European Commission

JRC TECHNICAL REPORT

On R&D sectoral intensities and convergence clubs

JRC Working Papers on Corporate R&D and Innovation No 01/2020

Cattaruzzo, S.

2020

Joint Research Centre

Contact information
Pietro Moncada-释Paternò-Castello
Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: jrc-b3-secretariat@ec.europa.eu
Tel.: +34 954488388
Fax: +34 954488316

EU Science Hub
https://ec.europa.eu/jrc

JRC120036

Seville: European Commission, 2020

How to cite this report: Cattaruzzo, S. (2020), *On R&D sectoral intensities and convergence clubs*, JRC Working Papers on Corporate R&D and Innovation No 01/2020, Joint Research Centre

On R&D sectoral intensities and convergence clubs

Sebastiano Cattaruzzo

Universitat Rovira i Virgili

January 9, 2020

Abstract

Sectoral convergence in R&D intensities among firms is a concept that, although rarely formalized, has been at the center of discussions of industrial and non-industrial actors, such as entrepreneurs, institutions and academics. Far from being a settled issue, the subject has seen very limited empirical attention. We start from the few current evidences, which point to the existence of some β-convergence together with diffused heterogeneity. We recover and integrate the literature from convergence clubs and extend the work introducing the use of Pavitt taxonomy, and new estimation techniques. Particularly, we apply the concept of weak sigma-convergence using a quite novel econometric factor model. Thanks to this, we provide evidences of both σ-convergence for within-sector intensities and of club convergence for across-sector intensities. Finally, the club classification according to "innovative effort" may be used as an alternative way to look at standard economic activities classifications.

The author wishes to thanks Julia Mazzei, Mercedes Teruel, Nicola Grassano, Michele Pezzoni and the participants to the EMAEE2019 hosted by SPRU for useful comments that contributed to the improvement of the paper. Further gratitude goes to the Industrial Research and Innovation unit at JRC-Seville for sharing the data. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 713679 and from the Universitat Rovira i Virgili (URV). The usual disclaimers apply.

1 Introduction

Typically, convergence studies have been carried out within the economic growth framework. This stream of literature has been prolific and largely studied by the academia, and from the 80s through the beginning of 2000, we have assisted to an intense debate on the so-referred "convergence hypothesis". Starting with the absolute convergence studies rejected by Romer (1986), Lucas (1988) and Barro (1991), we then assisted to the debate on conditional convergence, whose main supporters were Barro (1991), Mankiw et al. (1992) and Barro and Sala-i Martin (1995). Finally, the evidences collected by the latter researchers stimulated the formalization, both theoretical and empirical, of the "club convergence hypothesis", see for example Durlauf and Johnson (1995), and Quah (1996).

The idea was that countries sharing similar characteristics in their relevant, structural features (such as preferences, technologies, population growth), but with different initial level of per capita output would on the long-run, converge to some steady-states. The result was then achieved through a variety of assumptions incorporated in models ranging from simple two-sector formalization to more complex one, incorporating plausible elements such as human capital, income distribution, population growth, together with (only) some market imperfection.

Here, we recover the concept of convergence clubs, but we apply it to innovation studies. Coad (2018) carried out an interesting preliminary study on sectoral convergence in R&D intensities for the top-2500 global, private investors in research and development. Although only few evidences of β-convergence were found, we try to amend this research on several aspects. First, we integrate the literature with the consideration of Keith Pavitt's and Joseph Schumpeter's work and contextualize it. Then, given the pitfalls of conventional convergence analysis (Phillips and Sul, 2009), we apply a relatively recent technique to two measures of R&D intensity. Particularly, we look for the emergence of possible convergence clubs in our sample thanks to the non-linear factor model developed in Phillips and Sul (2007a). Then, we compare and contrast the results in light of both theoretical and policy implications.

The relevance of this work is multi-faceted by nature. On the academic side, the study is first of all, an attempt to overcome some of the drawbacks of standard sectoral classification by proposing a club clustering based on each companies' idiosyncratic R&D strategy. Also, it is a fruitful application of both advanced econometric techniques and of a Revised version of Pavitt taxonomy (1984), which is a very powerful systematization of firms in a knowledge economy. Further, our aim is to contribute constantly to the discovery and formalization of empirical regularities in economics of innovation, and some convergence dynamics as the ones emerging here may be good candidates. While, on the policy side, results might be extremely relevant to help designing plans aimed at reaching the ingloriously famous 3% target posed in 2000 with the Lisbon Agenda, and then incorporated in the policies of the Horizon 2020 program (Veugelers and Cincera, 2015).

The paper is structured as follows: Section 2 contains the literature review and the research questions we explore. Section 3 deals with the descriptive analysis of the data. Section 4 describes the methodology, then applied in Section 5; Section 6 concludes.

2 Literature Review

2.1 On R&D investment: determinants, patterns, and intensities

Thanks to economic research and to the advent of endogenous growth models, technical and technological innovations were recognized to have a fundamental role in growth and development (Neef, 1998). Thus, firms and countries have started devoting attention to how this could be stimulated. For this reason, both private businesses and public institutions invest considerable amounts of money and time in searching for these "innovations" (see Table 1).

The accounting item to which these investments refer is namely, research and development. At the beginning of the 1960s, starting from a document by Christopher Freeman and with the idea of standardizing definitions and data on research expenses, the OECD (Organization for Economic Development) has proposed the commonly-named "Frascati manual". Subject to frequent updates the last version was issued in 2015, and we find that "Research and Experimental Development comprise creative and systematic work in order to increase the stock of knowledge - including knowledge of humankind, culture and society - and to devise new applications of available knowledge". Further, the authors establish five core criteria for an activity to be an R&D one: 1) novel, 2) creative, 3) uncertain, 4) systematic, and 5)

Table 1: Global Gross Expenditure on R&D by sector of performance (OECD, 2018)

2010				2015			
		in mln $	%			in mln $	%
Public	Government	179,349.6	14.0%	Public	Government	206,599.2	13.0%
	Higher Education	215,054.7	16.8%		Higher Education	234,468.3	14.7%
Private	Business enterprises	858,047.4	**67.1%**	Private	Business enterprises	1,122,228.3	**70.6%**
	Other	26,586.0	2.1%		Other	26,866.5	1.7%
Total		1,279,037.5	100.0%	Total		1,590,162.3	100.0%

Please note that all amounts are expressed in 2010, millions of US Dollars.

transferable.
These criteria have key implications for studying research and development as an innovative effort. Particularly, the uncertainty of the process in one of its key feature and it frames the expenditure on R&D as an investment (Amoroso et al., 2017, Knight, 1921). Also, we believe that this measure is among the best candidate proxies for measuring what we define as "technological effort", or in broader terms, the intensity of innovative search.

