nicaragua (nic); Netherlands (nld); Norway (nor); New Zealand (nzl); Oman (omn); Pakistan (pak); Panama (pan); Peru (per); Philippines (phl); Papua New Guinea (png); Poland (pol); North Korea (prk); Portugal (prt); Paraguay (pry); Qatar (qat); Russia (rus); Saudi Arabia (sau); Sudan (sdn); Senegal (sen); Singapore (sgp); El Salvador (slv); Serbia (srb); Slovakia (svk); Slovenia (svn); Sweden (swe); Syria (syr); Thailand (tha); Turkmenistan (tkm); Trinidad and Tobago (tto); Tunisia (tun); Turkey (tur); Tanzania (tza); Ukraine (ukr); Uruguay (ury); United States (usa); Uzbekistan (uzb); Venezuela (ven); Vietnam (vnm); Yemen (yem); South Africa (zaf); Zambia (zmb); Zimbabwe (zwe).

The countries present in the HS92 dataset include: Angola (ago); Albania (alb); United Arab Emirates (are); Argentina (arg); Armenia (arm); Australia (aus); Austria (aut); Azerbaijan (aze); Benin (ben); Burkina Faso (bfa); Bangladesh (bgd); Bulgaria (bgr); Bahrain (bhr); Bosnia and Herzegovina (bih); Belarus (blr); Bolivia (bol); Brazil (bra); Canada (can); Switzerland (che); Chile (chl); China (chn); Cote d'Ivoire (civ); Cameroon (cmr); Republic of the Congo (cog); Colombia (col); Costa Rica (cri); Cuba (cub); Czech Republic (cze); Germany (deu); Denmark (dnk); Dominican Republic (dom); Algeria (dza); Ecuador (ecu); Egypt (egy); Spain (esp); Estonia (est); Ethiopia (eth); Finland (fin); France (fra); Gabon (gab); United Kingdom (gbr); Georgia (geo); Ghana (gha); Guinea (gin); Greece (grc); Guatemala (gtm); Hong Kong (hkg); Honduras (hnd); Croatia (hrv); Hungary (hun); Indonesia (idn); India (ind); Ireland (irl); Iran (irn); Israel (isr); Italy (ita); Jamaica (jam); Jordan (jor); Japan (jpn); Kazakhstan (kaz); Kenya (ken); Kyrgyzstan (kgz); Cambodia (khm); South Korea (kor); Kuwait (kwt); Laos (lao); Lebanon (lbn); Libya (lby); Sri Lanka (lka); Lithuania (ltu); Latvia (lva); Morocco (mar); Moldova (mda); Madagascar (mdg); Mexico (mex); Macedonia (mkd); Mali (mli); Burma (mmr); Mongolia (mng); Mozambique (moz); Mauritania (mrt); Mauritius (mus); Malawi (mwi); Malaysia (mys); Nigeria (nga); Nicaragua (nic); Netherlands (nld); Norway (nor); New Zealand (nzl); Oman (omn); Pakistan (pak); Panama (pan); Peru (per); Philippines (phl); Papua New Guinea (png); Poland (pol); North Korea (prk); Portugal (prt); Paraguay (pry); Qatar (qat); Russia (rus); Saudi Arabia (sau); Sudan (sdn); Senegal (sen); Singapore (sgp); El Salvador (slv); Serbia (srb); Slovakia (svk); Slovenia (svn); Sweden (swe); Syria (syr); Togo (tgo); Thailand (tha); Tajikistan (tjk); Turkmenistan (tkm); Trinidad and Tobago (tto); Tunisia (tun); Turkey (tur); Tanzania (tza); Uganda (uga); Ukraine (ukr); Uruguay (ury); United States (usa); Uzbekistan (uzb); Venezuela (ven); Vietnam (vnm); Yemen (yem); South Africa (zaf); Zambia (zmb); Zimbabwe (zwe)

Appendix D – Summary Tables

In this appendix we present two tables (D1 and D2) that show a yearly summary of the average of each variable of interest used along with the SITC and HS92 datasets. Each row shows: the number of countries (N_c); number of products (N_p); average size of the option set (O); average size of the product basket (P); number of observed newly developed products (N_j); average relative relatedness of newly developed products ($\Omega_{c,y\rightarrow y+2}$); average the relative complexity of newly developed products ($\Pi_{c,y\rightarrow y+2}$); average ECI of countries that developed at least one new product in year y; average correlation between the relatedness and complexity of the option set (ρ_{cy}); the average Log GDP *per capita*, average Log of Population size, and average Human Capital Index of countries that developed at least one new product in year y.

