Papers in Evolutionary Economic Geography

19.30

From variety to economic complexity: empirical evidence from Italian regions

Roberto Antonietti and Chiara Burlina

Utrecht University Human Geography and Planning

http://peeg.wordpress.com

From variety to economic complexity: empirical evidence from Italian regions

Roberto Antonietti* and Chiara Burlina°

*Corresponding author
"Marco Fanno" Department of Economics and Management
University of Padova
Via del Santo 33
35123 Padova, Italy
roberto.antonietti@unipd.it
ORCID ID: 0000-0002-2172-4062

°*Department of Mathematical Sciences, Mathematical Finance and Econometrics*
Catholic University of Milan
Via Necchi 9,
20123 Milan, Italy
chiara.burlina@unicatt.it
ORCID ID: 0000-0002-1914-8772

Abstract

Taking an evolutionary economic geography approach, we test whether the level of industry variety in a region affects its economic complexity. With reference to Italy, we measure variety using a Theil index of information entropy, and complexity with the Hidalgo and Hausmann index. Our results show that regions where variety grows faster also have a higher rate of growth in economic complexity. This relationship only holds in regions with low initial levels of variety and/or complexity, however, which are mainly located in the South of Italy. We suggest that product diversification, by increasing regional specialization in high-tech industries, can explain regional development and Italian North-South disparities.

Keywords: economic complexity, entropy, industry variety, unit root

JEL codes: O33; R11; R12

1. Introduction

Regional development has often been conceived as the outcome of specialization versus diversification economies. In the former, usually labelled as Marshall-Arrow-Romer (MAR) externalities, the concentration of an industry in a given region would facilitate knowledge exchange between similar firms, thus promoting knowledge specialization and greater productivity. In the latter, defined as Jacob externalities, regional economic performance would stem from local interaction between firms belonging to different industries, which drives the cross-fertilization of ideas and innovation, especially in highly urbanized areas. The literature is inconclusive as to which of the two factors prevail (for a review, see Beaudry and Schiffauerova, 2009).

Here we propose another approach that brings industry specialization and diversification together. We argue that a greater product or industry diversification pushes regions to specialize in high-quality products and industries because of the better chances of recombining different knowledge sources. This greater specialization in turn increases the level of regional economic complexity, which is ultimately linked to aggregate growth in productivity.

There is a growing consensus among scholars that economic complexity is related to wealth: countries producing a more diversified portfolio of highly sophisticated and less-ubiquitous products experience higher levels of income per capita. Economic complexity also favors regions' convergence in income per capita (Hausmann & Hidalgo, 2011; Hausmann et al., 2014). Two questions remain unanswered, however. Why is the level of economic complexity higher in some countries/regions than in others? What drives an increase or decrease in economic complexity?

In the recent literature, economic complexity has been studied in relation to: (i) a country's product diversification; (ii) growth in GDP per capita and rate of industrialization; (iii) economic development and income/wage inequality; and (iv) a region's export basket diversification (Cicerone, McCann, & Venhorst, 2019; Gao & Zhou, 2018; Pinheiro, Alshamsi, Hartmann, Boschma, & Hidalgo, 2018). In all these studies, economic complexity is taken as a predetermined, exogenous variable originating from a country's or region's unobserved set of skills and capabilities. Few papers have investigated the determinants of economic complexity. Sweet and Eterovic-Maggio (2015), for example, use economic complexity to approximate the aggregate innovative outcome of a country, and show that it depends on the stringency of the domestic intellectual propert rights (IPR) system. Javorcik, Lo Turco and Maggioni (2018) demonstrate that manufacturing firms in Turkey tend to introduce more complex products if they operate in regions and sectors with a greater propensity to supply foreign multinationals. Balland and Rigby (2017), and Balland et al. (2018), find instead in the US that complex knowledge and economic activities tend to co-agglomerate in large urban areas, and contribute to increasing spatial inequality across regions.

Identifying the determinants of economic complexity is important for two main reasons. From an academic perspective, it is worth knowing whether this variable can be considered as exogenous in explaining economic development (i.e. through the dynamics of GDP per capita). On the other hand, it might be better to treat it as endogenous in growth regressions using appropriate econometric techniques. This is crucial to preventing the complexity-growth relationship from being spurious. From a policy perspective, if regions make more sophisticated products as a way to improve the level of their technological and economic development, then it becomes important to identify the leverages of economic complexity and target industrial policies accordingly.

Despite a few efforts to clarify the picture, theoretical and empirical analyses on what generates and stimulates a region's or country's economic complexity remain scanty. This paper investigates the determinants of economic complexity, arguing that it is a structural characteristic of a region that depends on its level of product/industry diversification. Our focus is on Italian NUTS3 regions. Between 2008 and 2012, Italy ranked among the top twenty countries in the world by level of economic complexity[1]. Italian regions vary considerably in skill endowment, performance, and level of development. They are also characterized by a marked regional heterogeneity in terms of specializations, revealed comparative advantages, and density of goods in the product space (Cicerone, McCann and Venhorst, 2019), making it interesting to investigate the relationship between entropy and economic complexity at this level.

We measure industry variety through an information entropy index (Frenken, Van Oort, & Verburg, 2007), which captures the higher probability of a five-digit industry different from an initial distribution being randomly "picked" in a given region. In this sense, entropy represents the state of our "lack of knowledge" and is higher the more an economic system is diversified. This general information entropy measure can be split into a within-sector entropy (or related variety) and a between-sector entropy (or unrelated variety) component: the former refers to the weighted sum of the entropy at five-digit industry level within each two-digit sector, and captures the intensive margin of industry diversification; the latter refers to the entropy of the two-digit industry distribution, capturing the extensive margin of diversification.

We capture economic complexity using the Hidalgo and Hausmann (2009) index. For the period 2008-2016, we use NUTS3 regional export data at three-digit industry classification level provided by the Italian Statistics Institute (ISTAT) to build revealed comparative advantages and establish the proximity matrix among products (sectors). We define a region's economic complexity index (ECI) as the average knowledge intensity of the products (sectors) exported by the region, which underlies the state of sophistication of the knowledge the region has accumulated.

[1] https://atlas.media.mit.edu/en/rankings/country/eci/?year_range=2008-2012

Our empirical strategy develops in two steps. We first run a unit root test to check whether our variety and complexity indicators are stationary processes. Using this information, we then run a regression analysis where we add industry variety, a series of controls on regional human capital, population density, and number of inward greenfield projects involving foreign direct investments (FDI) to our focal regressor. To account for potential simultaneity, we adopt Lewbel's (2012) instrumental variable approach, which uses the conditional second moments of the endogenous variables and the heteroskedasticity of the first-stage regression residuals to generate the instruments. We also check whether the entropy-complexity relation depends on a region's initial level of entropy and economic complexity.

Our results confirm that regions where variety increased faster also had the largest growth in economic complexity. This result is robust to the inclusion of additional regional characteristics, fixed effects and simultaneity. This effect is not homogeneous across regions, however; it is only significant where the initial levels of variety and complexity are lower.

The rest of the paper is organized as follows: Section 2 presents the recent empirical literature on economic complexity; Section 3 describes the theoretical background and discusses the approaches used to explain the link between variety and economic complexity; Section 4 outlines our methodological approach, describing the data and the econometric analysis; in Section 5 we report and discuss our results; and Section 6 concludes.