One stylized fact about research and development is the high degree of concentration of its activities and investments. As reported in Table 1, this is especially true for private business enterprises, where very few firms compose the almost entirety of global expenditure in R&D. Following the Frascati manual, R&D activity is mainly measured by two factors: 1) the amount of money dedicated to it (R&D investment, henceforth), and 2) the number of employee dedicated to it (R&D personnel).

Investigating the effect of increasing research and development activities over innovative output is an extremely problematic topic (please refer to Griliches (1979) for the seminal article on the issue). Over the years, this attempt has entailed a huge variety of approaches and methodologies, whose almost none is free of objections. Difficulties in this field arise for several reasons, among these: measurability issues, presence of endogeneity and possible reverse causation, existence of sectoral-specific patterns, different sources of financing (public v. private), different types of research (basic v. applied), etc.

If some of the above issues will be hardly solved with "conventional" techniques, some of the others were deeply investigated and systematized by academics. Of particular interest to this paper is the work conducted on sectoral specificities and whose literature review we dedicated section 2.2. Indeed, despite all the obstacles, a big body of literature (and subsequent debate) has stemmed from investigations on R&D.

For obvious reasons, the first studies on the topic were largely based either on case studies or limited samples. Despite this limitation, researchers pioneered methods and approaches that still have relevance today. Some first attempts to evaluate R&D were made in Malcolm et al. (1959); nevertheless, for the first pioneering works on the topic we had to wait Mansfield and Brandenburg (1966), Scherer (1967), Leonard (1971) and Graboskwi and Baxter (1973). If the former authors were still constrained by data and either attempted to model the process theoretically (Scherer), or unraveled key characteristics such as the high degree of uncertainty (Mansfield), the latter authors gave another twist to the subject. Indeed, both in Leonard (1971) and in Graboskwi and Baxter (1973), key is the inclusion of a time-dimension in their studies, thanks to which some degree of causality was established going from R&D intensity to sales growth and competitors' answers were studied.

Focusing on investment, many academics have tried to unravel the determinants of expenditure on R&D. The seminal study on the topic is Hall and Hayashi (1989), whose authors noted that investment theories where lacking of any consideration of research and development expenses, making up to 20% of the gross investment expenditure of manufacturing firms in the United States. Although integrated by further evidences offered in Hall (1993), these early studies failed to explain thoroughly the conundrum, as they were largely based on Tobin's q theories and thus, including several and delicate assumptions that do not always hold (Domns and Dunne, 1998).

Roughly in the same years, another school of thought, namely evolutionary economists, proposed a different approach based on some key characteristics of the innovation process. Particularly, the intrinsic uncertainty and novelty of the research process that make calculations based on returns on investment rarely true. Consequently, how would profit-motivated agents decide how much resources allocate to this process? According to Dosi (1988), their commitment must be based upon two elements: the perception of opportunities, and the existing incentives (appropriability mechanisms, relative prices, market and broader socio-economic conditions).

Since clearly it is impossible to precisely account for this large degree of uncertainty and routinely assign it some value, firms are likely to work with quite general and event-independent rules, such as spending some given percentages of sales on research and development. Also, occasionally firms may come up with meta-rules that could react to unanticipated shocks on interest or profit rates by cutting specific areas of research (Dosi, 1988). This elaboration by the author is based upon both managerial evidence and findings of Griliches et al. (1986), where the authors find evidences for representing firm-level R&D investment as a "martingale with a relative low variance". One of the few attempts to model this framework theoretically has been carried out in Yildizoglu (2002), who proposed an evolutionary computational model for competing R&D strategies. Indeed, one of the outcomes of the model was exactly that companies

"are not simply randomizing, and the dispersion is decreasing in time"

Thus, this hints at the existence of some convergence mechanisms, likely due to learning and competition mechanisms.

Following this work, few studies have focused exclusively on R&D investment and its dynamic properties at firm-level, with the notable and extremely recent exception of Coad (2018). The study analyzes patterns in R&D investment within sectors using firm-level data and it contains several interesting features. First, the author focuses on R&D intensities[1], rather than pure investment; thus, he incorporates the heuristic approach (Simon, 1956) while also correcting for scale effects. Second, he attempts to incorporate all the existing incentives and the features that are known to influence the perception of opportunities with the aim of isolating the most the "rule of thumb" used by the firms. Particularly, the author looks whether there exists some type of convergence at sectoral level, as the possible outcome of a combination of sectoral-specific factors and rivalrous behaviors by the firms (Graboskwi and Baxter, 1973).

From a more general (and aggregate) standpoint, there exist much more work and results. In general, countries seem to have a tendency both toward specialization (Archibugi and Pianta, 1994, Patel and Pavitt, 1994) and convergence in R&D intensity (Archibugi and Pianta, 1994). This has implications in the sectoral composition and impacts directly the path toward the 3% target. Indeed, as noted by Moncada-Paternò-Castello et al. (2011), the trend of research and development intensity in the European Union have been largely flat between 2002 and 2008, and timidly increasing for the years starting from 2010 (Veugelers and Cincera, 2015). However, the difference between the European Union and its competitors lies exactly on the sectoral composition of its industries, which contain much less firm in high-tech sectors (Moncada-Paternò-Castello et al., 2016).

