From entropy 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

The aim of this paper is to analyse the relationship between entropy and economic complexity at the level of Italian NUTS 3 regions, borrowing elements from condensed-matter physics and recent contributions of evolutionary economic theory. Entropy is measured through a Theil index of industry variety, and economic complexity through the Hidalgo and Hausmann (2009) index using regional export data at three-digit industry level. Following the recent debates on this topic, entropy and economic complexity are considered two distinct phenomena. Higher levels of entropy induce higher levels of economic complexity. Results show that regions where entropy grows faster are also those where the growth rate of economic complexity is higher. Moreover, this relation holds only in regions with low initial levels of entropy and/or complexity, mainly located in the South of Italy, and that is robust to endogeneity using a Lewbel (2012) IV-GMM approach.

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

JEL codes: O33; R11; R12

1. Introduction

The concepts of entropy and complexity have been recently applied to economic development, international trade and industrial dynamics, to explain how economic activities are distributed across specializations (Frenken, Van Oort, & Verburg, 2007) and how economic systems engage in productions that require wider, and context-specific, sets of knowledge and skills (Hidalgo & Hausmann, 2009). There is increasing consensus among scholars that economic complexity is related to higher wealth: countries producing a diversified portfolio of highly sophisticated and less ubiquitous products are those experiencing higher levels of income per capita. Economic complexity favours the process of convergence in income per capita among regions (Hausmann & Hidalgo, 2011; Hausmann et al., 2014). However, two questions remain unanswered: why is the level of complexity in some countries/regions higher than in others? What drives the increase or decrease of complexity?

In this paper we try to answer these questions using two different, but complementary, explanations based on condensed-matter physics and on evolutionary economics. We posit that the increase in economic complexity in a region is strictly related to the increase in the level of entropy: as the regional economic structure diversify, the level of sophistication of the regional product basket increases.

Recent literature study economic complexity in relation with: product diversification in a country (Pinheiro, Alshamsi, Hartmann, Boschma, & Hidalgo, 2018), GDP per capita growth and speed of industrialization (Ferrarini & Scaramozzino, 2016; Hidalgo & Hausmann, 2009; Pugliese, Chiarotti, Zaccaria, & Pietronero, 2017; Sbardella et al., 2018), economic development and income/wage inequality (Gao & Zhou, 2018; Hartmann, Guevara, Jara-Figueroa, Aristarán, & Hidalgo, 2017; Sbardella, Pugliese, & Pietronero, 2017), and export basket diversification in a region (Cicerone, McCann, & Venhorst, 2019).

However, in all these studies economic complexity is taken as a pre-determined, exogenous variable, the origins of which just rely on the unobserved set of skills and capabilities a

country/region is endowed with. Only few papers have investigated the determinants of economic complexity. Sweet and Eterovic-Maggio (2015), for example, show that economic complexity, used to approximate the aggregate innovative outcome of a country, depends on the stringency of the domestic IPR system. Javorcik, Lo Turco and Maggioni (2018) show that manufacturing firms in Turkey tend to introduce more complex products when operating in regions and sectors with a higher propensity to supply foreign multinationals. Balland and Rigby (2017) and Balland et al. (2018), instead, find that complex knowledge and economic activities tend to co-agglomerate in large urban areas in the US, contributing to increase spatial inequality across regions.

Identifying the determinants of economic complexity is important for two main reasons. First, from a purely 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), or, differently, whether it should be treated as endogenous in growth regressions using appropriate econometric techniques. This point is crucial in order to avoid the complexity-growth relationship to be spurious and path-dependent. Second, from a policy perspective, if raising the product sophistication of a region is a way to improve its level of technological and economic development, then it becomes relevant identifying the leverages of economic complexity and target the industrial policies accordingly.

However, despite the few efforts, the theoretical and empirical analyses of what generates and stimulates economic complexity in a region, or a country, remain scanty. This paper investigates the determinants of economic complexity, arguing that it is a structural characteristic of a region depends on its level of entropy. Our focus is on Italian NUTS3 regions. Between 2008 and 2012, Italy is ranked within the top twenty countries for its level of economic complexity[1].

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

Italian regions show a high heterogeneity in skill endowment, performance and level of development. However, they are also characterized by a strong regional heterogeneity in terms of specializations, revealed comparative advantages, and density of the product space (Cicerone, McCann and Venhorst, 2019), which makes it interesting investigating the relationship between entropy and complexity at this level.

We measure entropy through a Theil index of industry variety (Frenken et al., 2007), which captures the higher probability in a region to randomly "pick" a different five-digit industry from an initial distribution. In this sense, entropy represents the state of our "lack of knowledge" and increases as much as an economic system is diversified. This general entropy measure can be further split into a within-entropy (or related variety) and a between-entropy (or unrelated variety) component: while the former refers to the weighted sum of the entropy at five-digit 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 instead the extensive margin of diversification.

We capture economic complexity using the Hidalgo and Hausmann (2009) index. For the period 2008-16, we use NUTS3 regional export data at three-digit industry level of classification provided by the Italian Statistical Institute (ISTAT) to build revealed comparative advantages and the proximity matrix among products (sectors). We define regional economic complexity (ECI) as the average knowledge intensity of the products (sectors) that the region exports, which underlies the state of sophistication of the knowledge accumulated in the region itself.

Our empirical strategy develops in two steps. We first run a unit root test to check whether our entropy and complexity indicators are stationary processes. Using this information, we subsequently run a regression analysis where we add to our focal regressor, industry variety, a series of controls on regional human capital, population density and number of inward greenfield FDI projects. To account for potential simultaneity, we adopt the 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. In addition, we check whether the entropy-complexity relation depends on the initial level of entropy and economic complexity of the region.

Our results confirm that regions where entropy increased faster are also those where the level of economic complexity grew the most. This result is robust to the inclusion of additional regional characteristics, fixed effects and simultaneity. However, this effect is not homogeneous across regions, but is significant only where the initial level of entropy and complexity is lower.

The rest of the paper is structured as follows: Section 2 presents the recent empirical literature on economic complexity; Section 3 presents the theoretical background and discusses the approaches used to explain the link between entropy and economic complexity; Section 4 presents the methodological approach, describing the data and the econometric analysis; the results are presented and discussed in Section 5, while Section 6 concludes.