2. Related literature

2.1 Economic complexity as a driver of economic development

Several attempts have been made over time to explain Italy's regional development and economic disparities, from the seminal papers of the 1960s (Eckaus, 1961; Schachter, 1967) to very recent contributions (Filippetti, Guy, & Iammarino, 2019; Missiaia, 2019). North-South growth differentials might be explained by: social capital and income disparities (Calcagnini &

Perugini, 2019; Helliwell & Putnam, 1995); sectoral specialization (Basile & Ciccarelli, 2018; Boschma & Iammarino, 2009; Guerrieri & Iammarino, 2007); growth in employment rates (Basile, Girardi, Mantuano, & Russo, 2018; Cellini & Torrisi, 2014; Cracolici, Cuffaro, & Nijkamp, 2007; Lagravinese, 2015); and productivity gaps (Accetturo, Di Giacinto, Micucci, & Pagnini, 2018; Rungi & Biancalani, 2019).

Despite the abundance of literature on the topic, to the best of our knowledge, no studies conducted so far have tried to explain economic differentials in terms of economic complexity. In this paper we aim to disentangle this phenomenon by linking regions' economic complexity with the level of sophistication and variety of their products (Frenken et al., 2007; Inoua, 2016; van Dam & Frenken, 2019).

Starting from the seminal contributions from Hidalgo et al. (2007), and Hidalgo and Hausmann (2009), the concept of economic complexity has been increasingly associated with that of economic development. In their framework, complexity arises as the outcome of two characteristics: the diversity of a country's product/export portfolio, which is greater the larger the number of different products it exports; and the ubiquity of a product, which is higher the lower the number of countries producing or exporting it. The underlying mechanism is that countries differ in level of economic complexity (and development) because they are endowed with different sets of skills and capabilities.

Relying on this framework, many scholars in the last few years have examined the role of economic complexity in explaining aggregate economic outcomes like growth in GDP per capita or income inequality. Across countries, Hidalgo and Hausmann (2009) identify a positive correlation between their measures of economic complexity and the rate of growth in GDP per capita. Their findings are confirmed by Felipe et al. (2012), who consider more than 5000 products and 124 countries, and by Ferrarini and Scaramozzino (2016), Pugliese et al. (2017), and Sbardella et al. (2018), who use the 'fitness' index as an alternative measure of economic complexity. In a sample of over 150 countries considered between 1963 and 2008, Hartmann et al. (2017) finds more economic complexity associated with less income inequality. Sbardella et al. (2017) show that the relationship between economic complexity and wage inequality has an inverted U-shape: it increases in the earliest stages of development and decreases in the latest, when the level of economic complexity is higher. Gao and Zhou (2018) also document a negative correlation between economic complexity and income inequality for Chinese regions in 2000-15.

Other recent studies look at the possible influence of economic complexity on countries' ability to diversify their product portfolio or develop new specializations in unrelated industries. Pinheiro et al. (2018) find that countries with intermediate levels of development and economic complexity are the most likely to engage in unrelated diversification.

All the above-mentioned studies use economic complexity as an exogenous predictor, however. They see it as a path-dependent process where the development of new products or industries is the outcome of a process that recombines existing skills and capabilities (Fontana, 2010; Hidalgo et al., 2007). To our knowledge, very few studies have considered economic complexity as a dependent variable, and explicitly investigated why some regions or countries have a higher tendency to increase their level of knowledge complexity than others. Sweet and Eferovic-Maggio (2015) look at whether a stronger IPR system triggers innovation, using the economic complexity index to capture the innovative output of a sample of 94 countries from 1965 to 2005. Their results show that more stringent IPR laws make a country better able to raise the level of sophistication of its products, but only if the country already has high levels of development, human capital and economic complexity.

In a study of Turkish manufacturing firms between 2006 and 2009, Javorcik et al. (2018) show that firms' ability to upgrade the quality (and complexity) of their products depends on the amount of downstream inward FDI in their region. Multinational enterprises therefore act as agents of structural change (Neffke, Hartog, Boschma, & Henning, 2018) by improving firms' average level of product sophistication.

Another possible driver of economic complexity is spatial agglomeration, specifically in the form of urban economies and/or scale effects in large cities. In the US, Balland and Rigby (2017), and Balland et al. (2018) demonstrate that complex knowledge-intensive activities such as patenting tend to be concentrated in large metropolitan areas. This is because complex activities and processes require a deeper division of labor (i.e. greater specialization), and can be developed more efficiently where knowledge is more coordinated and knowledge spillovers are easier, as in large cities (Antonietti, Cainelli, & Lupi, 2013).

In this paper, we propose an alternative explanation, positing that a region's increasing economic complexity is the outcome of an ever-higher level of entropy. To elucidate our approach, we borrow elements from condensed-matter physics and evolutionary economics. From the former, we take the argument that the patterns of complex (trade) relationships emerging between firms resemble the emergence of complex macroscopic arrangements of particles. From the latter, we draw the idea that the evolution of complexity depends on the entry and exit of products in a region's export basket of, which in turn depends on the degree of diversification of an economy.

2.2 The link between variety and complexity

Since the beginning of the 21st century, scholars have realized the need to combine economics with complexity science to overcome certain possible limitations of neoclassical economic theory. Economies can be seen as complex systems inasmuch as they have two main features: they comprise a number of heterogeneous agents organized into a great variety of groups; and they exhibit nonlinear patterns, because agents act at different times and in different places (Blume & Durlauf, 2006; Fontana, 2010).

Evolutionary economics and combinatorial calculus try to explain the relationship between variety and complexity from another point of view. Inoua (2016), and Van Dam and Frenken (2019), among others, make the point that countries and regions develop by accumulating skills and capabilities that complement the available natural resources. The ability of a country or region to create products depends on the capabilities available, and on the chances to combine them. Inoua (2016) shows that, in the long run, an economy's growth rate is proportional to the log diversification of its knowhow, which is directly proportional to the log diversification of its products. Judging from COMTRADE export data concerning 160 countries around the world, the ECI (which reflects the average product sophistication in a given country) correlates strongly with the economy's degree of diversification, as measured using a technology development index.

Van Dam and Frenken (2019) take a similar stance, assuming that a product is the outcome of a string of capabilities, the number of which defines the product's length. If a country/region is endowed with n capabilities, then it can produce $\binom{n}{s}$ different combinations of length s and a total amount of products $d(n)$ equating to 2^n, for an average product length $\bar{s}(n)$of $n/2$. The term d represents the variety of products that an economy achieves with its n capabilities, and it follows that both $\ln(d)$ and the average product length $\bar{s}$are proportional to n. Combining these two elements, we have that $\ln(d)$ is proportional to $\bar{s}(n)$; in other words, product variety is proportional to the average product length.

The entry and exit of products explain the dynamics of a country's economic complexity and its link with entropy. As a country develops, it produces new and more sophisticated products, income increases, and so do wages. At the same time, some products (usually the most basic) exit from the country's product portfolio because they become too expensive to produce. Van Dam and Frenken (2019) demonstrate that, when a country abandons its less-sophisticated goods and services, its average economic complexity increases.