As explained above, alleged convergence results could easily be an artifact of aggregation and also justified by the peculiar period of study. Similarly, a mix of sectoral-specific factors and features of the innovative process may result in some type of convergence also at sectoral level. As noted in Montobbio (2003), at sectoral level, R&D intensities are quite stable across countries and through time, while it is possible to identify some of the sources of variations. Among these, it appears that the variance between sectors composes more than two-thirds of the total variance, thus making sectoral differences much stronger than those among countries. We believe that the remaining one third can be largely captured by the previously explained factors.

All of the above considered, the analysis of highly disaggregated data, at firm-level, is almost fundamental to identify sector-specific trends, which are likely to exist and emerge due to their implicit characteristics.

[1] R&D intensity shall be regarded as one possible formalization of the intuition by Dosi (1988). Indeed, defining intensity, we pose investment in R&D at the numerator of a ratio, whose denominator is net sales, thus obtaining *some* quantity proportional to the volume of firm sales. Interestingly, the first application of this measure is likely to be in Smith and Creamer (1968), then recovered by Leonard (1971) and Graboskwi and Baxter (1973), as a way to eliminate scale effects.

2.2 Why considering also other classifications of economic sectors

As mentioned above, besides the rivalrous behavior among firms in the same sector, likely we can theorize the presence of specific sectoral features, especially with regard to technical change and innovative behavior. However, these sectors may not straightforwardly reflect the standard industrial classifications used. Indeed, as noted by Pavitt (1982) and Coad (2018) (among others), dealing with large R&D investors implies also dealing with large conglomerates that may well operate in different markets. This is the main reason for us to apply a complementary classification of industries, which is possibly more focused on innovative activity.

In 1984, Pavitt proposed a pioneering study of industries in which he realizes that sectors differ, and can be classified, according to the relative importance given to the four basic dimension of technical advancement: 1) economically expensive and formalized processes of search, 2) informal processes of information diffusion and of technological capabilities, 3) phenomena of learning-by-doing or learning-by-using, and 4) the possible adoption of innovations from other industries.

The multifaceted nature of the above dimensions clearly makes the thorough understanding of this interaction hard to grasp. In order to deal with this puzzle, a large literature, both theoretical and empirical, emerged. Among these contributions, some focused on the systems of innovation at sectoral level (Breschi and Malerba, 1997, Malerba, 2002, 2005), some focused on the intrinsic characteristics of sectors according to their knowledge approach (Dosi, 1988, Pavitt, 1984), while some others on capturing specific relations controlling for sectoral specificities (Bottazzi and Secchi, 2003, Castellani et al., 2018, Evangelista, 2000, Montobbio, 2003). Also, Peneder (2010) proposed a possible way of classifying firms according to their innovation strategies and related technological regime, applying cluster analysis to cross-sectional data.

Concerning the process of knowledge creation and diffusion, besides the determinants we already mentioned in section 2.1, the literature showed how features such as tacitness, cumulativeness, and appropriability are central in its characterization (Dosi, 1988, Dosi et al., 2008, Nelson and Winter, 1982). Particularly, the existence and nature of sector-specific, appropriability mechanisms (e.g. patents, secrecy, lead-time) inevitably affect the amount that firms invest in R&D. As described in Breschi and Malerba (1997), these mechanisms characterize industrial sectors that can be ranked according to their relative appropriability conditions. In the same line, also factors such as tacitness and cumulativeness tend to be very sector-specific, both in nature and in their influence on the firms.

With purpose, all the concepts above have been reported without properly defining industrial sectors, but with the aim of demonstrating the existence of several, various, sector-specific features that interact in the moment of setting an R&D intensity level by a managerial board. Our belief is that this framework does allow for a successful use of an appropriate taxonomy to classify industrial sectors according to their innovative (or technological) effort.

The first attempts to do so date back to Scherer (1982). There, the author, thanks to inter-temporal data on U.S. firms, proposes a very first mapping of industrial sectors according to their technological inputs and outputs. The resulting matrix had the aim of explaining the main technological flows undergoing in the U.S. manufacturing and services industry. Although preliminary, this study has been pioneering in the understanding and characterization of industrial sector according to their technological contents.

2.2.1 On Typologies and Taxonomies, Pavitt (1984) and Schumpeter (1911, 1942) as reconciling tools

At this point, it should be clear that despite the ubiquitous micro-economic heterogeneity (Dosi et al., 2010), a variance-reducing tool that have proven to be significant is the use of sectors to study industries. Using an appropriate categorization of an economy is an extremely tricky concept; however, it requires a deep knowledge of what choosing a particular categorization over one other may imply.

The main fork in terms of possible classification methods consists of focusing either on product portfolios or on production processes, with the latter being much more frequently used than the former. Typical industry classifications (e.g. NACE, NAICS, SIC, ICB) are ways of categorizing economic activities according to the underlying production processes.

Although fruitful under many aspects, these classifications tend to, at least partially, overlook most of the aspects linked with the innovative processes. Indeed, one of the reasons why innovation is at the very core of economic growth is also its intrinsically complex and linked nature. By this, we mean that seldom innovations are developed in an industrial sector and there they stay; the intuition formalized by Pavitt (1984) and his colleagues at SPRU was that inter-sectoral flows were not only fundamental to economic activity, but they were also shaping it.

Quoting Pavitt (1984), the aim of his taxonomy was to"describe and try to explain similarities and differences amongst sectors in the sources, nature and impact of innovations, defines by the sources of knowledge inputs, by the size and principal lines of activity of innovating firms, and by the sectors' of innovations' production and main use".

When it comes to describing what the Pavitt taxonomy is we shall start from its etymological roots. As well pointed out by Archibugi (2001), taxonomies differ from categories in a key dimension: the formers are intended to classify phenomena with the objective of maximizing difference among groups. Thus, they are considered useful "if they are able to reduce the complexity of the population studied into easily recallable macro-classes".