2. Related literature

In the last two decades the concept of economic complexity has become very popular within the economic disciplines. A search on Scopus reveals that, between 2000 and 2018, the production of scientific documents (in economics, econometrics, finance, business, management and accounting) reporting "economic complexity" in the abstract, title or keywords has grown by a factor of 9, passing from 41 to 396, as shown in Figure 1.

[INSERT FIGURE 1 ABOUT HERE]

Starting from the seminal contributions of Hidalgo et al. (2007) and Hidalgo and Hausmann (2009), the concept of economic complexity has been increasingly associated to 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 increases the higher is the number of different exported product; and the ubiquity of a product, which increases the lower is the number of countries producing or exporting that product. The underlying mechanism is that countries differ in their level of economic complexity (and development) because they are endowed with different sets of skills and capabilities.

Relying on this framework, in the last few years many scholars assessed the role of economic complexity in explaining aggregate economic outcomes, like the growth of GDP per capita or income inequality. Hidalgo and Hausmann (2009) find a positive correlation between their measures of economic complexity and the growth rate of GDP per capita in a cross-country setting. This result is confirmed by Felipe et al. (2012) on more than 5000 products and 124 countries, and by Ferrarini and Scaramozzino (2016), Pugliese et al. (2017) and Sbardella et al. (2018) who use fitness as an alternative measure of economic complexity. A higher level of economic complexity has been found associated to a lower income inequality by Hartmann et al. (2017), on a sample of over 150 countries between 1963 and 2008. Sbardella et al. (2017) show that the relationship between economic complexity and wage inequality is inverted U-shaped: it increases during the earliest stages of development and it decreases during the latest, when the level of economic complexity is higher. The negative correlation between complexity and income inequality is also documented by Gao and Zhou (2018) for Chinese regions in 2000-15.

Recently, other studies looked at the possible role of economic complexity in affecting the capability of countries to diversify their product portfolio, or to develop new specializations in unrelated industries. In this respect, Pinheiro et al. (2018) find that the highest probability of

unrelated diversification belongs to countries with an intermediate level of development and economic complexity.

All these papers, however, use complexity as an exogenous predictor and postulate that it is 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 focused on economic complexity as a dependent variable, trying to explain explicitly why only some regions or countries show a higher tendency to increase their level of knowledge complexity. Among these, Sweet and Eferovic-Maggio (2015) analyse whether a stronger intellectual property rights (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 underline that more stringent IPR laws increase the capability of a country to increase the level of sophistication of its products. However, this impact holds only for countries where the level of development, human capital and complexity are high.

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

Another possible driver of economic complexity is spatial agglomeration, specifically in the form of urban economies and/or scale effects emerging in large cities. In this respect, Balland and Rigby (2017) and Balland et al. (2018) document that complex knowledge-intensive activities, like patenting, tend to concentrate in large metropolitan areas in the US. The reason is that complex activities and processes require a deeper division of labour, i.e. a higher

specialization, and can be developed more efficiently where knowledge coordination is higher and knowledge spillovers are easier, as in large cities (Antonietti, Cainelli, & Lupi, 2013).

In this paper, we provide an alternative explanation, and we posit that increasing complexity in a region is the outcome of its ever-increasing level of entropy. In doing so, we borrow elements from condensed-matter physics and evolutionary economics: from the former, we take the argument that the emerging patterns of complex (trade) relationships among firms can be assimilated to the emergence of complex macroscopic arrangements of particles. From the latter, we take the idea that the evolution of complexity depends on the entry and exit of products in the export basket of a region, which, in turn, depends on the degree of diversification of an economy.

3. The link between entropy and complexity

3.1 From condensed-matter physics...

To explain the relationship between entropy and complexity, we focus on condensed-matter physics and in particular on the second law of thermodynamics. This law states that the level of entropy of an isolated system is always increasing as long as the system evolves over time. According to some physicists (Baranger 2001; Carroll 2016), entropy represents the tendency of things to break down and of particles to wiggle and (re)arrange themselves into new building blocks. In this respect, an ever-increasing entropy becomes essential for the emergence of complex systems (Damasceno, Engel, & Glotzer, 2012; Glotzer & Solomon, 2007; Zhang & Glotzer, 2004). Carroll (2016, p. 235) argues that: "*The appearance of complexity isn't just compatible with increasing entropy; it relies on it.* [...] *The only reason complex structures form at all is because the universe is undergoing a gradual evolution from very low entropy to very high entropy. "Disorder" is growing, and that's precisely what permits complexity to appear end endure for a long time"*.

Zhang and Glotzer (2004), Glotzer and Solomon (2007) and Damasceno, Engel and Glotzer (2012), studied the self-assembly of patchy particles into ordered complex shapes, like icosahedra, square pyramids, quasicrystals, chains or staircases starting from very simple arrangements, like tetrahedra. The main finding of their simulations is that particles self-assemble into complex forms because of increasing entropy, where this should not be taken in terms of "disorder", but rather in terms of "options" for a particle to wiggle. The higher the level of entropy of a system, the higher the number of options for a particle to rearrange its position, and the higher is the chance to combine these positions into a high-order complex shape (Carroll, 2016).

The entropy-complexity relation can be also explained using the analogy with the Polya's urn adopted by Loreto et al. (2016). In this respect, entropy can be associated with the variety of balls of different colour: an increase in entropy implies a higher possibility to pick each time a ball of a different colour, and then to increase the number of colours in the urn. This process opens new possibilities, that, in turn, can be recombined and generate other novelties following a process of expansion of what Kauffman (1995) defines "adjacent possible". The outcome is that increasing entropy triggers innovation, and complexity.

3.2 ... to the Evolutionary Economics Approach

From the beginning of the 21st century, among scholars has emerged the need to combine economics and complexity science, to overcome some possible limitation linked to the neoclassical economics theory. Economies can be considered complex systems as far as they present two main characteristics: [i] they consist in a number of heterogenous agent organised in a great variety of groups; and [ii] they exhibit non-linear patterns, because agents act in different moment of space and time (Blume & Durlauf, 2006; Fontana, 2010). Thus, if we compare balls, or particles, to firms, it follows that a higher entropy, which is associated to a

higher industry variety, increases the options for firms to create links. These links can take different forms: trade relationships, joint ventures, alliances, buyer-supplier relationships, R&D networks, and so on. In addition, firms in a highly diversified region have many options to (re)arrange in complex networks of specializations and (trade) relationships, and generate new, and more sophisticated, products (Holland, 2014; Simon, 1969).