Table D1 – Descriptive summary of features used to analyze the SITC dataset between 1984 to 2011.

Year	N_c	N_p	O	P	N_j	$\Omega_{c,y\rightarrow y+2}$	$\Pi_{c,y\rightarrow y+2}$	ECI	ρ	Log GDP_{pc}	Log Pop	HC
1984	102	747	646,7	100,3	93	1	-0,57	0,04	-0,6	3,78	7,07	2,01
1985	101	750	649,1	100,9	94	0,96	-0,55	0,03	-0,61	3,78	7,08	2,04
1986	102	754	652,7	101,3	94	0,96	-0,54	0,02	-0,62	3,77	7,09	2,05
1987	102	756	653,4	102,6	89	0,92	-0,5	0,05	-0,62	3,77	7,09	2,07
1988	102	759	653,9	105,1	88	0,87	-0,35	0,11	-0,6	3,77	7,1	2,09
1989	102	759	651,8	107,2	91	0,88	-0,37	0,11	-0,6	3,77	7,11	2,1
1990	102	759	650,7	108,3	95	0,9	-0,37	0,1	-0,6	3,78	7,12	2,13
1991	102	759	649,1	109,9	92	0,88	-0,36	0,12	-0,6	3,78	7,13	2,15
1992	118	759	648,6	110,4	94	0,8	-0,23	0,06	-0,61	3,79	7,1	2,23
1993	121	760	643,2	116,8	97	0,79	-0,2	0,01	-0,58	3,8	7,09	2,27
1994	121	760	640,6	119,4	97	0,81	-0,25	0,02	-0,58	3,8	7,1	2,29
1995	121	760	639,8	120,2	97	0,78	-0,23	0,02	-0,58	3,81	7,1	2,31
1996	121	760	637,5	122,5	112	0,74	-0,19	0,03	-0,56	3,83	7,11	2,34
1997	121	760	636	124	114	0,73	-0,16	0,02	-0,56	3,84	7,12	2,36
1998	121	760	637,3	122,7	112	0,76	-0,2	-0,01	-0,55	3,84	7,12	2,38
1999	121	760	638,2	121,8	107	0,67	-0,15	0,06	-0,55	3,85	7,13	2,4
2000	123	761	639,5	121,5	112	0,74	-0,17	0,05	-0,56	3,87	7,12	2,42
2001	123	762	638,1	123,9	111	0,76	-0,21	0,08	-0,57	3,88	7,12	2,44
2002	123	762	637,9	124,1	108	0,68	-0,16	0,1	-0,57	3,89	7,13	2,46
2003	123	761	637,2	123,8	111	0,68	-0,17	0,11	-0,56	3,91	7,14	2,48
2004	123	762	637,3	124,7	105	0,71	-0,13	0,11	-0,54	3,93	7,14	2,5
2005	123	762	636,8	125,2	103	0,7	-0,11	0,14	-0,54	3,98	7,15	2,52

2006	124	763	636,1	126,9	108	0,72	-0,14	0,12	-0,52	4	7,15	2,55
2007	124	763	635,5	127,5	96	0,68	-0,11	0,23	-0,51	4,03	7,16	2,56
2008	124	763	635,1	127,9	88	0,63	0	0,3	-0,51	4,06	7,17	2,58
2009	124	763	634,7	128,3	103	0,6	-0,09	0,16	-0,51	4,04	7,17	2,6
2010	124	763	634,6	128,4	102	0,54	-0,07	0,19	-0,49	4,08	7,18	2,62
2011	124	763	633,3	129,7	98	0,59	-0,07	0,18	-0,49	4,11	7,18	2,64

Table D2 – Descriptive summary of features used to analyze the HS92 dataset between 2001 to 2011.