But, how does this taxonomy work in practice? Differently from production processes classification, the idea behind this taxonomy is to trace the product portfolio of firms, and by understanding the complex interplay of relations among them, among and inside the firms, to create a number of groups that approach innovation in the most similar way. Firstly conducted in the 1980s by Keith Pavitt using survey data from UK manufacturing firms, this exercise led to the creation of 4 basic categories, to which he then added a fifth one. Unfortunately, a direct application of this approach would require an *ad hoc* database as for Pavitt (1984), so in this empirical exercise we will work with what we define a "revised Pavitt taxonomy", which is applied indirectly to the 4-digit ICB sectors, present in our database.

The categories are designed as follows: 1) supplier-dominated firms, 2) specialized suppliers of capital goods and equipments, 3) scale-intensive firms active in mass production industries, and 4) science based firms, which tend to be extremely R&D intensive as their *locus* of production of knowledge is often in-house laboratories.
Subsequently, the author realized the emergence of information-intensive firms, which in sectors such as tourism, banking and retailing were accumulating technology by advanced systems of data-processing.

Although initially disregarded by Pavitt because of its low relative weight in the economy, the service sector turned out to be another important origin of knowledge and innovation. For this reason, with the advent of globalization that made services easily tradable, Miozzo and Soete (2001) proposed a re-assessment of Pavitt taxonomy that include also services. Previously, Pavitt uniformly assigned "services" to the supplier-dominated firms category. On the contrary, several authors have tried to amend this shortage. For example, Miozzo and Soete created customized (but sometimes overlapping) categories for services only, consisting of: 1) supplier-dominated firms, 2) scale-intensive physical networks and information networks sector, and 3) science-based and specialized suppliers sectors.
Following these works, the most recent appraisal of the taxonomies can be found in Bogliacino and Pianta (2016). The two authors run a battery of empirical tests on what they define as a "revised Pavitt taxonomy", taking into account all the proposal and amendments the original formulation went through in time. Eventually, they find that the taxonomy performs very well and that concerning services, it would be recommendable to include them in the scale intensive sector, as they are very statistically similar.

Other classifications based on the product portfolio have been proposed and used in the recent years. One peculiar and compelling example of this is the so-called Text-based Network Industry Classification (TNIC) proposed in Hoberg and Phillips (2016). The focus of this classification was on competition in the product market, but the concept is very similar as they try to group firms according to some similarity measure in some given dimension. Another notable example of purely data-driven classification applied to the technological field can be found in Gkotsis et al. (2018).

Nevertheless, in this work, we apply a revised version of Pavitt taxonomy as the core of our exercise regards technological effort, or R&D intensity. The reasons to do so are obviously multiple, first we do not embrace purely data-driven approaches when explanations are needed. Further, as appraised in Bogliacino and Pianta (2016), this taxonomy performs very well in identifying "(a) levels and types of innovative efforts, (b) proximity in a multi-dimensional technological domain, (c) determinants of innovative performances and (d) the relationship between innovative and economic performance". Useless to say, we are interested in point a.

Further, as a complementary view, we will also try to connect our findings to the "technological regimes" individuated by Schumpeter (1911, 1942), leading to the distinction between a "Schumpeter

Mark I" and a "Schumpeter Mark II" regime[2]. The first regime has the following distinguishing features: relatively low accumulativeness of knowledge, lower innovative opportunities and innovations carried out often by new entrants. On the contrary, Mark II is characterized by more cumulative innovative activities and higher opportunity, making often a few incumbents "serial innovators".

2.3 On Convergence Clubs in Economics and the Pitfalls of (some) Convergence Approach

The last missing ingredients of this paper is the idea of "convergence clubs". After spending at least a couple of decades in debating the "convergence hypothesis" in macroeconomics, economists seem to have achieved quite an agreement in the formalization of "convergence clubs". Two seminal papers in this sense are Durlauf and Johnson (1995) and Quah (1996).

The two papers are interesting exercises because allowing for the possibility of multiple regimes in the common linear model, they found a way to achieve multiple steady-states that could mimic more realistically the behavior of countries in their path of economic growth. The underlying idea was that instead of looking forcedly for an hypothetical absolute convergence, it made more sense to allow for multiple regimes and analyze whether units achieved any type of convergence within.

Following progresses in the econometric field that allowed for more flexible estimation techniques (Bianchi (1997), for example), the concept of convergence club gained importance and empirical credibility.

In an attempt of giving explanation for the presence of these "country clubs" in the paths of economic growth, academics from the innovation field proposed an explanation based on technology clubs (Castellacci, 2008, Castellacci and Archibugi, 2008). In these works, countries can be grouped according to their different levels of absorptive capacity and innovative ability. The empirical exercises obtained great matching with the data that posed this explanation as a very plausible one to explain different growth path undertook by cities, countries or unions in history (see for a similar line of explanation Dosi and Tranchero (2018)).

It is our intention to argue that as technology clubs have been proven to exist as an empirically testable concept, the same can be done to firms, particularly to the ones occupying the very top spots in terms of global R&D intensity. Particularly, this can achieve even more explanatory power when considering intensity in research and development spending as a possible measure of technological effort.

Also, as highlighted in Phillips and Sul (2009), conventional approaches to β-convergence imply a key assumption that is very rarely satisfied in real world: homogeneity of technological progress. Meaning that the typical studies of β-convergence provide unbiased estimates if and only if, we assume 1) that the speed of convergence is homogeneous across observations (transitional homogeneity), and 2) that there are no omitted variables or endogeneity concerns. More realistically, those processes are fairly heterogeneous and occur with different speeds of convergence. It is with this in mind that we develop our econometric strategy.

3 Data

For the sake of this exercise, we use a peculiar dataset, whose representativeness must be taken into consideration when interpreting the results. The starting dataset is the "EU Industrial R&D Investment Scoreboard"[3] that contains information regarding the top 2500 global, private R&D investors that all together make up for roughly 90% of world private expenditure in research and development (Hernández et al., 2018). The data is shaped as a panel, following firms' activities for 18 years (2000-2017). Other than R&D investment, the data contains information regarding net sales, number of employees, operating profit and ICB sector at the 3-digit level.