Complementary to the physic approach, evolutionary economics and combinatorial calculus try to explain the relationship between entropy and complexity from another point of view. Inoua (2016) and Van Dam and Frenken (2019), among the others, observe that countries and regions develop by accumulating skills and capabilities, that are complementary to available natural resources. The ability of a country, or region, to create products depends on the number of available capabilities, and on the possibility to combine them. Inoua (2016) shows that, in the long run, the growth rate of an economy is proportional to the log diversification of knowhow that, in turn, is directly proportional to the log diversification of products. Using COMTRADE export data on 160 countries in the world, the ECI, which reflects the average product sophistication in a country, is found highly correlated to the degree of diversification of the economy, measured through a technology development index.

Van Dam and Frenken (2019) provide a similar argument, assuming that a product is the outcome of a string of capabilities, the number of which defines the product 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 $d(n)$ equal to 2^n, for an average product length $\overline{s}(n)$of $n/2$. The term d represents the variety of products that an economy realizes with its n capabilities, and it follows that both $\ln(d)$ and the average product length $\overline{s}$ are proportional to n. Combining these two elements, we have that $\ln(d)$ is proportional to $\overline{s}(n)$: 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, new and more sophisticated products are produced, income increases, and wages increase as well. At the same time, some products, usually the simplest, exit from the product portfolio because they become too expensive to be realized. Van Dam and Frenken (2019) demonstrate that, when a country loses its less sophisticated goods and services, the average economic complexity increases.

This fact allows reproducing the dynamics of diversification and the hump-shaped relationship between variety and economic development. In the initial stage, of development the rate of product variety grows exponentially, and exceeds the speed of accumulation of capabilities. This means that a country is generating new, and even more sophisticated products, without losing any of them. In this phase, both variety and complexity increase, the latter increasing linearly in the number of capabilities. In the transition stage, countries lose their simplest products, but the economies are still diversifying, but at a lower rate. However, the economic complexity is still increasing, because product exit makes the average product length higher. In the last stage, the number of products exists increases so much that the overall variety decreases, while economic complexity reaches its maximum.

Combining all these elements, we derive the following: countries and regions develop accumulating new capabilities; these capabilities are recombined to generate new products; as far as the number of different capabilities increases, the chances to recombine different pieces of knowhow increase and the average length of products increases as well, implying a general increase in economic complexity. As countries and regions reach a certain level of development and diversification, product exits overcome entries, variety decreases, and complexity stabilizes.

Therefore, our first hypothesis that regions where industry variety, or total entropy, increases are also those where the level of complexity increases too.

However, the effect of entropy on complexity can vary according to the initial state of the system. In general, we do expect the relation to be stronger in regions where the initial levels of entropy and complexity are low, whereas we do not expect any significant relationship in highly developed regions, where both diversification and complexity are high. In regions where both diversification and complexity are low, an increase in entropy can have a high chance to improve the degree of sophistication of production. On the other hand, in regions where both entropy and complexity are high the scope for an increase in total entropy to further stimulate complexity is limited, or absent.

For these reasons, our second hypothesis is that the impact of increasing entropy on economic complexity is stronger in regions with a low initial level of entropy and/or complexity.

4. Empirical analysis

4.1 Data and variables

We test our hypothesis using two main data sources on 103 Italian administrative provinces (NUTS 3 regions) between 2008 and 2016. To compute our industry variety index, we use data from the Archive of Italian Active Firms (ASIA), administered by ISTAT. This dataset provides information on the number of plants and employees for each NUTS 3 region and five-digit industry, using the NACE Rev. 2 classification. The data to compute the economic complexity index, instead, come from the Coeweb archive on foreign trade statistics (ISTAT) and concern yearly exports of Italian NUTS 3 regions and (NACE Rev. 2) three-digit industry level. We also use the ISTAT-ASTI database (Atlas of Territorial Infrastructure Statistics) to collect information on additional regional characteristics.

Following Frenken et al. (2007), our entropy measure is given by the Theil index of industry variety (VAR), computed at five-digit level. The total level of industry variety is the sum of two components: related (or within-entropy) and unrelated (or between-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 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 on the intensive margin respectively. The former occurs when the number of new industries rises, while the latter occurs when the distribution of employment across existing industries becomes more equal, implying convergence in the employment shares across sectors in a region.

The economic complexity index, ECI, is computed using the eigenvector method as in Balland (2017). We use a revisited version of his knowledge complexity index, where we replace patent data with export data at the province level, in line with Hidalgo and Hausmann (2009). Following Reynolds et al. (2018), we use export data as an indicator of international competitiveness; therefore, our ECI is 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 of province *p* in (three-digit) industry *i*; if the index is higher than 1 (RCA>1), the province has a revealed comparative advantage in that industry. From the RCA index we derive the ubiquity and diversity measures: the former corresponds to the number of provinces with comparative advantage in an industry, while the latter to the number of industries in which a province has a comparative advantage. Putting these 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. Thus, our ECI index can be defined as the average knowledge intensity of the products (sectors) that are exported by a province, which underlies the state of sophistication and diversification of the knowledge accumulated in that area.

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 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 higher the more diversified is the industry mix of the region. Differently, the ECI index does not capture diversification. Despite being 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: specifically, the dot product between the knowledge diversity index $k_{c,0}$ and the ECI is zero, implying that the two measures are independent. This means that the ECI captures a different kind of information with respect to diversity: in particular, it is highly related to the specialization of countries in high or low-quality products, where high-ECI countries are those mostly specialized in technologically advanced, and unique, products, whereas low-ECI countries are those specialized in low-quality products. Therefore, the ECI is closer to a measure of (trade) specialization (in high or low-quality goods) than of diversification, and not an algebraic transformation of the entropy index.

To corroborate the idea that the two indicators capture different phenomena, we provide their yearly correlation. Table 1 shows that the pairwise correlations between the two indicators are modest[2].

[INSERT TABLE 1 ABOUT HERE]

As an additional check, 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 (low) levels of both entropy and complexity (i.e. lying in the first and third quadrant). Figure 2 shows that there are several regions (one-third of the sample) positioned in the second and fourth quadrant, where entropy is low (i.e. below the mean), but complexity is high (i.e. above the mean), and where entropy is high, but complexity is low, respectively.