This fact allows us to reproduce the dynamics of diversification, and the hump-shaped relationship between variety and economic development. In the initial phase of development, the rate of product variety grows exponentially, faster than the accumulation of capabilities. This means that a country is generating new and ever more sophisticated products, without losing any of those it previously produced. In this phase, both variety and complexity increase, the latter increasing linearly with the capabilities. In the transition phase, countries abandon their more basic products, and their economies continue to diversify, but at a slower rate. Their economic complexity continues to rise, however, because the exiting of products increases the average product length. In the last phase, the number of products exiting increases to such a degree that overall variety declines, while economic complexity peaks.

Combining all these elements, we derive the following hypotheses: countries and regions develop by accumulating new capabilities; these capabilities are recombined to generate new products; as long as the number of different capabilities increases, the chances of recombining different pieces of know-how increase as well, and so does the average product length, leading to a general increase in economic complexity. As countries and regions reach a certain level of development and diversification, their product exits exceed their product entries, variety declines, and complexity reaches its upper limit.

So, our first hypothesis is that the level of economic complexity rises in regions where industry variety (or total entropy) increases. The effect of variety on complexity can vary, however, depending on the baseline state of the system. In general, we can expect the relation to be stronger in regions with initially low levels of variety and complexity, and negligible in highly developed regions with high levels of diversification and complexity. In regions where diversification and complexity are low, an increase in the former has a good chance of making their production more sophisticated. On the other hand, in regions where variety and complexity are high, there is little or no room for any additional increase in total entropy to further stimulate complexity.

Hence our second hypothesis that the impact of an increase in variety on economic complexity is greater in regions with a low initial level of variety and/or complexity.

4. Empirical analysis

4.1 Data and variables

We test our hypotheses using two main data sources regarding 103 Italian provinces (NUTS 3 regions) for the years from 2008 to 2016. To compute our industry variety index, we use data from the Archives of Active Italian Firms (ASIA) administered by the ISTAT. This dataset provides information on the numbers of plants and employees in each NUTS 3 region and five-digit industry, using the NACE classification (Revision 2). The data used to compute the ECI come instead from the Coeweb archives on foreign trade statistics (ISTAT) concerning annual exports from Italian NUTS 3 regions and three-digit industries (NACE Rev. 2). We also use the ISTAT-ASTI database (Atlas of Territorial Infrastructure Statistics) to collect information on other regional characteristics.

Following Frenken et al. (2007), our industry variety (VAR) index is given by a Theil information entropy index, computed at five-digit industry level. The total measure of industry variety is obtained from the sum of two components, i.e. related (or within-sector entropy) and unrelated (or between-sector entropy) variety: VAR=RV+UV. We compute related variety as the weighted sum of the entropy at five-digit industry level (p_i), within each two-digit industry class (P_g), while unrelated variety captures entropy at the two-digit industry level:

[1] $RV = \sum_{g=1}^{G} P_g H_g, \ where\ H_g = \sum_{i \in S_g} \frac{p_i}{P_g} log_2 \left(\frac{1}{\frac{p_i}{P_g}} \right)$

[2] $UV = \sum_{g=1}^{G} P_g log_2 \left(\frac{1}{P_g} \right)$.

Following Cadot, Carrère and Strauss-Kahn (2011), this distinction between UV and RV corresponds to a diversification pattern based on the extensive and intensive margin, respectively. The former occurs when the number of new industries rises, the latter when the distribution of employment across existing industries becomes more balanced, implying a convergence in the employment shares across sectors in a region.

The ECI is computed using the eigenvector method, as in Balland (2017). We use a revisited version of his knowledge complexity index, replacing patent data with export data at province level, like Hidalgo and Hausmann (2009). As done by Reynolds et al. (2018), we use export data as an indicator of international competitiveness. Our ECI is therefore based on three-digit industries in which each province has a revealed comparative advantage in terms of export activity. We compute the revealed comparative advantage (RCA) as follows:

[3] $RCA_{pi} = \frac{X_{pi}}{\sum_p X_{pi}} / \frac{\sum_i X_{pi}}{\sum_{pi} X_{pi}}$

where X_{pi} represents the value of exports from province p in (three-digit) industry i; if the index is higher than 1 (RCA>1), the province has a revealed comparative advantage in the industry concerned. We derive ubiquity and diversity measures from the RCA index: the former corresponds to the number of provinces with an RCA in an industry; the latter to the number of industries in which a province has an RCA. Putting the two measures together in a proximity matrix between industries and provinces, we obtain the ECI as follows:

[4] $ECI_p = \frac{K_p - \langle K \rangle}{std(K)}$,

where K_p represents the eigenvector associated with the second largest eigenvalue of the proximity matrix, obtained using the method of reflections, while $\langle K \rangle$ is its average. Our ECI can thus be defined as the average knowledge intensity of the products (sectors) exported by a province, which underlies the state of sophistication and diversification of the province's accumulated knowledge.

It is worth noting that the two indicators capture two distinct features of a regional economic system. The entropy index is a measure of the general diversification of an economy, and it increases with the number of five-digit industries, or with a more even distribution of employment across five-digit industries. The higher the index, the more the region's industry mix is diversified. The ECI does not capture diversification. Although it is computed starting from the "diversity" index, Kemp-Benedict (2014), and Mealy, Farmer and Teytelboym (2019) demonstrate that the Hausmann and Hidalgo index and the initial diversity index are orthogonal: the dot product between the knowledge diversity index $k_{c,0}$ and the ECI is zero, meaning the two measures are independent. In other words, the ECI captures a different type of information from diversity. In particular, it is closely related to a country's specialization in high- or low-quality products: countries with a high ECI specialize mostly in technologically advanced, quality products, while those with a low ECI produce more basic, lower-quality goods. The ECI is therefore more a measure of (trade) specialization (in high- or low-quality goods) than of diversification, and it is not an algebraic transformation of the variety index.

To support the concept that the two indicators capture different phenomena, we first correlate them on a yearly basis: Table 1 shows that the pairwise correlations between the two indicators are modest[2].

[INSERT TABLE 1 ABOUT HERE]

Second, we observe the positioning of the Italian provinces in the entropy-complexity space: if the two measures captured the same phenomenon, we would expect all the regions to be characterized by high (or low) levels of entropy and complexity (i.e. they would be in the first

[2] As Kemp-Benedict (2014) stressed, the fact that the diversity index and the ECI are orthogonal does not prevent their correlation from being different from zero.

or third quadrants). Figure 1 shows that several regions (one in three in our sample) occupy the second and fourth quadrants, where entropy is low (i.e. below the mean) and complexity is high (i.e. above the mean), or entropy is high and complexity is low, respectively.

[INSERT FIGURE 1 ABOUT HERE]

Finally, the ECI and industry variety variables follow two different trends: while the latter has a stochastic trend with evidence of a unit root, economic complexity is stationary (as we shall see later, in Section 4.2 and Table 3). If both indicators measured the same diversification pattern, we would expect their data generating process (or trend) to be similar, which is not the case.

Figure 2 shows the geographical distributions of the ECI (in levels) in 2008 and 2016, and of the average yearly ECI growth rate between 2008 and 2016, where the colored regions in the third map are those where the average ECI growth rate is positive. It is noteworthy that, while the highest levels of economic complexity are in the central and northern regions, most of the regions where the ECI increased are in the South of Italy.

[INSERT FIGURE 2 ABOUT HERE]

Figure 3 shows the regions where the ECI increased or decreased between 2008 and 2016. Many of them show a rising economic complexity, especially among those lying below the red line. Only a few regions show a decline in their ECI. In 2016, the highest ECIs can be found in central and northern regions like Prato (PO), Biella (BI) and Bologna (BO), and the lowest in southern regions like Siracusa (SR), Ragusa (RG) and Foggia (FG).