However, we decide to sub-set this data and work with only incumbent firms, which contribute for only 35-40% to the global private expenditure in R&D. Although potentially reducing the scope of this work, we decide to undertake this decision for several reasons.

[2]Actually, the labels Mark I and Mark II have been originally introduced by Nelson and Winter (1982) to offer a more synthetic representation the theoretical models presented by J. Schumpeter and characterized by forces of creative destruction/accumulation.

[3]Available at https://iri.jrc.ec.europa.eu/rd_monitoring

First, we believe that entry/exit dynamics in the strongly unbalanced original version of the data may significantly bias the estimates. Particularly, the vast majority of the observations in the unbalanced version are only for one or two years. Second, as noted in Graboskwi and Baxter (1973), the very top R&D investors show much more responsiveness to each other, if compared to non-top ones. This is crucial for the exercise as we believe that competition and imitation are at the basis of possible convergence mechanisms. In Figure 1, we plot the kernel density functions for the different versions of our sample, which shows the largely skewed nature of this distribution and the fact that incumbents seem to contribute if not more, at least equally to the volume of expenses. Finally, the applied technique performs significantly better with a balanced sample.

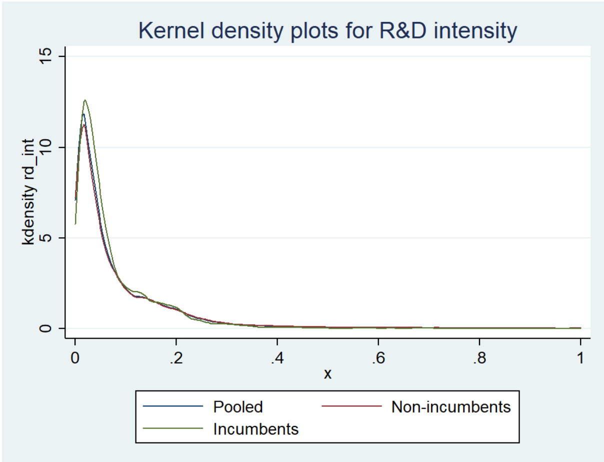

Figure 1: Balanced v. unbalanced sample comparison

An empirical work with similar data was conducted by Patel and Pavitt (1997) and it led to the confirmation of two key characteristics of large, leading innovating firms: technical complexity and hysteresis. Also, the same study left unexplained a considerable portion of variance that the authors noted in large firms' innovation activities. With this study, our aim is to follow this path and attempt to systematize the different approaches that leading, global R&D firms undertake to face complexity and uncertainty.

4 Methodology

In this paper, we apply the factor model developed in Phillips and Sul (2007a). The approach not only allows for the estimation of possible convergence in the data, but also it offers techniques for the estimation of clusters (or convergence clubs) from the data. Being purely data-driven, the factor model is quite free of imposing assumptions, as it can accommodate very different, also nonlinear functional forms without losing the needed flexibility in modeling firms' idiosyncratic behavior over time and across section.

With the aim of appraising both within-sector and across-sector convergence, we use two different dependent variables: R&D intensity, defined as the simple ratio of R&D to sales, and the R&D intensity gap, defined as each firm's distance from the mean sectoral intensity. Thus, we define the following for a firm i operating in sector j at time t:

$$RD\_intensity_{i,t} = RD\_expenses_{i,t}/Sales_{i,t} \tag{1}$$

$$RD\_intensity\_gap_{i,t} = RD\_intensity_{i,t} - RD\_intensity_{j,t} \tag{2}$$

In order to provide convergence estimates for equations 2 and 1, we use the method for time-series data developed in Phillips and Sul (2007a). The model, formulated as a nonlinear time-varying factor model, has seen several applications in macro-studies of convergence (Bartkowska and Riedl, 2012, Blanco and Presno, 2019, Borsi and Metiu, 2014, Camarero et al., 2013, Monfort et al., 2013, Panopoulou and Pantelidis, 2009, Phillips and Sul, 2007b), but much less in micro-panels. Nevertheless, especially when applied to micro-data, the approach provides a great deal of flexibility in modeling the idiosyncratic behavior over time and across sections, while also allowing for a common, unknown growth component (Phillips and Sul, 2007a). Further, the model is not only tailored to estimate possible sigma-convergences, but it is also able to accommodate the estimation of convergence groups (or clubs) starting from panel data.

Start from considering the standard expression for a simple, single factor model:

$$X_{it} = \delta_i \mu_t + \epsilon_{i,t} \tag{3}$$

where δ_i is the firms' individual distance between some common factor μ_t and the systematic part of X_{it}, while $\epsilon_{i,t}$ is the error term. The amendments coming from Phillips and Sul (2007a) are mainly two.

First, they allow the term δ_i to be time-varying, so $\delta_{i,t}$. Then, they incorporate the random component of the error term in $\delta_{i,t}$. The final representation of this new, factor model is

$$X_{it} = \delta_{i,t}\mu_t \tag{4}$$

The above representation is very easily interpretable as in a traditional decomposition of panel data, we would have two components: a common trend component given by μ_t and then, $\delta_{i,t}$, which measures the relative share of μ_t for firm i at time t. These amendments are very relevant in allowing for a diffused heterogeneity in firms' dynamics of research and development investment, while also dealing with the pitfalls of the typical convergence approaches (Phillips and Sul, 2009).

Particularly, this estimation procedure is applied with the aim of analyzing each firm's transition path in R&D intensities, together with its associated growth. The transition path corresponds to the measurement of the relative share of firm i R&D intensity in total intensities. In other words, we have that $h_{it} = RD\_intensity_{i,t}/\overline{RD\_intensity_t}$, where $\overline{RD\_intensity_t}$ refers to the cross-sectional average of R&D intensity for the firms under consideration. Thus, Phillips and Sul (2007a) show that the quantity h_{it} eliminates the common growth components to the first order by providing a measurement of each individual firm's share in common R&D component growth. Additionally, being h_{it} time-variant, it also follows the evolution of this share as the time passes.