[INSERT FIGURE 2 ABOUT HERE]

Finally, ECI and entropy variables follow two different trends: we find that while entropy is characterized by a stochastic trend with the presence of a unit root, economic complexity is stationary (as will be shown later in Section 4.2 and Table 3). If both indicators measured the same diversification pattern, we would expect their data generating process (or their trend) to be similar, which is not the case.

Figure 3 shows the geographical distribution of the ECI (in levels) in 2008 and 2016, and of the average yearly growth rate of ECI between 2008 and 2016, where in the third map coloured

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

regions are those where the average growth rate of ECI is positive. Interestingly, we note that while the highest levels of economic complexity are for regions in the Centre and North of Italy, in the South of Italy we find a higher number of provinces where ECI has increased.

[INSERT FIGURE 3 ABOUT HERE]

Figure 4 shows in which provinces the ECI has increased or decreased between 2008 and 2016. We register an increase of economic complexity in many provinces, specifically those lying below the red line, whereas we find few provinces where ECI has decreased. Interestingly, we also observe that in 2016 the ECI was the highest 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 4 ABOUT HERE]

We also include other regional characteristics that can confound the relationship between regional entropy and economic complexity. First, we control for population density (DEN), computed as resident population per squared kilometre, which captures the role that dense urban areas can play in favouring the spatial concentration of complex activities (Balland et al., 2018). Second, we control for the regional stock of human capital (HK), computed as the number of bachelor and master's degree graduates on total resident population in the region. Third, we add the flows of inward greenfield FDI projects (IGFDI), that are 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 come from fDi Markets, a database administered by the Financial Times Ltd. and that provides up-to-date information on

greenfield FDI projects operated by foreign multinationals across the world. For our purposes, we take information on yearly inward FDI projects in Italian NUTS 3 regions, that we cumulate over the focal period 2008-16 to build the corresponding stock[3].

Finally, we also control for the R&D intensity of the region, measured as total R&D expenditure 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 the econometric analysis, we transform all the variables in natural logarithms, except ECI and HK for which we have, respectively, a series of negative values and some missing values for some regions and years. As regard ECI, we adopt the following transformation: lnECI=ln(ECI+1), which allows reproducing a kernel distribution that is very close to that of ECI, as shown in Figure 5.

[INSERT FIGURE 5 ABOUT HERE]

For what concerns HK, we first replace the few missing values with 0s, and then we operate the following transformation: lnHK=ln(HK+1). Variables' description and summary statistics are reported in Table 2.

[INSERT TABLE 2 ABOUT HERE]

4.2 Econometric strategy

Our econometric strategy is developed in two steps. In the first step, we test whether lnECI and lnVAR are stationary processes, to avoid estimating spurious regressions. To do so, we use two

[3] We also consider the yearly flows of inward FDI instead of the stock. We do not find any change in results.

unit root tests: the Im–Pesaran–Shin (Im, Pesaran, & Shin, 2003) and the cross-sectional Im, Pesaran and Shin (CIPS) test proposed by Pesaran (2007). We first test for the presence of a unit root for lnECI and lnVAR in levels: in case of non-rejection of H_0, we proceed with the test on the variables in first difference. In the Im-Pesaran and Shin (2003) test, we use only 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.

However, de-meaning the data is not sufficient to eliminate the problem, as some panels (i.e. provinces) can react differently to common unobserved shocks. Therefore, we also use the Pesaran (2007) test, which consists in extending an individual augmented Dickey-Fuller (ADF) regressions with the cross-sectional means of the lagged levels and first differences of the individual regressor (i.e. lnECI and lnVAR respectively) that are 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 variables are in levels, but rejects 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 reported in Table 3.

[INSERT TABLE 3 ABOUT HERE]

We find that all the tests reject the null hypothesis of non-stationarity of lnECI (and ΔlnECI) at 1% level: in other words, between 2008 and 2016, lnECI is a stationary, I(0), process. Differently, all the tests do not reject the null hypothesis on lnVAR, whereas they reject it (at 1% level) on ΔlnVAR and $\Delta_2$lnVAR. We conclude that industry variety is a non-stationary, I(1), process. Therefore, our two focal variables show an opposite behaviour: while the trend in (log) entropy is stochastic, the economic complexity index tends to follow a mean-reverting

process. The stationarity of economic complexity is not surprising. As described in Section 3, recent studies have shown that complexity can increase or decrease over time. This is due either to product exit, when countries/regions do not find convenient to keep the simplest products in their portfolio (van Dam and Frenken 2019), or because of the natural tendency of highly complex systems to stabilize or even decrease their level of complexity (Carroll, 2016).

Therefore, to avoid estimating a spurious relation between lnECI and lnVAR in levels, 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 that 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 difference and transformed in natural logarithm. In this way, we only use stationary variables and we control for unobserved province-specific fixed effects, like regional infrastructure quality and endowment or like the quality and type of local institutions. Since VAR and ECI are expressed in logs and in first differences, we should interpret the parameter β_1 as the elasticity of the growth rate of economic complexity with respect to the growth rate of entropy. To mitigate the presence of spatial autocorrelation, we not only rule out the province fixed effects, but we also cluster the standard errors at NUTS 2 regional level[4].

[4] Unfortunately, the limited amount of available years in our panel does not allow testing for panel Granger causality, so we decided to proceed with a regression analysis.

Table 4 shows the pairwise correlations among the regressors: since they are all very low, we are confident that multicollinearity is not an issue.

[INSERT TABLE 4 ABOUT HERE]

Despite following the theoretical framework described in Section 2, we should consider entropy as an exogenous, pre-determined variable, the growth of which is governed by a path-dependent process (Fontana, 2010). Thus, we control for a potential simultaneity bias arising if economic complexity and entropy are determined, or affected, by a common unobserved factor. In absence of suitable external instruments that correlate with entropy but not with economic complexity, we adopt the Lewbel's instrumental variable approach (Lewbel, 2012). This method uses the conditional second moments of ΔlnVAR (or ΔlnRV and ΔlnUV) to address potential endogeneity. Identification occurs when the residuals of the first-stage regression are heteroskedastic and at least a subset of the regressors used for estimating equation [5] is correlated with the variance of these residuals but is independent from the covariance between these first-stage residuals and the residuals from the second-stage regression, $\Delta\varepsilon_{it}$. If this condition is satisfied, instruments are computed multiplying the first-stage residuals by the mean-centred 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.