[INSERT FIGURE 3 ABOUT HERE]

We also consider other regional features that might confound the relationship between regions' entropy and their economic complexity. First, we control for population density (DEN), computed as the resident population per square kilometer, which captures the role that densely populated urban areas can have in favoring the spatial concentration of complex activities (Balland et al., 2018). Second, we control for the regions' stock of human capital (HK), computed as the proportion of the total resident population in the region with a university degree (bachelors and masters). Third, we add the flows of inward greenfield FDI projects (IGFDI), as a proxy for the role of foreign-owned multinationals in improving the average complexity of resident firms' products (Javorcik, Lo Turco and Maggioni, 2017). Information on FDI is drawn from fDi Markets, a database administered by the Financial Times Ltd., which provides up-to-date details of greenfield FDI projects across the world. For our purposes, we cumulatively obtain information on yearly inward FDI projects in Italian NUTS 3 regions over the years 2008-16 to establish the corresponding stock[3].

Finally, we control for regions' intensity of R&D, measured as their total expenditure on R&D per unit of GDP, at current prices. Unfortunately, this information is not available for NUTS 3 regions, but only at NUTS 2 regional level.

Before proceeding with our econometric analysis, we transform all the variables into natural logarithms, with the exception of the ECI and HK, as we have a series of negative values for the former, and some values missing for some regions and years for the latter. As regards the ECI, we adopt the following transformation: lnECI=ln(ECI+1), which allows us to reproduce a kernel distribution coming very close to that of the ECI, as shown in Figure 4.

[3] We also consider the yearly flows of inward FDI instead of the stock, but the results remain the same.

[INSERT FIGURE 4 ABOUT HERE]

As for HK, we first replace the few missing values with 0, then we operate the following transformation: lnHK=ln(HK+1). A description of the variables and summary statistics are provided in Table 2.

[INSERT TABLE 2 ABOUT HERE]

4.2 Econometric strategy

Our econometric strategy is developed in two steps. In the first, we test whether lnECI and lnVAR are stationary processes, to avoid estimating spurious regressions. To do so, we use two unit root tests: the Im-Pesaran-Shin test (Im, Pesaran, & Shin, 2003), and the cross-sectional Im-Pesaran-Shin (CIPS) test (Pesaran, 2007). We first test for the presence of a unit root for lnECI and lnVAR: if H_0 is not rejected, we proceed with the test on the variables in first difference. In the Im-Pesaran-Shin (2003) test, we only use one-time lag, and we add a time trend and a series of cross-sectional means for each province to mitigate the cross-sectional dependence.

It is not enough to de-mean the data to eliminate the problem, however, because some panels (i.e. provinces) can react differently to common unobserved shocks. So, we also use the CIPS test, which consists in extending individual augmented Dickey-Fuller regressions with the cross-sectional means of the lagged levels and first differences of the individual regressor (i.e. lnECI and lnVAR, respectively) used as proxy for the unobserved common factors. The test is based on the null hypothesis that the variable under investigation has a unit root. We first test for the presence of a unit root in our focal variables in levels, and then in their first differences. If the test does not reject H_0 when the variables are in levels, but does reject it when they are in first differences, then we conclude that they are integrated of order 1, i.e. non-stationary. The results of the two tests are given in Table 3.

[INSERT TABLE 3 ABOUT HERE]

We find that all the tests reject the null hypothesis of lnECI (and ΔlnECI) being non-stationary at 1% level. In other words, lnECI is a stationary, or I(0) process. On the other hand, none of the tests reject H_0 on lnVAR, whereas they do reject it (at 1% level) on ΔlnVAR and $\Delta_2$lnVAR. We conclude that industry variety is non-stationary, or I(1). Our two focal variables thus show an opposite behavior: while the trend in (log) entropy is stochastic, the ECI tends to follow a mean-reverting process.

So, to avoid estimating a spurious relation between lnECI and lnVAR, in the second step we estimate the following equation where all variables are computed in first differences:

[5] $\Delta lnECI_{it} = \beta_1 \Delta lnVAR_{it} + \boldsymbol{X}'_{it}\beta_2 + \theta_t + \Delta\varepsilon_{it}$,

where i is the NUTS 3 region, t is the year, θ_t is a vector of year dummies to control for the business cycle and possible macroeconomic shocks (like those related to the 2008 economic crisis), and **X** is the vector of additional control variables, all computed in first differences and transformed into natural logarithms. In this way, we only use stationary variables and we control for unobserved province-specific fixed effects, such as the availability and quality of a region's infrastructure, or the type and quality of the local institutions. Since VAR and ECI are expressed in logs and first differences, we should interpret the parameter β_1 as the elasticity of the rate of growth in economic complexity vis-à-vis the rate of growth in entropy. To mitigate any spatial autocorrelation, in addition to ruling out the province-specific fixed effects, we also cluster the standard errors at NUTS 2 regional level[4].

Table 4 shows the pairwise correlations between the regressors. They are all very low, so we are confident that multicollinearity is not an issue.

[INSERT TABLE 4 ABOUT HERE]

Following the theoretical framework described in Section 2, we should consider entropy as an exogenous, predetermined variable, the growth of which is governed by a path-dependent process (Fontana, 2010). We nonetheless control for a potential simultaneity bias arising if economic complexity and entropy are determined or affected by a common unobserved factor. In the absence of suitable external instruments correlating with variety but not with economic complexity, we adopt Lewbel's instrumental variable approach (Lewbel, 2012). This method uses the conditional second moments of ΔlnVAR (or ΔlnRV and ΔlnUV) to address potential endogeneity. Parameters' identification occurs when the residuals of the first-stage regression are heteroskedastic, and at least a subset of the regressors used to estimate equation [5] correlates with the variance of these residuals, but is independent of the covariance between these first-stage residuals and the residuals from the second-stage regression, $\Delta\varepsilon_{it}$. If this condition is met, instruments are computed by multiplying the first-stage residuals by the mean-centered regressors. To test for the heteroskedasticity of the first-stage residuals we use a Breusch-Pagan test, where the null hypothesis is that errors are homoskedastic.

[4] Unfortunately, the limited number of years available in our panel prevents us from testing for Granger causality, so we proceed with a regression analysis instead.

We proceed as follows. In the first stage, we regress ΔlnECI on the other regressors, and we compute the residuals. Then we use the Breusch-Pagan statistic to test whether these residuals are heteroskedastic. We compute the instruments by mean-centering each regressor and multiplying it by the first-stage residuals. In the second stage, we estimate equation [5] using the instruments generated and a GMM approach. To test for the validity of our instrumentation strategy, we use the Kleibergen-Paap F statistic, and run the Hansen J test on the over-identifying restrictions. We also test for the exogeneity of our entropy measures using a difference in the Sargan test, where the null hypothesis that ΔlnVAR (or ΔlnRV and ΔlnUV) is exogenous is tested using a statistic distributed as a chi-squared with a number of degrees of freedom corresponding to the number of endogenous variables.