The technique improves significantly the potential of convergence studies by focusing on an individual transition parameter, h_{it}, which is functional of $\delta_{i,t}$, and it represents the transition path for individual i in relation to the other individuals in the panel. Then, the convergence test is built upon the time series linear regression of a cross section variance ratio of the individual h_{it} on $log(t)$. In the next sub-section, starting from this relatively simple concept, we explain the complete procedure proposed by Phillips and Sul (2007a).

4.1 Procedure

As mentioned above, the procedure we apply comes from Phillips and Sul (2007a) and its application is carried out following the statistical packaged developed in Du (2017). The approach implies first a regression test on the whole sample with the aim of appraising possible evidences of convergence in the data. This is what the authors refer as *log-t regression test*, or relative convergence, and it implies testing the following condition:

$$\lim_{t\to\infty} \frac{X_{it}}{X_{jt}} = 1, \text{for all } i \text{ and } j \tag{5}$$

With the aim of appraising the following condition, Phillips and Sul (2007a) proposed an approach based on a non-linear, time-varying factor model that is not assumptions demanding and accommodates well the heterogeneous behavior of the units of observation.

Nevertheless, instead of testing it directly on the time-series, first the trend component is extracted from it, and the test is run on the idiosyncratic part of the series. With this specific aim, the method proposes to apply a version of the well-known Hodrick-Prescott filter (Hodrick and Prescott, 1997, Whittaker, 1923), tailored to account for the individual heterogeneity that panel data contain, differently from simple time-series contexts. Also this method is demonstrated to be free of imposing distributional assumptions and performs well even when the time span is short (Phillips and Sul, 2007a).

At this point, it is already possible to conclude whether there are evidences of convergence for the whole sample or not. If the answer is negative, the method continues by looking for convergence in subgroups of the data. This data-driven part is very conveniently framed to avoid the formation of a-priori groups and instead, making the grouping a matter for direct empirical determination.

The sub-procedure for club convergence grouping starts with a cross-sectional sorting of the panel according to the value of last period. This is because usual evidence of convergence obviously are most apparent in the final time series observations. Then, the data is initially clustered according to these values, creating sub-groups of size k, $G_k = \{1, ..., k\}$ for $\{k = 2, ..., N\}$.

The following step involves re-running the log-t test, whose condition is in equation 5, on the single sub-groups, and thus, obtaining the t_k statistics relative to G_k. Then, the algorithm choose k^* to maximize t_k over all values for which $t_k > c$ for $k = 2, ..., N$ and where c is some critical value. In our case, we pick c to be -1.65, corresponding to the 5-percent significance level for the one-sided t-test involved. If the inequality test is not satisfied, the highest j individuals from the group will be dropped and will be part of the formation of a new sub-group. This continues until the creation of a divergent sub-group is the only option, which works as a stopping rule. Once this initial classification is done, the method continues iteratively by adding individual observations from adjacent groups with the purpose of reducing the number of clubs identified in the first part, and using as stopping rule the condition $t_k > c$.

So, to recap schematically, the procedure works as follows:

1. log-t regression test for whole s> -1.65, stop here and conclude there are evidences of σ-convergence for the sample
2. otherwise, sort the observation by last period values
3. form core groups maximizing the value of the test statistic
4. try merging groups, by adding individuals to the adjacent clubs, following $t_k > c$
5. stop when the remaining individuals form a divergent group, or all observations have been assigned to groups

5 Results

Our result section is divided into two parts: a first part, which focuses on within sector convergence, while a second one, which deals with across sectors convergence. On the one hand, we find evidences of relative convergence using the research and development intensity gap indicator, we do not find the same for the whole sample without controlling for sectoral averages. However, for the latter, we run the procedure for club detection and we are able to identify a total of 4 R&D intensity clubs, for each of which some clear and identifiable patterns emerge.

5.1 Within sectors

As we anticipated, in this part of the exercise we use the measure of innovative activity outlined in equation 2, thus we control for sectoral averages using the information available at both the 3-digit and 2-digit ICB sector. In this analysis, we apply the log-t regression test to the whole sample of incumbent R&D global investors and indeed, we find evidences of relative convergence.

Table 2: Log-t regression test results on the whole, balanced sample for R&D intensity industry gap

Variable	Coefficient	SE	T-statistic
R&D intensity gap 3-digit	2.6686	0.7950	3.3567
R&D intensity gap 2-digit	-1.0253	1.2782	-0.8022

Number of observations: 472. Number of time periods: 18.

Being the hypothesis of relative convergence accepted when using the R&D intensity gap measure under both sectoral aggregations, there is no need of seeking any additional club classification. This is an interesting finding because although with different techniques, in the study that motivated this in-depth analysis (Coad, 2018), there were no evidences of it.

5.2 Across sectors

On the contrary, when we run the same test that resulted in a significant coefficient in Table 2 on the "pure" measure of R&D intensity, we face a significant rejection of the convergence hypothesis. As reported in Table 3, the estimated coefficient for our series is negative and not significant. Nevertheless, we had no reasons for expecting a general convergence of R&D intensities of all sectors.

As a consequence, following the approach developed by Phillips and Sul (2007a), we proceed to the estimation of possible club convergences. The first core group classification results in 10 clubs.[4]

[4]For the sake of conciseness, we do not report the initial classification here, but interested readers can find it in the Appendix.

Table 3: Log-t regression test results on the whole, balanced sample for R&D intensity

Variable	Coefficient	SE	T-statistic
R&D intensity	-1.073	0.039	-27.889

Number of observations: 472. Number of time periods: 18.

Then, continuing with the recursive procedure aimed at merging possibly close groups and explained in Section 4.1, we obtain a more compact representation of the long-term R&D dynamics undergoing among the largest investors.

Table 4: Log-t regression test results on the final club classification for the whole, balanced sample using R&D intensity

R&D int.	Club 1	Club 2	Club 3	Club 4
Coefficient	0.033	-0.106	0.687	0.476
T-statistic	0.953	-1.052	8.044	2.712

Number of observations: 472. Number of time periods: 18.