Year	N_c	N_p	O	P	N_j	$\Omega_{c,y\rightarrow y+2}$	$\Pi_{c,y\rightarrow y+2}$	ECI	ρ	Log GDP_{pc}	Log Pop	HC
2001	127	1205	1018,7	186,3	120	0,74	-0,2	0,02	-0,56	3,84	7,12	2,4
2002	127	1206	1018,8	187,2	119	0,77	-0,22	0,03	-0,55	3,85	7,13	2,42
2003	127	1206	1020,5	185,5	117	0,73	-0,17	0,08	-0,55	3,86	7,14	2,44
2004	127	1206	1020,1	185,9	120	0,69	-0,1	0,05	-0,54	3,89	7,14	2,45
2005	128	1208	1020,3	187,7	122	0,71	-0,12	0,02	-0,52	3,93	7,15	2,48
2006	128	1208	1019,6	188,4	122	0,68	-0,13	0,02	-0,52	3,96	7,15	2,5
2007	128	1208	1017,8	190,2	117	0,68	-0,11	0,05	-0,51	3,98	7,16	2,52
2008	128	1207	1015,1	191,9	117	0,67	-0,14	0,08	-0,5	4,01	7,17	2,53
2009	128	1207	1015,1	191,9	117	0,68	-0,12	0,09	-0,5	3,99	7,17	2,55
2010	128	1207	1015,6	191,4	118	0,66	-0,06	0,07	-0,49	4,03	7,18	2,57
2011	128	1207	1017,4	189,6	118	0,69	-0,11	0,08	-0,47	4,06	7,19	2,59

References

Abramovitz, M. (1986). Catching up, forging ahead, and falling behind. *The Journal of Economic History, 46*(2), 385-406.

Alshamsi, A., Pinheiro, F. L., & Hidalgo, C. A. (2018). Optimal diversification strategies in the networks of related products and of related research areas. *Nature Communications, 9*(1), 1328.

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy*: University of Michigan Press.

Atkinson, A. B., & Stiglitz, J. E. (1969). A new view of technological change. *The Economic Journal, 79*(315), 573-578.

Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American Economic Review, 86*(3), 630-640.

Autant-Bernard, C. (2001). The geography of knowledge spillovers and technological proximity. *Economics of Innovation and New Technology, 10*(4), 237-254.

Bahar, D., Hausmann, R., & Hidalgo, C. A. (2014). Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics, 92*(1), 111-123.

Bahar, D., Rosenow, S., Stein, E., & Wagner, R. (2019). Export take-offs and acceleration: Unpacking cross-sector linkages in the evolution of comparative advantage. *World Development, 117*, 48-60.

Balland, P.-A., & Boschma, R. (2021). Complementary interregional linkages and Smart Specialisation: an empirical study on European regions. *Regional Studies*, 1-12. doi:10.1080/00343404.2020.1861240

Balland, P.-A., Boschma, R., Crespo, J., & Rigby, D. L. (2019). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies, 53*(9), 1252-1268.

Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional Studies, 39*(1), 61-74.

Boschma, R. (2017). Relatedness as driver of regional diversification: A research agenda. *Regional Studies, 51*(3), 351-364.

Boschma, R., Balland, P.-A., & Kogler, D. F. (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.

Boschma, R., & Capone, G. (2015). Institutions and diversification: Related versus unrelated diversification in a varieties of capitalism framework. *Research Policy, 44*(10), 1902-1914.

Boschma, R., & Capone, G. (2016). Relatedness and diversification in the European Union (EU-27) and European Neighbourhood Policy countries. *Environment and Planning C: Government and Policy, 34*(4), 617-637.

Boschma, R., & Lambooy, J. (1999). Evolutionary economics and economic geography. *Journal of Evolutionary Economics, 9*(4), 411-429.

Boschma, R., Minondo, A., & Navarro, M. (2013). The Emergence of New Industries at the Regional Level in S pain: A Proximity Approach Based on Product Relatedness. *Economic Geography, 89*(1), 29-51.

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research Policy, 32*(1), 69-87.

Brezis, E. S., Krugman, P. R., & Tsiddon, D. (1993). Leapfrogging in international competition: A theory of cycles in national technological leadership. *The American Economic Review, 83*(5), 1211-1219.

Caviggioli, F., Jensen, P., & Scellato, G. (2020). Highly skilled migrants and technological diversification in the US and Europe. *Technological Forecasting and Social Change, 154*, 119951. doi:https://doi.org/10.1016/j.techfore.2020.119951

Chatterjee, S., & Wernerfelt, B. (1991). The link between resources and type of diversification: Theory and evidence. *Strategic Management Journal, 12*(1), 33-48.

Cicerone, G., McCann, P., & Venhorst, V. A. (2020). Promoting regional growth and innovation: relatedness, revealed comparative advantage and the product space. *Journal of Economic Geography, 20*(1), 293-316.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 128-152.

Coniglio, N. D., Lagravinese, R., Vurchio, D., & Armenise, M. (2018). The pattern of structural change: testing the product space framework. *Industrial and Corporate Change, 27*(4), 763-785.