To test our second hypothesis, i.e. that entropy affects complexity more strongly where entropy and/or complexity are low, we re-estimate equation [5] on different sub-samples. First, we split our sample into two groups, depending on whether the regions' initial level of entropy or economic complexity (in 2008) is below or above the median. Regions are classified as "low VAR" (or "low ECI") if their level of entropy (or complexity) is below the median in 2008, or as "high VAR" ("high ECI") if it is above the median. According to our second hypothesis, we should only expect the estimated coefficient of ΔlnVAR to be positive and statistically significant (or stronger in magnitude) in the former group.

As shown in Figure 1, we also split our initial sample into four groups, combining the levels of variety and complexity in 2008, and ranking the regions as: "low VAR - low ECI" if their initial levels of industry variety and complexity were both below the corresponding medians; "high VAR - high ECI" if both variety and complexity were above the median; "low VAR - high ECI" if variety was below, but complexity was above the median; and "high VAR - low ECI" if variety was above, but complexity was below the median. We expect to find a stronger impact of industry variety on the ECI in the "low VAR - low ECI" group.

5. Results

Tables 5 and 6 show the results of the OLS and IV-GMM estimates, respectively. In Table 5, Column 1, we find that the estimated coefficient of ΔlnVAR is positive and statistically significant (at 5% level). Specifically, a 10% increase in industry variety (total entropy) corresponds to an average 1.3% rise in economic complexity. Column 2 shows that this effect is explained by both related (within-sector entropy) and unrelated (between-sector entropy) variety, the coefficient of the latter being almost twice that of the former. Columns 3 and 4 show that the previous results are robust to the inclusion of additional regional features like population density, human capital, and inward greenfield FDI. Among these regressors, only the coefficient of HK is statistically significant, but negative, meaning that a region's economic complexity grows faster when its accumulated HK is lower. The AIC and BIC statistics show that a model without these additional regional controls fits our data better, however.

[INSERT TABLE 5 ABOUT HERE]

Table 6 shows the results when we estimate equation 5 using the IV-GMM approach proposed by Lewbel (2012). Columns 1 and 2 confirm the results shown in Table 5, where the level of economic complexity grows more in regions where the entropy growth rate is higher. Specifically, our IV estimates show that a 10% increase in industry variety corresponds to an average 1.2% increase in economic complexity. We still find that most of this relationship is explained by the between-sector entropy growth rate, with an estimated coefficient of 0.095 as opposed to 0.035 for within-sector entropy growth. Columns 3 and 4 confirm that the results in Columns 1 and 2 are robust to the inclusion of additional regional controls.

In all the specifications, the value of the F statistic is much larger than 10, or than the 5% Stock and Yogo critical value for weak identification, while the Hansen J statistic never rejects the null hypothesis that the model is correctly specified. The difference in the Sargan statistic also never rejects the null hypothesis of exogeneity of our variety measures. This confirms our claim that entropy is a pre-determined variable, and its evolution in a given region is unaffected by the rate of growth in economic complexity.

[INSERT TABLE 6 ABOUT HERE]

We conclude that the empirical analysis conducted thus far corroborates our first hypothesis: the level of economic complexity in a region increases more rapidly where the level of entropy grows faster, particularly at the extensive margin. Therefore, firms in regions where unrelated variety grows faster have more opportunities to recombine and cross-fertilize with adjacent but different pieces of knowledge to generate new, more sophisticated products and/or new specializations.

As a final step, we check whether these results depend on a region's initial level of variety and economic complexity.

The results in Table 7 support our second hypothesis. In the upper part of the table, the estimated coefficients of ΔlnVAR are positive and significant only for low-variety or low-complexity regions, whereas we find no significant result for high-variety *or* high-complexity regions. Columns 2, 4, 6 and 8 show that these results are driven mainly by increases in unrelated (between-sector) variety, whereas the estimated coefficient of related variety is only weakly or not at all statistically significant. The lower part of the table refines this result: when we further split the sample, we find the relationship between lnVAR and lnECI significant and positive only for "low VAR - low ECI" regions, i.e. NUTS 3 regions featuring a low variety *and* a low complexity (Column 5 in Table 7).

A closer look at these regions shows that they are located mainly in the South of Italy, in Calabria (15%), Sicily (16%), and Puglia (9%), but also in Tuscany (9%) and Liguria (9%).

[INSERT TABLE 7 ABOUT HERE]

6. Conclusions

This paper investigates the relationship between industry variety and economic complexity at regional level, in an effort to suggest another explanation for regional economic development and disparities. Taking an evolutionary economic geography approach, we find that a greater increase in the level of entropy (proxied by industry diversification) makes economic complexity grow faster. This relationship should be stronger in regions with initially lower level of variety and/or complexity, where there is more room for the degree of industry diversification and product sophistication to grow.

We test our hypothesis on a panel of Italian NUTS 3 regions over a period of nine years, between 2008 and 2016. We first test whether our economic complexity and variety variables are stationary. Since the unit root tests reveal that economic complexity is a stationary process whereas industry variety is characterized by the presence of a unit root, we estimate a series of linear models using first-difference estimators to avoid identifying spurious relations using variables in levels.

Our results show that, as a region's level of entropy rises, so does its overall level of economic complexity. We also split total regional entropy into between-sector (unrelated) and within-sector (related) industry variety, finding the effect of the former larger than that of the latter. Our results are robust to the inclusion of additional regional features like human capital, population density and inward greenfield FDI projects, and also to endogeneity.

In line with our expectations, we find that the positive relationship between increasing variety and rising complexity only holds in regions with initially low levels of variety and complexity. These results can have two main implications. From an academic standpoint, to the best of our knowledge, this is the first contribution to identify the determinants of economic complexity at regional level. In doing so, we show that industry diversification can be seen as an exogenous phenomenon coinciding with more options for economic agents to specialize in new, more sophisticated products. At the same time, we show that a growth in variety is the most important factor explaining a region's rising level of economic complexity.

From a policy perspective, our paper provides useful pointers concerning the link between industry diversification and regions' economic development. Since economic complexity correlates positively with economic development, diversifying a region's knowledge portfolio is one of the most important ways to generate innovation, new specializations, and wealth. While this might be less important for mature and complex regional systems, it is crucial for less developed areas, where there is more room for both variety and complexity to grow. In short, our explanation of regions' economic development allows for diversification and specialization (in high-tech industries) to co-exist.

References

Accetturo, A., Di Giacinto, V., Micucci, G., & Pagnini, M. (2018). Geography, productivity, and trade: does selection explain why some locations are more productive than others? *Journal of Regional Science*, *58*(5), 949–979.

Antonietti, R., Cainelli, G., & Lupi, C. (2013). Vertical disintegration and spatial co-localization: The case of KIBS in the metropolitan region of Milan. *Economics Letters*, *118*(2), 360–363.

Balland, P.-A., Boschma, R., Crespo, J., & Rigby, D. L. (2018). Smart specialization policy in the European Union: relatedness, knowledge complexity and regional diversification. *Regional Studies*, 1–17.

Balland, P.-A., & Rigby, D. (2017). The geography of complex knowledge. *Economic Geography*, *93*(1), 1–23.

Basile, R., & Ciccarelli, C. (2018). The location of the Italian manufacturing industry, 1871–1911: a sectoral analysis. *Journal of Economic Geography*, *18*(3), 627–661.

Basile, R., Girardi, A., Mantuano, M., & Russo, G. (2018). Interregional migration of human capital and unemployment dynamics: evidence from Italian provinces. *German Economic Review*.

Beaudry, C., & Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, *38*(2), 318–337.