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 generated instruments and a GMM approach. To test for the validity of our instrumentation strategy, we report the Kleibergen-Paap F statistic and the Hansen J test on the over-identifying

restrictions. In addition, we test for the exogeneity of our entropy measures using a difference in Sargan test, where the null hypothesis that ΔlnVAR (or ΔlnRV and ΔlnUV) is exogenous is tested using a test statistic that is distributed as a chi-squared with a number of degrees of freedom that corresponds 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 in two groups, according to whether the initial (i.e. in 2008) levels of entropy or economic complexity is below or above the median. Regions are classified as "low VAR" ("low ECI") if their level of entropy (complexity) in 2008 is below the median, and "high VAR" ("high ECI") if it is above the median. According to our second hypothesis, we should expect the estimated coefficient of ΔlnVAR to be positive and statistically significant (or stronger in magnitude) only in the former group.

Following Figure 2, we also split our initial sample in four groups, combining the 2008 level of entropy and complexity and ranking regions using the median of the two indicators: "low VAR-low ECI" regions are those characterized by initial levels of both entropy and complexity below their corresponding median; "high VAR-high ECI", on the contrary, are those regions where both entropy and complexity are above the median; "low VAR-high ECI" are regions with an initial level of entropy below the median but a level of complexity above the median, and finally "high VAR-low ECI" are regions with above-median level of entropy and below-median level of economic complexity. We do expect to find a stronger impact of variety on 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, we find that a 10% increase in total (five-digit) entropy corresponds to an average 1.3% increase in economic complexity. Column 2 shows that this effect is explained by both related (within-entropy) and unrelated variety (between-entropy), the coefficient of the latter being almost twice than that of the former. We find that a 10% increase in the growth rate of related variety corresponds to an average 0.45% increase in the growth rate of economic complexity, while when unrelated variety grows by 10% in a region, economic complexity grows by an average 0.73%. Columns 3 and 4 show that the previous results are robust to the inclusion of additional regional characteristics like population density, human capital and inward greenfield FDI projects. Among these regressors, only the coefficient of human capital is found to be statistically significant but negative, meaning that economic complexity in a region grows faster when human capital accumulation is lower. However, the AIC and BIC statistics show that a model without additional regional controls better fits our data.

[INSERT TABLE 5 ABOUT HERE]

Table 6 shows the results when we estimate equation 5 through the IV-GMM approach proposed by Lewbel (2012). Columns 1 and 2 confirm the results shown on Table 5, where the level of economic complexity grows more in regions where the growth rate of entropy is higher. Specifically, our IV estimates show that a 10% increase in the growth rate of industry variety corresponds to an average 1.2% increase in economic complexity. Moreover, we still find that the major part of this relationship is explained by the growth rate in between-entropy, with an estimated coefficient of 0.095 with respect the 0.035 of within entropy. Columns 3 and 4 confirm that 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 and of the 5% Stock and Yogo critical value for weak identification; while the Hansen J test never rejects the null hypothesis that the model is correctly specified. Moreover, the difference in Sargan test does never reject the null hypothesis of exogeneity of our entropy measures. This confirms our idea that entropy is a pre-determined variable, the evolution of which in a region is not affected by the speed of growth of economic complexity.

[INSERT TABLE 6 ABOUT HERE]

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

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

Results in Table 7 corroborate our second hypothesis. In the upper part of the table, the estimated coefficients of ΔlnVAR are positive and significant only for low-entropy or low-complexity regions, whereas we do not find any significant result for the high-entropy *or* high-complexity regions. Columns 2, 4, 6 and 8 show that these results are mainly driven by increases in unrelated variety (between-entropy), whereas the estimated coefficient of related variety is weakly, or not, statistically significant. The bottom part of the table refines this result: when we further split the sample, we find that the relationship between lnVAR and lnECI is significant and positive only for "low VAR-low ECI" regions, i.e. NUTS 3 regions characterized by low entropy *and* low complexity (Column 5 in Table 7).

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

[INSERT TABLE 7 ABOUT HERE]

6. Conclusions

This paper investigates the relationship between entropy and economic complexity at regional level. This relation can be viewed from two perspectives: one related to condensed-matter physics, the other on evolutionary economics and combinatorial calculus. From both perspectives, we draw that a higher increase in the level of entropy, proxied by industry diversification, induces economic complexity to grow faster. This relation should be stronger in regions with initial low level of entropy and/or complexity, where the scope to further increase the degree of industry diversification and of product sophistication is higher.

We test our hypothesis on a panel of Italian NUTS 3 regions and nine years, between 2008 and 2016. We first test for the stationarity of our economic complexity and entropy variables. 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 a first-difference estimators to avoid identifying spurious relations in levels.

Our results reveal that, as the level of entropy grows faster in a region, so does the overall level of economic complexity. We also split total regional entropy into a between (or unrelated variety) and a within (or related variety) component, and we find that the effect of the former is larger than that of the latter. Our results are robust to the inclusion of additional regional characteristics like human capital, population density and inward greenfield FDI projects, and to endogeneity.

Moreover, in line with expectations, we find that the positive relationship between increasing entropy and complexity holds only in regions characterized by an initial low level of entropy and complexity.

These results can have two main implications. From an academic point of view, to the best of our knowledge, this is the first contribution identifying the factors explaining the regional heterogeneity in economic complexity. In doing so, we show that entropy can be thought as an exogenous phenomenon, which corresponds to increasing options for economic agents to specialise in new and more sophisticated products. At the same time, we show that increasing entropy is the most important factor explaining increasing complexity in a region.

From a policy perspective, our paper provides useful advices on the linkage between industry diversification and regional economic development. Since economic complexity is positively correlated with economic development, diversifying the knowledge portfolio of the region is one of the most relevant leverages to generate innovation, new specializations and wealth. If this could be less relevant for mature and complex regional systems, it becomes crucial for less developed regions, where the scope for both entropy and complexity to grow is higher.

References

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. https://doi.org/10.1016/J.ECONLET.2012.11.031

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. https://doi.org/10.1080/00343404.2018.1437900

Balland, P.-A., & Rigby, D. (2017). The Geography of Complex Knowledge. *Economic Geography*, *93*(1), 1–23. https://doi.org/10.1080/00130095.2016.1205947

Baranger, M., & Institute, N. E. C. S. (2001). *Chaos, Complexity, and Entropy: A Physics Talk for Non-physicists*. New England Complex Systems Institute. Retrieved from https://books.google.it/books?id=duzsMgEACAAJ

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.