Thanks to the results reported in Table 4, we are able to study the characteristics of the firms that belong to each of the identified clubs and to see if there exists any identifiable characteristic. Indeed, what we find by complementing the basic accounting items with the information coming from the two industrial classification that we use (ICB and Revised Pavitt) is a quite clear picture of how the firms under consideration behave in the competitive market for innovation.

In particular, foregoing momentarily the standard industry classification we apply, thanks to this exercise we are able to identify five, clearly distinct and delineated clubs of R&D intensities, according to the individual transition paths that each firm followed. Table 5 contains the descriptive statistics relative to the final year of observation for each of the identified convergence clubs, while Figure 2 reports the time dynamics relative to each club. Indeed, as expected, the results can be systematized and interpreted, quite straightforwardly.

Below, with the aim of providing a general description, we explore each club under both the accounting figures and main sector of operation (using both ICB and Revised Pavitt):

1. **R&D specialists.** This club is made up of just 11 enterprises, very sparse across industrial sectors, but with approx. the 80% being either science-based or knowledge-intensive business service, according to our Revised Pavitt taxonomy. Their average R&D intensity is very high, but driven upward by a couple of outliers (resulting in very high variance and a considerably smaller median). Finally, these firms tend to be small in both employment and sales.

2. **STEMmers.** The second club is also the most numerous group and firms mainly belong to the STEM fields[5], with particular concentration on chemicals, engineering, IT, and pharmaceuticals. Firms here invest in R&D an amount which ranges from 5 to 10% of their sales, although the data still with considerable variance. Also, they tend to be quite big in sales and in employment, particularly.

3. **Good, old manufacturers.** This group is the second-largest in the sample and it has a compelling sectoral composition under both our classification. Indeed, if under ICB categories, it is mainly composed of industrial machineries suppliers and basic chemical production; using the Revised Pavitt taxonomy, we find that 82 out of 93 firms belong to either scale-intensive, specialized-supplier or supplier-dominated industries. These companies apply a moderate intensity in terms of R&D ($\approx 2\%$) and also tend to have a modest volume of turnover.

4. **Scale&Energy.** Half of this group is composed by companies that are either in construction, basic materials, oil or electricity. Using other categories, almost half of the companies in this

[5]STEM is an acronym that stands for Science, Technology, Engineering and Mathematics.

group operates in scale-intensive industries, while a considerable, remaining group works in physical networks, or as specialized suppliers. Here, firms are not very R&D intensive (around 0.6%) , but at the same time, they are characterized by considerably high levels of turnover and employment.

Table 5: Descriptive statistics for each final club

Final club	R&D intensity	R&D investment	Net Sales	Employment	Operating Profit	
1 -	1.1	852	3,505	8,882	-.013	mean
R&D Specialists	.24	385	1,732	8,500	.15	median
	2.5	1,443	5,208	8,001	.47	sd
	11	11	11	11	11	n
2 -	.092	1,221	17,123	51,284	.13	mean
STEMmers	.065	388	6,224	20,800	.11	median
	.074	2,123	31,090	77,867	.13	sd
	317	317	317	313	316	n
3 -	.020	273.36	13,798.72	51,296.91	.107	mean
Good, old manufacturers	.019	159.79	8,228.5	29,853	.11	median
	.005	357.09	16,167.55	67,450.69	.085	sd
	74	74	74	74	74	n
4 -	.007	198.52	38,732.26	73161.56	.105	mean
Scale&Energy	.006	92.15	18,744.05	32,730.5	.083	median
	.004	270.14	52,168.66	111,000	.088	sd
	70	70	70	68	70	n

The statistics presented refer to the final year of the sample: 2017. All monetary values are expressed in thousand Euros at the end of year 2017.

After the estimation and description of these four convergence clubs on the basis of each companies' innovative effort, we try to match some of these findings with empirical regularities in the sectoral composition of innovation. Particularly, foregoing the first group of "R&D Specialists" because of their limited number and across-sector dispersion, we try to link the characteristics of each group with the relative technological regimes elaborated by Schumpeter (1911, 1942) and also explained in Breschi and Malerba (1997).

Indeed, following the findings by Breschi et al. (2000), who studied the characteristics of economic sectors trying to link them to the Mark I/Mark II distinction, we analyze how the clubs we identified could fit into either one of the two regimes. Although incomplete due to the lack of entry/exit dynamics, this mental exercise is carried out by linking the ICB sectoral distribution of firms in each club (Table 17) to the empirical findings in Breschi et al. (2000) with the aim of appraising the representativeness of our proposed, classification.

Following the above mentioned study, we would have that "Good, old manufacturers" fall into Mark I regime, due to the large presence of producers of industrial machineries, food and household goods. Indeed, this club show a quite moderate R&D intensity that could be a sign of relatively low technological opportunities. On the contrary, "STEMmers" and "Scale&Energy" producers would fall into a Mark II regime. This outcome would be due the presence of oil and gas producers, chemical and pharmaceutical companies, consumer electronics suppliers and automobiles and parts makers. Because of the bigger technological opportunity, these sectors should exhibit more R&D intensive firms, operating in a regime