Blume, L. E., & Durlauf, S. N. (2006). *The Economy as an Evolving Complex System III*. (L. E. Blume & S. N. Durlauf, Eds.). Oxford: Oxford University Press.

Boschma, R., & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy. *Economic Geography*, *85*(3), 289–311.

Cadot, O., Carrère, C., & Strauss-Kahn, V. (2011). Export diversification: what's behind the hump? *Review of Economics and Statistics*, *93*(2), 590–605.

Calcagnini, G., & Perugini, F. (2019). Income distribution dynamics among Italian provinces. The role of bank foundations. *Applied Economics*, *51*(29), 3198–3211.

Cellini, R., & Torrisi, G. (2014). Regional resilience in Italy: a very long-run analysis. *Regional Studies*, *48*(11), 1779–1796.

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, *20*(2), 249–272.

Cicerone, G., McCann, P., & Venhorst, V. A. (2019). Promoting regional growth and innovation: relatedness, revealed comparative advantage and the product space. *Journal of Economic Geography*.

Cracolici, M. F., Cuffaro, M., & Nijkamp, P. (2007). Geographical distribution of unemployment: an analysis of provincial differences in Italy. *Growth and Change*, *38*(4), 649–670.

Eckaus, R. S. (1961). The North-South differential in Italian economic development. *Journal of Economic History*, *21*(3), 285–317.

Felipe, J., Kumar, U., Abdon, A., & Bacate, M. (2012). Product complexity and economic development. *Structural Change and Economic Dynamics*, *23*(1), 36–68.

Filippetti, A., Guy, F., & Iammarino, S. (2019). Regional disparities in the effect of training on employment. *Regional Studies*, *53*(2), 217–230.

Fontana, M. (2010). Can neoclassical economics handle complexity? The fallacy of the oil spot dynamic. *Journal of Economic Behavior & Organization*, *76*(3), 584–596.

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, *41*(5), 685–697.

Gao, J., & Zhou, T. (2018). Quantifying China's regional economic complexity. *Physica A: Statistical Mechanics and Its Applications*, *492*, 1591–1603.

Guerrieri, P., & Iammarino, S. (2007). Dynamics of export specialization in the regions of the Italian Mezzogiorno: persistence and change. *Regional Studies*, *41*(7), 933–948.

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., Chung, S., Jimenez, J., … Yildirim, M. A. (2014). *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. (R. Hausmann, C. A. Hidalgo, S. Bustos, M. Coscia, S. Chung, J. Jimenez, … M. A. Yildirim, Eds.). Cambridge, MA, USA: MIT Press.

Helliwell, J. F., & Putnam, R. D. (1995). Economic growth and social capital in Italy. *Eastern Economic Journal*, *21*(3), 295–307.

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

(2014). *Complexity: A Very Short Introduction*. Oxford: Oxford University Press.

Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, *115*(1), 53–74.

Inoua, S. (2016). A Simple Measure of Economic Complexity. Available at: https://arxiv.org/abs/1601.05012, *mimeo*.

Javorcik, B. S., Lo Turco, A., & Maggioni, D. (2018). New and improved: does FDI boost production complexity in host countries? *The Economic Journal*, *128*(614), 2507–2537.

Kemp-Benedict, E. (2014). *An Interpretation and Critique of the Method of Reflections* (MPRA working paper No. 60705).

Lagravinese, R. (2015). Economic crisis and rising gaps North-South: evidence from the Italian regions. *Cambridge Journal of Regions, Economy and Society*, *8*(2), 331–342.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, *30*(1), 67–80.

Mealy, P., Farmer, J. D., & Teytelboym, A. (2019). Interpreting economic complexity. *Science Advances*, *5*(1), eaau1705.

Missiaia, A. (2019). One market fits all? Market access and the origins of the Italian north–south divide. *Regional Studies, Regional Science*, *6*(1), 92–100.

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.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, *22*(2), 265–312.

Pinheiro, F. L., Alshamsi, A., Hartmann, D., Boschma, R., & Hidalgo, C. A. (2018). *Shooting High or Low: Do Countries Benefit from Entering Unrelated Activities?* (Papers in Evolutionary Economic Geography No. 18–07). Utrecht.

Reynolds, C., Agrawal, M., Lee, I., Zhan, C., Li, J., Taylor, P., … Roos, G. (2018). A sub-national economic complexity analysis of Australia’s states and territories. *Regional Studies*, *52*(5), 715–726.

Rungi, A., & Biancalani, F. (2019). Heterogeneous firms and the North–South divide in Italy. *Italian Economic Journal*, 5(3), 325–347.

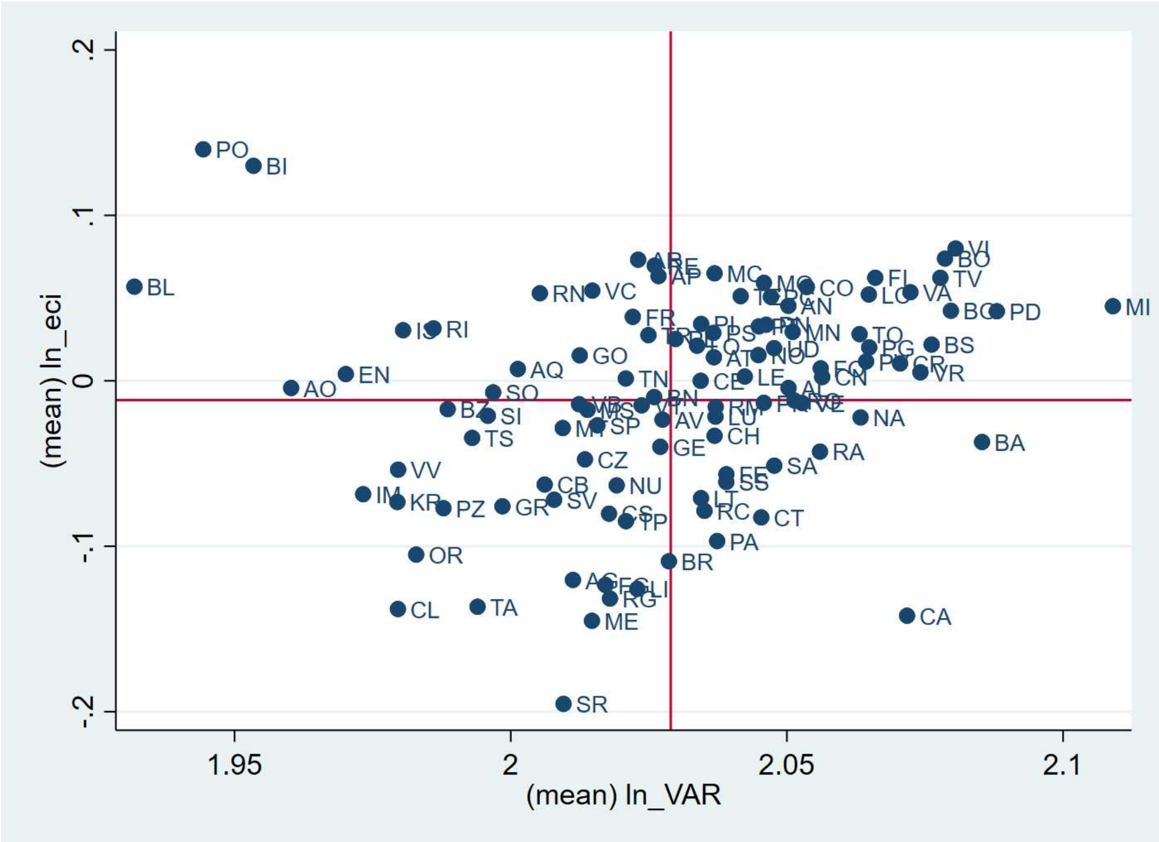
Schachter, G. (1967). Regional development in the Italian dual economy. *Economic Development and Cultural Change*, *15*(4), 398–407.