Cadot, O., Carrère, C., & Strauss-Kahn, V. (2011). Export Diversification: What's behind the Hump? *Review of Economics and Statistics*, *93*(2), 590–605. https://doi.org/10.1162/REST_a_00078

Carroll, S. (2016). *The Big Picture: On the Origins of Life, Meaning, and the Universe Itself*. London: Onerworld Publications. https://doi.org/10.1063/PT.3.3399

Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, *20*(2), 249–272. https://doi.org/10.1016/S0261-5606(00)00048-6

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. https://doi.org/10.1093/jeg/lbz001

Damasceno, P. F., Engel, M., & Glotzer, S. C. (2012). Predictive Self-Assembly of Polyhedra into Complex Structures. *Science*. American Association for the Advancement of Science. https://doi.org/10.2307/23267917

Felipe, J., Kumar, U., Abdon, A., & Bacate, M. (2012). Product complexity and economic development. *Structural Change and Economic Dynamics*, *23*(1), 36–68. https://doi.org/10.1016/J.STRUECO.2011.08.003

Ferrarini, B., & Scaramozzino, P. (2016). Production complexity, adaptability and economic growth. *Structural Change and Economic Dynamics*, *37*, 52–61. https://doi.org/10.1016/J.STRUECO.2015.12.001

Fontana, M. (2010). Can neoclassical economics handle complexity? The fallacy of the oil spot dynamic. *Journal of Economic Behavior & Organization*, *76*(3), 584–596. https://doi.org/10.1016/J.JEBO.2010.08.010

Frenken, K., Van Oort, F., & Verburg, T. (2007). Related Variety, Unrelated Variety and Regional Economic Growth. *Regional Studies*, *41*(5), 685–697. https://doi.org/10.1080/00343400601120296

Gao, J., & Zhou, T. (2018). Quantifying China's regional economic complexity. *Physica A: Statistical Mechanics and Its Applications*, *492*, 1591–1603. https://doi.org/10.1016/J.PHYSA.2017.11.084

Glotzer, S. C., & Solomon, M. J. (2007). Anisotropy of building blocks and their assembly into complex structures. *Nature Materials*, *6*(8), 557–562. https://doi.org/10.1038/nmat1949

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. https://doi.org/10.1016/J.WORLDDEV.2016.12.020

Hausmann, R., & Hidalgo, C. A. (2011). The network structure of economic output. *Journal of Economic Growth*, *16*(4), 309–342. https://doi.org/10.1007/s10887-011-9071-4

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. Retrieved from https://books.google.it/books?hl=en&lr=&id=cp-NAgAAQBAJ&oi=fnd&pg=PA10&ots=RWJzGaQQPu&sig=VvdpyysvazGt1IJQ6UpH3AuYyUw&redir_esc=y#v=onepage&q&f=false

Hidalgo, C. A., & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10570–10575. https://doi.org/10.1073/pnas.0900943106

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. https://doi.org/10.1126/science.1144581

Holland, J. H. (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. https://doi.org/10.1016/S0304-4076(03)00092-7

Inoua, S. (2016). A Simple Measure of Economic Complexity. Retrieved from

http://arxiv.org/abs/1601.05012

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. https://doi.org/10.1111/ecoj.12530

Kauffman, S. (1995). *At Home in the Universe: The Search for Laws of Self-organization and Complexity*. Oxford University Press. Retrieved from https://books.google.it/books?id=o-Owb5IDkSQC

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

Lewbel, A. (2012). Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. *Journal of Business & Economic Statistics*, *30*(1), 67–80. https://doi.org/10.1080/07350015.2012.643126

Loreto, V., Servedio, V. D. P., Strogatz, S. H., & Tria, F. (2016). Dynamics on Expanding Spaces: Modeling the Emergence of Novelties (pp. 59–83). Springer, Cham. https://doi.org/10.1007/978-3-319-24403-7_5

Mealy, P., Farmer, J. D., & Teytelboym, A. (2019). Interpreting economic complexity. *Science Advances*, *5*(1), eaau1705. https://doi.org/10.1126/sciadv.aau1705

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. https://doi.org/10.1080/00130095.2017.1391691

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. https://doi.org/10.1002/jae.951

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. Retrieved from https://arxiv.org/abs/1801.05352

Pugliese, E., Chiarotti, G. L., Zaccaria, A., & Pietronero, L. (2017). Complex Economies Have a Lateral Escape from the Poverty Trap. *PLOS ONE*, *12*(1), e0168540. https://doi.org/10.1371/journal.pone.0168540

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. https://doi.org/10.1080/00343404.2017.1283012

Sbardella, A., Pugliese, E., & Pietronero, L. (2017). Economic development and wage inequality: A complex system analysis. *PLOS ONE*, *12*(9), e0182774. https://doi.org/10.1371/journal.pone.0182774

Sbardella, A., Pugliese, E., Zaccaria, A., & Scaramozzino, P. (2018). The Role of Complex Analysis in Modelling Economic Growth. *Entropy*, *20*(11), 883. https://doi.org/10.3390/e20110883

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA, USA: M.I.T. Press.

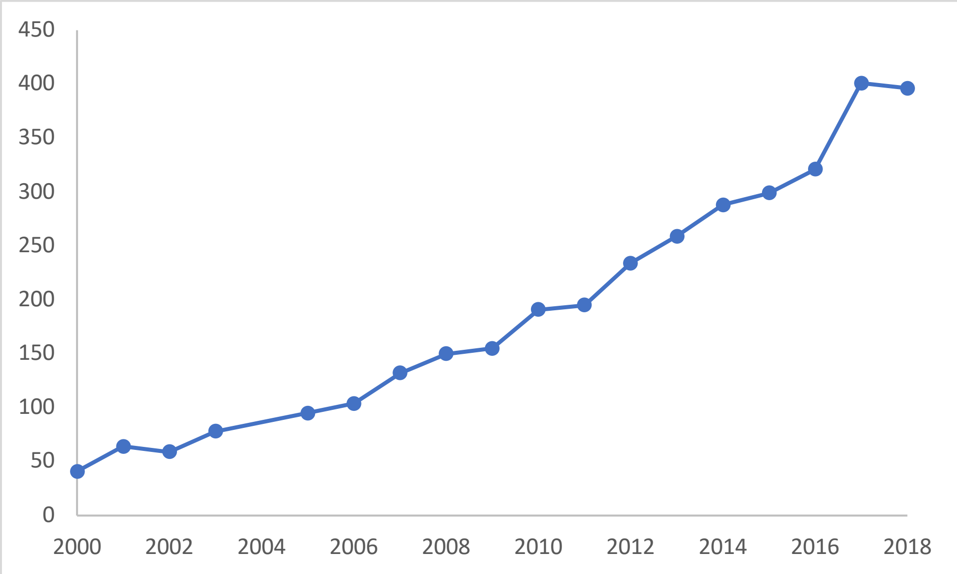
Sweet, C. M., & Eterovic Maggio, D. S. (2015). Do Stronger Intellectual Property Rights Increase Innovation? *World Development*, *66*, 665–677. https://doi.org/10.1016/J.WORLDDEV.2014.08.025