Cortinovis, N., Crescenzi, R., & Van Oort, F. (2020). Multinational enterprises, industrial relatedness and employment in European regions. *Journal of Economic Geography, 20*(5), 1165-1205.

Crescenzi, R., Gagliardi, L., & Iammarino, S. (2015). Foreign multinationals and domestic innovation: Intra-industry effects and firm heterogeneity. *Research Policy, 44*(3), 596-609.

Delgado, M., Porter, M. E., & Stern, S. (2015). Defining clusters of related industries. *Journal of Economic Geography, 16*(1), 1-38.

Diodato, D., Neffke, F., & O'Clery, N. (2018). Why do industries coagglomerate? How Marshallian externalities differ by industry and have evolved over time. *Journal of Urban Economics, 106*, 1-26.

Dosi, G. (1984). Technical Change in the International Environment: The Dynamics of Trade and Investment. In *Technical Change and Industrial Transformation* (pp. 218-283): Springer.

Eichengreen, B., Park, D., & Shin, K. (2013). *Growth slowdowns redux: New evidence on the middle-income trap*. Working Paper Series, (18673). National Bureau of Economic Research, http://www.nber.org/papers/w18673.

Elekes, Z., Boschma, R., & Lengyel, B. (2019). Foreign-owned firms as agents of structural change in regions: the case of Hungary 2000-2009. *Regional Studies, 53*(11), 1603-1613.

Fagerberg, J., & Srholec, M. (2008). National innovation systems, capabilities and economic development. *Research Policy, 37*(9), 1417-1435.

Farinha, T., Balland, P.-A., Morrison, A., & Boschma, R. (2019). What drives the geography of jobs in the US? Unpacking relatedness. *Industry and Innovation, 26*(9), 988-1022.

Fassio, C., Montobbio, F., & Venturini, A. (2019). Skilled migration and innovation in European industries. *Research Policy, 48*(3), 706-718.

Feenstra, R. C., Inklaar, R., & Timmer, M. P. (2015). The next generation of the Penn World Table. *The American Economic Review, 105*(10), 3150-3182.

Feenstra, R. C., Lipsey, R. E., Deng, H., Ma, A. C., & Mo, H. (2005). *World trade flows: 1962-2000*. Working Paper Series, (11040). National Bureau of Economic Research, http://www.nber.org/papers/w11040.

Felipe, J., Abdon, A., & Kumar, U. (2012). Tracking the middle-income trap: What is it, who is in it, and why?

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research Policy, 30*(7), 1019-1039.

Freeman, C. (1982). *The economics of industrial innovation*. United States.

Frenken, K., & Boschma, R. (2007). A theoretical framework for evolutionary economic geography: industrial dynamics and urban growth as a branching process. *Journal of Economic Geography, 7*(5), 635-649.

Gao, J., Jun, B., Pentland, A., Zhou, T., & Hidalgo, C. A. (2021). Spillovers across industries and regions in China's regional economic diversification. *Regional Studies*. doi:https://doi.org/10.1080/00343404.2021.1883191

Gaulier, G., & Zignago, S. (2010). Baci: international trade database at the product-level (the 1994-2007 version).

Gill, I. S., & Kharas, H. (2015). *The middle-income trap turns ten*: The World Bank.

Guevara, M. R., Hartmann, D., Aristarán, M., Mendoza, M., & Hidalgo, C. A. (2016). The research space: using career paths to predict the evolution of the research output of individuals, institutions, and nations. *Scientometrics, 109*(3), 1695--1709.

Hartmann, D., Bezerra, M., Lodolo, B., & Pinheiro, F. L. (2020). International trade, development traps, and the core-periphery structure of income inequality. *EconomiA, 21*(2), 255-278.

Hartmann, D., Guevara, M. R., Jara-Figueroa, C., Aristarán, M., & Hidalgo, C. A. (2017). Linking economic complexity, institutions, and income inequality. *World Development, 93*, 75-93.

Hartmann, D., Zagato, L., Gala, P., & Pinheiro, F. L. (2021). Why did some countries catch-up, while others got stuck in the middle? Stages of productive sophistication and smart industrial policies. *Structural Change and Economic Dynamics, 58*, 1-13.

Hausmann, R., & Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth, 16*(4), 309-342.

Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2014). *The atlas of economic complexity: Mapping paths to prosperity*: Mit Press.

Hidalgo, C. A. (2021). Economic complexity theory and applications. *Nature Reviews Physics*, 1-22.