Sweet, C. M., & Eterovic Maggio, D. S. (2015). Do stronger intellectual property rights increase innovation? *World Development*, *66*, 665–677.

van Dam, A., & Frenken, K. (2019). Variety, complexity and economic development. *Papers in Evolutionary Economic Geography (PEEG)*.

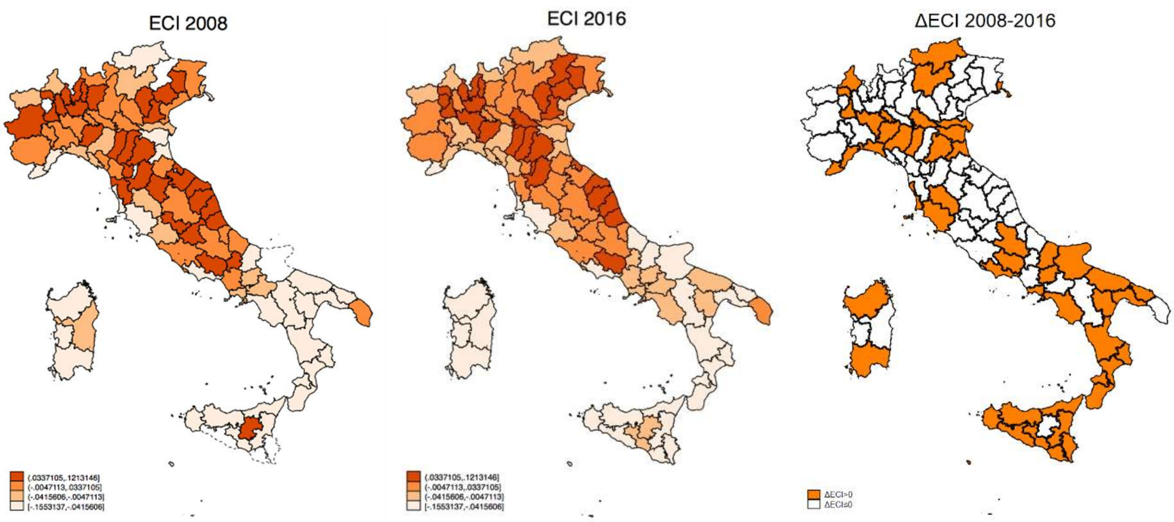
FIGURES AND TABLES

Figure 1. The Italian NUTS 3 regions in the variety-complexity space



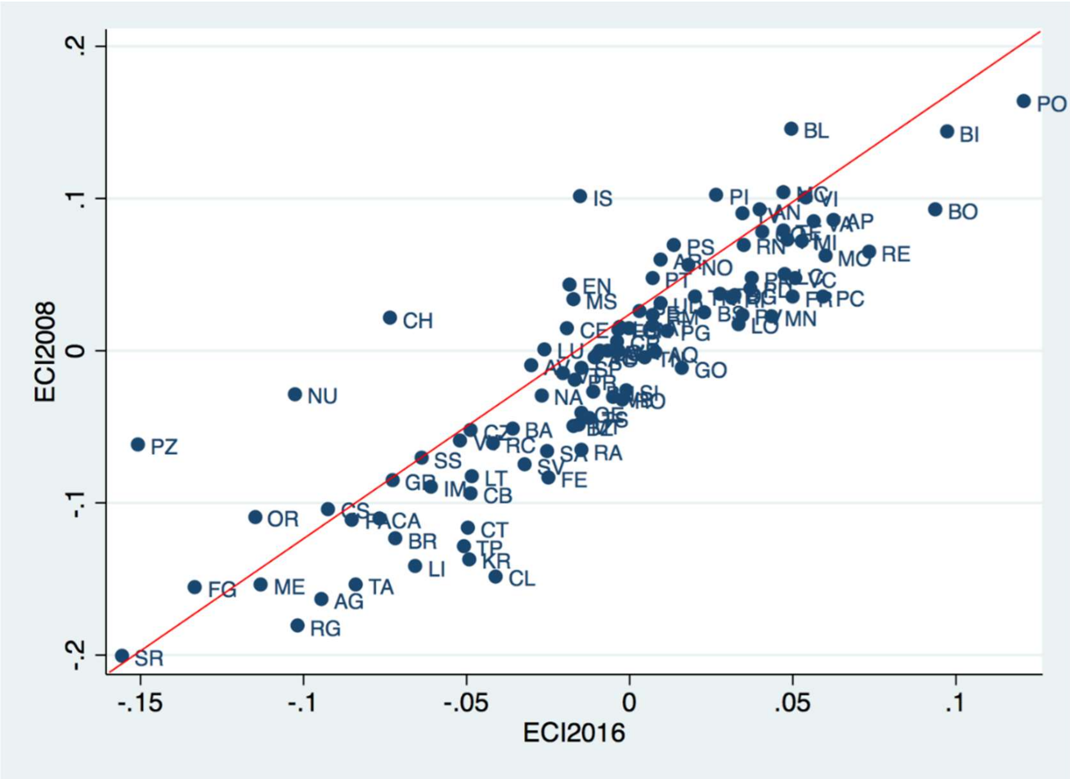
Source: authors' elaborations of ASIA and Coeweb (ISTAT) data

Figure 2. Geography of the ECI: in 2008 and 2016, and variation over the period



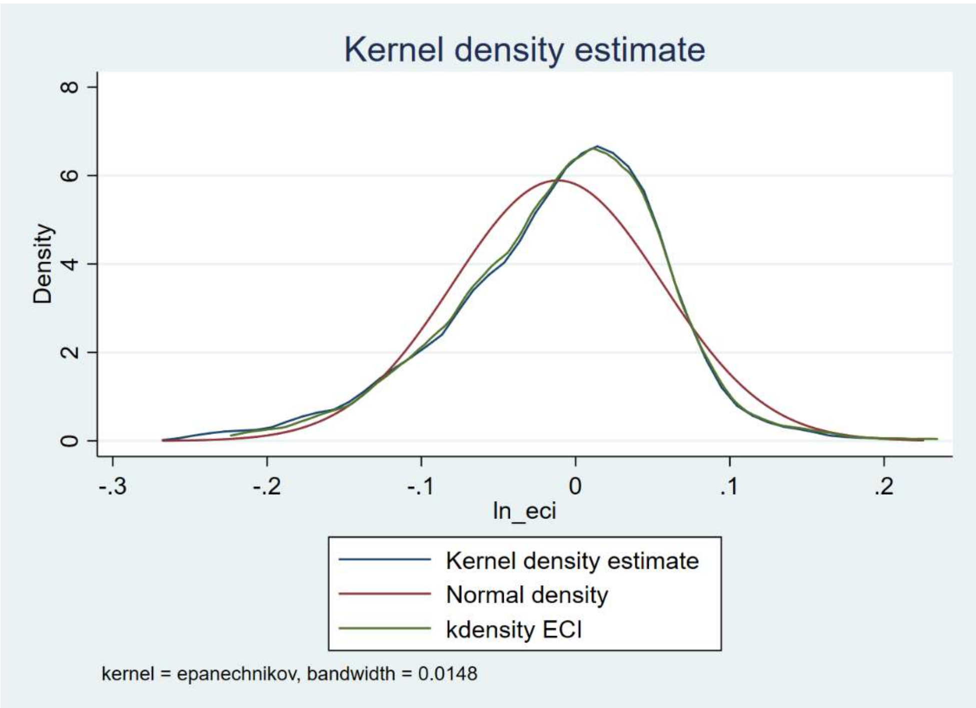
Source: authors' elaborations of Coeweb (ISTAT) data