van Dam, A., & Frenken, K. (2019). Variety, Complexity and Economic Development. *Papers in Evolutionary Economic Geography (PEEG)*. Retrieved from https://ideas.repec.org/p/egu/wpaper/1912.html

Zhang, Z., & Glotzer, S. C. (2004). Self-Assembly of Patchy Particles. *Nano Letters*, *4*(8), 1407–1413. https://doi.org/10.1021/nl0493500

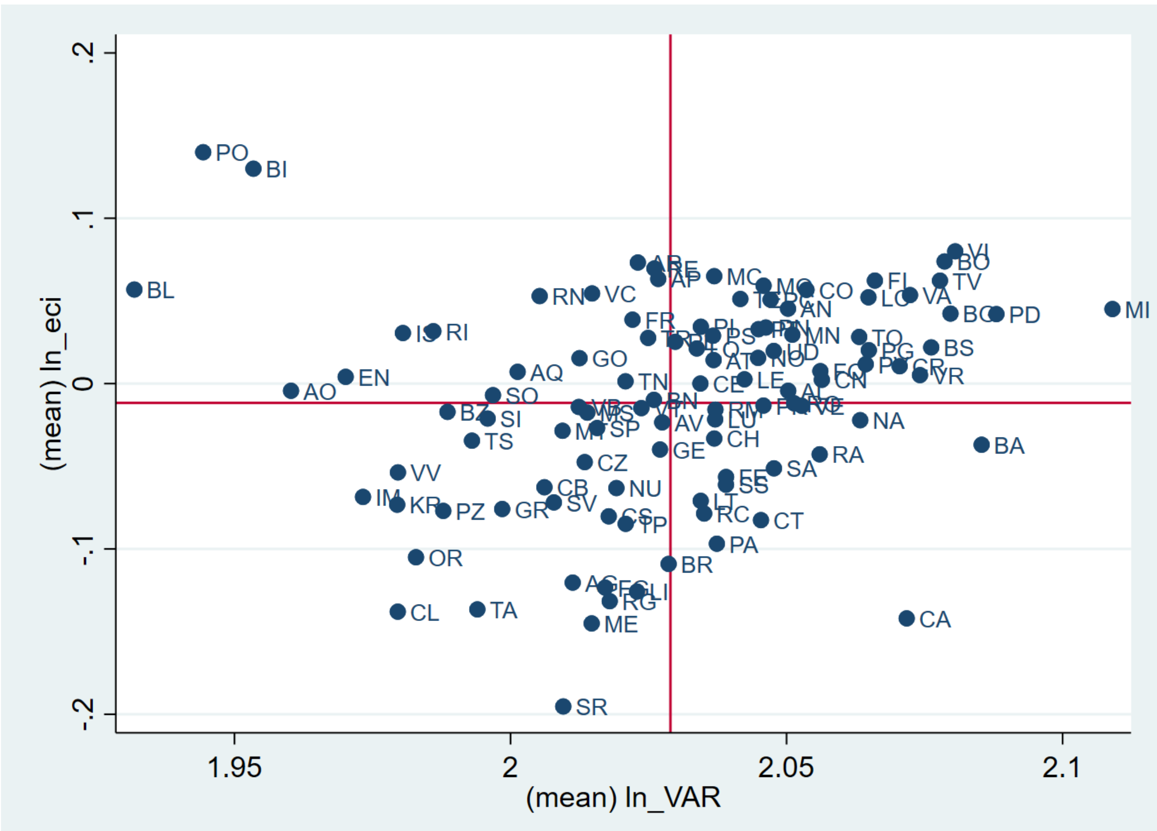
FIGURES AND TABLES

Figure 1. Scientific output on economic complexity in economic disciplines



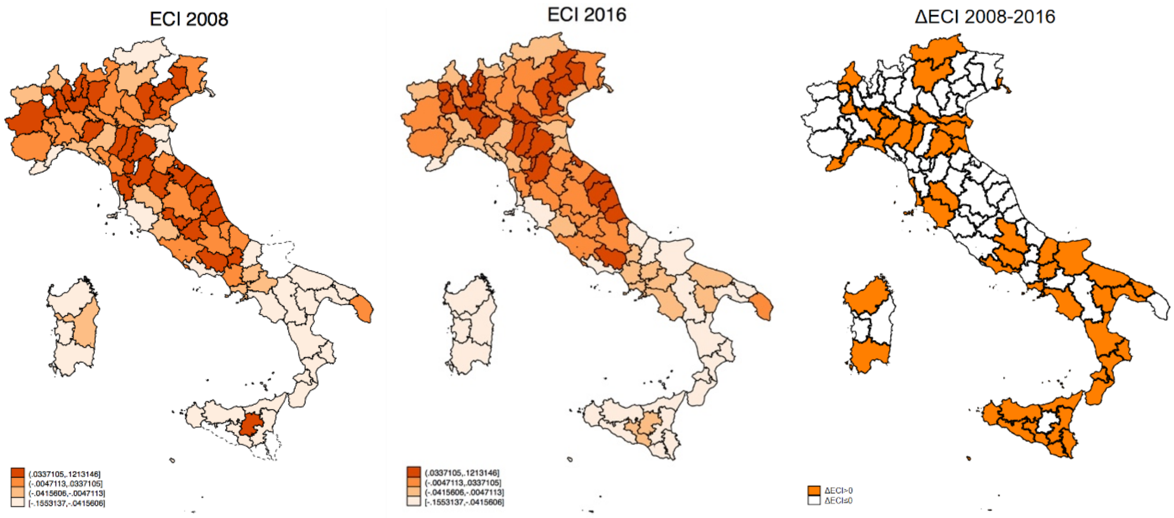
Source: authors' elaborations on Scopus data (latest access: 25 February 2019)