Hidalgo, C. A., Balland, P.-A., Boschma, R., Delgado, M., Feldman, M., Frenken, K., . . . Zhu, S. (2018, 2018//). *The Principle of Relatedness.* Paper presented at the Unifying Themes in Complex Systems IX, Cham.

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences, 106*(26), 10570-10575.

Hidalgo, C. A., Klinger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science, 317*(5837), 482-487.

Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics, 108*(3), 577-598.

Jara-Figueroa, C., Jun, B., Glaeser, E. L., & Hidalgo, C. A. (2018). The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms. *Proceedings of the National Academy of Sciences, 115*(50), 12646-12653.

Kharas, H., & Kohli, H. (2011). What is the middle income trap, why do countries fall into it, and how can it be avoided? *Global Journal of Emerging Market Economies, 3*(3), 281-289.

Kogler, D. F., Rigby, D. L., & Tucker, I. (2013). Mapping knowledge space and technological relatedness in US cities. *European Planning Studies, 21*(9), 1374-1391.

Kogut, B., & Zander, U. (1993). Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies, 24*(4), 625-645.

Lall, S. (1992). Technological capabilities and industrialization. *World Development, 20*(2), 165-186.

Lee, K. (2013). *Schumpeterian analysis of economic catch-up: Knowledge, path-creation, and the middle-income trap*: Cambridge University Press.

Lee, K., & Lim, C. (2001). Technological regimes, catching-up and leapfrogging: findings from the Korean industries. *Research Policy, 30*(3), 459-483.

Montresor, S., & Quatraro, F. (2017). Regional branching and Key enabling technologies: Evidence from European patent data. *Economic Geography, 93*(4), 367-396.

Moreno, R., Paci, R., & Usai, S. (2005). Spatial spillovers and innovation activity in European regions. *Environment and planning A, 37*(10), 1793-1812.

Neffke, F., Hartog, M., Boschma, R., & Henning, M. (2018). Agents of structural change: the role of firms and entrepreneurs in regional diversification. *Economic Geography, 94*(1), 23-48.

Neffke, F., & Henning, M. (2013). Skill relatedness and firm diversification. *Strategic Management Journal, 34*(3), 297-316.

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

Pasinetti, L. L. (1981). Structural change and economic dynamics. *A Theoretical Essay on the Dynamics of the Wealth of Nations, Cambridge*.

Perez, C., & Soete, L. (1988). Catching up in technology: entry barriers and windows. *Technical Change and Economic Theory. CF Giovanni Dosi, Richard Nelson, Gerald Silverberg, Luc Soete. London, Pinter*, 458-479.

Petralia, S., Balland, P.-A., & Morrison, A. (2017). Climbing the ladder of technological development. *Research Policy, 46*(5), 956-969.

Rigby, D. L. (2015). Technological relatedness and knowledge space: entry and exit of US cities from patent classes. *Regional Studies, 49*(11), 1922-1937.

Rigby, D. L., & Essletzbichler, J. (1997). Evolution, process variety, and regional trajectories of technological change in US manufacturing. *Economic Geography, 73*(3), 269-284.

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy, 98*(5, Part 2), S71-S102.

Romer, P. M. (1994). The origins of endogenous growth. *Journal of Economic Perspectives, 8*(1), 3-22.

Saviotti, P. P., & Frenken, K. (2008). Export variety and the economic performance of countries. *Journal of Evolutionary Economics, 18*(2), 201-218.

Saviotti, P. P., & Pyka, A. (2004). Economic development by the creation of new sectors. *Journal of Evolutionary Economics, 14*(1), 1-35.

Schumpeter, J. (1942). Creative destruction. *Capitalism, Socialism and Democracy, 825*, 82-85.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization, 1*(1), 161-176.

Weitzman, M. L. (1998). Recombinant growth. *The Quarterly Journal of Economics, 113*(2), 331-360.

Xiao, J., Boschma, R., & Andersson, M. (2018). Industrial diversification in Europe: The differentiated role of relatedness. *Economic Geography, 94*(5), 514-549.

Yoguel, G. (2015). Carlota Perez, Luc Soete-Catching up in technology: entry barriers and windows of opportunity. *Revista Brasileira de Inovação, 14*(2), 257-288.

Zhu, S., He, C., & Zhou, Y. (2017). How to jump further and catch up? Path-breaking in an uneven industry space. *Journal of Economic Geography, 17*(3), 521-545.