Figure 3. Relative ECI positions of Italian provinces, 2008-16



Source: authors' elaborations of Coeweb (ISTAT) data

Figure 4. Kernel density of ECI and lnECI



Source: authors' elaborations of Coeweb (ISTAT) data

Table 1. Yearly correlations between variety and ECI

Year	Correlation
2008	0.134
2009	0.176
2010	0.072
2011	0.115
2012	0.223
2013	0.189
2014	0.205
2015	0.293
2016	0.280
Total	0.171

Table 2. Summary statistics and description of variables

VARIABLES	Source	Description	Mean	Std. Dev.	Min.	Max.
lnECI	Coeweb (ISTAT)	Economic complexity index, based on value of exports of three-digit industries (NACE Rev. 2)	-0.0116	0.0677	-0.253	0.211
lnVAR	ASIA (ISTAT)	Entropy index of five-digit industries: RV+UV	2.029	0.0364	1.833	2.160
lnUV	ASIA (ISTAT)	Unrelated variety index at two-digit industry level	1.618	0.0390	1.481	1.695
lnRV	ASIA (ISTAT)	Related variety index for five-digit industries within two-digit industries	0.940	0.0745	0.605	1.223
lnDEN	ASTI (ISTAT)	Resident population per km^2	0.559	0.770	-0.989	3.283
lnHK	ASTI (ISTAT)	Proportion of university graduates (bachelors + masters) in resident population	0.00358	0.00419	0	0.0216
lnIGFDI	FDI Markets	Stock of inward greenfield FDI projects	0.373	0.676	0	4.277
lnR&D	ISTAT	Annual R&D expenditure as proportion of GDP	-4.418	0.728	-5.523	0.673

Table 3. Unit root tests

Im-Pesaran-Shin (2003) unit root test		
Options included	**Variable**	**Statistic**
Time trend + cross-sectional means	lnECI	-12.285***
Time trend + cross-sectional means	ΔlnECI	-9.4076***
Time trend + cross-sectional means	lnVAR	1.245
Time trend + cross-sectional means	ΔlnVAR	-9.051***
Time trend + cross-sectional means	$\Delta_2$lnVAR	-10.05***

CIPS unit root test		
Options included	**Variable**	**Statistic**
Constant + time trend	lnECI	-3.134***
Constant + time trend	ΔlnECI	-3.672***
Constant + time trend	lnVAR	-2.063
Constant + time trend	ΔlnVAR	-3.212***
Constant + time trend	$\Delta_2$lnVAR	-4.115***

Notes: in the Im-Pesaran-Shin (2003) test the number of lags is set to 1. In the CIPS test (Pesaran, 2007) the number of lags is determined using the Portmanteau test for white noise. The relevant 10%, 5%, and 1% critical values are, respectively: -2.66, -2.75 and -2.91.

Table 4. Correlation matrix

	ΔlnVAR	ΔlnRV	ΔlnUV	ΔlnDEN	ΔlnHK	ΔlnIGFDI	ΔlnR&D
ΔlnVAR	1						
ΔlnRV	0.6425	1					
ΔlnUV	0.8135	0.0789	1				
ΔlnDEN	0.0557	0.0248	0.0535	1			
ΔlnHK	-0.0031	-0.0358	0.0257	0.0457	1		
ΔlnIGFDI	0.014	-0.0102	0.0277	0.0010	-0.0145	1	
ΔlnR&D	0.099	-0.054	0.1696	0.0152	-0.0418	0.0023	1

Table 5. Role of variety on economic complexity: OLS estimates, 2008-16

	(1)	(2)	(3)	(4)
ΔlnVAR	0.127**		0.123**	
	[0.046]		[0.046]	
ΔlnRV		0.045*		0.044*
		[0.023]		[0.023]
ΔlnUV		0.073***		0.072***
		[0.018]		[0.018]
ΔlnDEN			0.007	0.007
			[0.029]	[0.030]
ΔlnHK			-2.222**	-2.238**
			[1.004]	[1.004]
ΔlnIGFDI			0.002	0.002
			[0.002]	[0.002]
ΔlnR&D			-0.007	-0.007
			[0.010]	[0.010]
Year dummies	Yes	Yes	Yes	Yes
N	824	824	824	824
R^2	0.005	0.005	0.008	0.008
AIC	-3572.66	-3570.548	-3569.596	-3567.503
BIC	-3534.946	-3528.12	-3517.741	-3510.933

*** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at NUTS2 regional level.

Table 6. Economic complexity and variety: Lewbel's (2012) IV-GMM estimates

	(1)	(2)	(3)	(4)
ΔlnVAR	0.119***		0.129***	
	[0.039]		[0.041]	
ΔlnRV		0.035**		0.041***
		[0.011]		[0.015]
ΔlnUV		0.095***		0.073*
		[0.018]		[0.042]
ΔlnDEN			-0.011	-0.011
			[0.025]	[0.021]
ΔlnHK			-2.545***	-2.706***
			[0.673]	[0.604]
ΔlnIGFDI			0.002	0.003*
			[0.002]	[0.001]
ΔlnR&D			-0.002	-0.002
			[0.007]	[0.006]
Year dummies	Yes	Yes	Yes	Yes
N	824	824	824	824
R^2	0.005	0.004	0.007	0.007
Hansen test (p-value)	0.342	0.299	0.356	0.225
Endogeneity test (p-value)	0.813	0.679	0.566	0.819
Kleibergen-Paap F statistics	7627.9	2174.5	2346.5	2073.6
First stage: Breusch-Pagan test (p-value)				
Dep. var. = ΔlnVAR	0.000		0.000	
Dep. var. = ΔlnRV		0.000		0.000
Dep. var. = ΔlnUV		0.000		0.000

*** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at NUTS3 regional level.

Table 7. Role of variety on economic complexity by level of entropy and ECI

	(1)	(2)	(3)	(4)
	Low VAR	Low ECI	High VAR	High ECI
ΔlnVAR	0.193***	0.158***	0.030	0.025
	[0.051]	[0.052]	[0.056]	[0.064]
Year dummies	Yes	Yes	Yes	Yes
N	408	416	416	408
R^2	0.026	0.121	0.032	0.153
	(5)	(6)	(7)	(8)
	Low VAR Low ECI	Low VAR High ECI	High VAR Low ECI	High VAR High ECI
ΔlnVAR	0.228***	-0.179	0.034	0.067
	[0.049]	[0.335]	[0.046]	[0.079]
Year dummies	Yes	Yes	Yes	Yes
N	280	128	136	280
R^2	0.150	0.160	0.097	0.179

*** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at NUTS2 regional level.