Figure 2. The Italian provinces in the entropy-complexity space



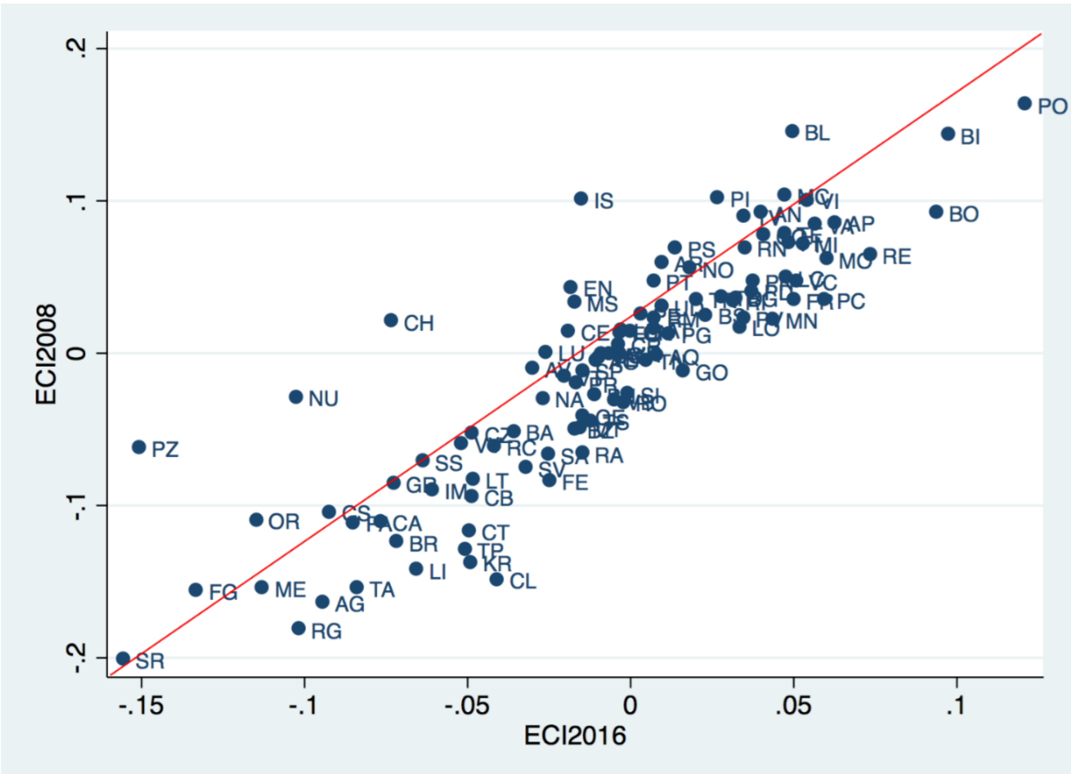
Source: authors' elaborations on ASIA and Coeweb (ISTAT) data

Figure 3. The geography of ECI: 2008, 2016 and mean Δ2008-16



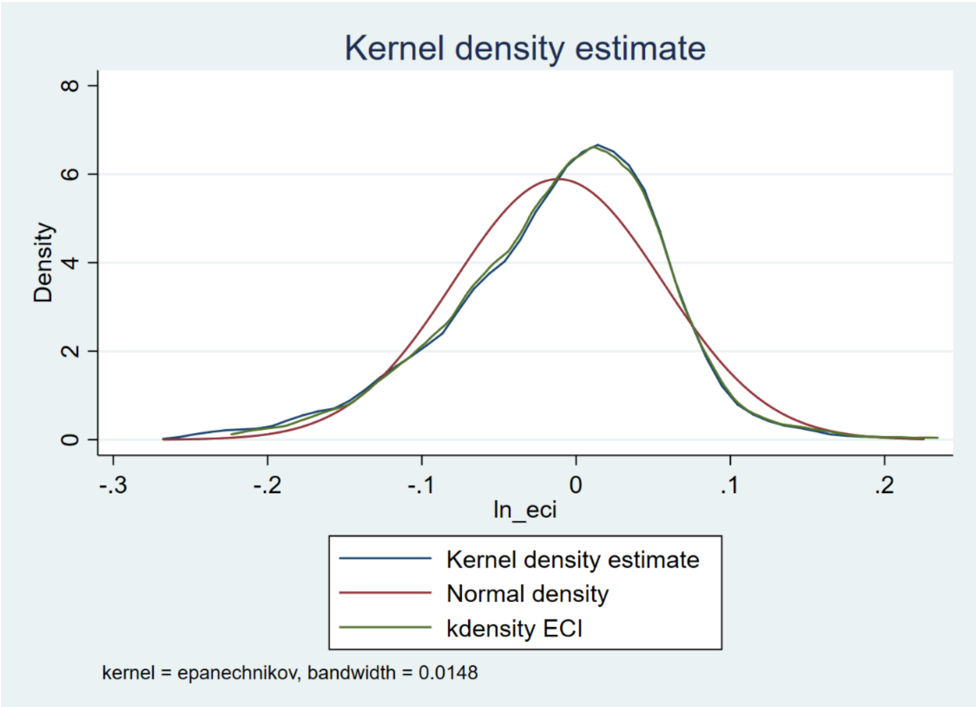
Source: authors' elaborations on Coeweb (ISTAT) data

Figure 4. The relative ECI position of Italian provinces, 2008-16



Source: authors' elaborations on Coeweb (ISTAT) data

Figure 5. Kernel density of ECI and lnECI



Source: authors' elaborations on Coeweb (ISTAT) data

Table 1. Correlations between entropy and ECI

Year	corr (lnVAR, lnECI)
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 variables' description

VARIABLES	Source	Description	Mean	Std. Dev.	Min.	Max.
lnECI	Coeweb (ISTAT)	Economic Complexity Index, based on the 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 of 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)	Number of (bachelor + master) university graduates per 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 on 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***
Pesaran (2007) 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 Pesaran (2007) test 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. The role of entropy on economic complexity: OLS-FD 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 entropy: Lewbel's (2012) IV 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. The role of entropy 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.