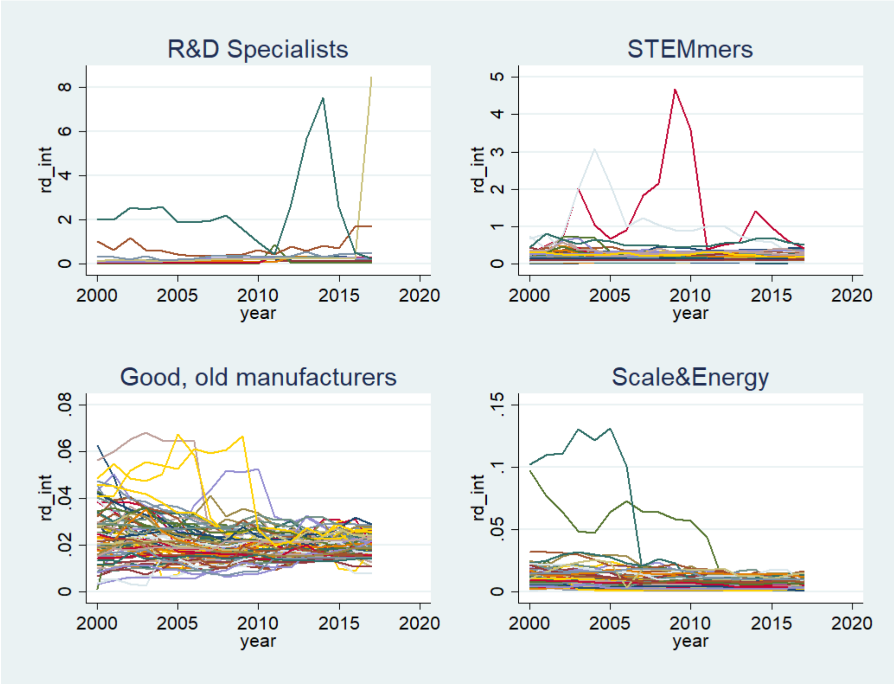

Figure 2: Time dynamics relative to each of the identified groups[6]

where cumulativeness is extremely important. Finally, for "STEMmers" we can interpret a more exploratory approach[7], aimed at maintaining their leading position in the market for technology. On the contrary, it is supposable that companies involved in "Scale & Energy" seek more exploitative strategies thanks to the position they achieved in their respective sectors. This is a possible justification for the lower R&D intensity.

5.3 Robustness checks and sensitivity analysis

Considering the high degree of novelty of the employed technique and the lack of previous applications to micro-studies, we propose a battery of robustness checks aimed at exploring the sensitivity of the convergence approach under consideration. Unfortunately, the exploration is limited by the need of a balanced sample for estimation, but we still propose considerably different sub-samples at test. Specifically, we repeat the estimations, both within and across sectors, for the following three sub-samples of the original 18-year data: 1) a 17-year sub-sample, which the last year of the original sample eliminated, 2) a 90% random sub-sample of the original one, and 3) an 80% random sub-sample of the original one.

Regarding within sectors convergence, we can see the results summarized in Table 11. Out of the six, additional estimations we see that 5 confirms strictly the evidence of within sector convergence in R&D intensities. Also, the "conflicting" result offers a t-statistic of -1.76, which is extremely close to the assumed threshold of -1.65, corresponding to the 5-percent significance level for the underlying, one-sided t-test.

When looking at across sectors convergence, the aim is to discover whether changes to the composition of our sample affect consistently the associated estimations and the following club identification. Presenting each result, we compare it to the original classification both in terms of allocation by Pavitt sector and of firms' characteristics.

[6]From Figure 2, despite the presence of some outliers, the reduction in dispersion is quite clear. One big potential of this dataset is that we have the name of each company; thus, we can control for unexpected behaviors and understand their origins. For instance, in the case of "R&D Specialists", two cases leap out: 1) Exelixis, a pharma company, which bet everything on a drug for prostate cancer, but after failing phase-3 development, was forced to let go 70% of its workforce in 2014, and 2) Altaba, an IT company, which closed its activities in the last year of sample, and having almost zero sales, the reported R&D intensity peaked.

[7]For a thorough explanation of the exploratory v. exploitative approaches in technological regimes, please refer to Colombelli et al. (2013)

With the first sub-sample, composed of all the original firms without the last year of observation (2017), we obtain still a 4-club classification with very similar firms' characteristics (see Table 12) and a distribution by Pavitt sector on the very same line (Table 8). Results start to vary slightly more when starting to drop firms from the sample.

For the 90%-subsample, the classification shrinks to a 3-club grouping. In Tables 9 and 13, we can see how the change in the sample composition led to a slightly different classification. Indeed, what happens is that firms previously in the "STEMmers" club mix with the ones composing the backbone of the manufacturing system. Finally, with the 80%-subsample the obtained grouping is a 5-club one. This is mainly due to two outliers: one is Exelis, of which we talk in Footnote 6, while the other is one Morphosys, which is a relatively small and very high R%D intensity company.

Overall, we can judge the outcome of these checks quite positive, as the results showed to be largely stable for the within-sector convergence estimations. Also, the club classification approach showed consistent results even when changing the sample composition. A further confirmation of this, which also summarizes largely our robustness checks, consists of the correlation results that can be found in table 15. As estimated, the club allocation due to the different samples correlates quite robustly at firm-level, with values ranging from 0.73 to 0.88.

6 Concluding remarks

Despite the delicacy in terms of external validity for these findings, we believe that the above study provides an interesting appraisal of the long-term dynamics relative to research and development investment across leading firms. We try to combine novel econometric techniques that to our knowledge have not yet been applied to micro-econometric study, and thanks to these, we find interesting empirical results that are also consistent with the economic theory on innovation studies.

The overall sectoral tendency by companies to reduce their R&D gap toward the sectoral average is confirmed thanks to our σ-test of relative convergence. This points at the fact that at least among leading innovation firms, there exist mechanisms of imitation and adaptation that push their innovative efforts toward common paths on the long run. However, it is unlikely for the results to be present also among non-leading firms.

Further, with the aim of offering possible classifications that go beyond the simple, standard industrial ones, we run the algorithm and identify four, clearly distinct convergence clubs among firms. Differently from previous convergence studies, the result is achieved by relaxing the assumption of homogeneous convergence speed (among others), and on the contrary, by estimating the idiosyncratic components of each firms' convergence path. This alternative approach to classification can be very helpful, especially when dealing with large, contemporary firms, which very often operate under conglomerates that often transcend the classical industrial sectors.

The identified convergence clubs are composed by firms that share very distinct characteristics, both in terms of R&D intensities and of basic, accounting figures. These data merged with the knowledge coming from their main sector of operations helped us to portrait this representation of the long-term dynamics relative to the leading, incumbent global investors on R&D.

Further work can be done to either extend these techniques to a non-balanced version of this data, but also to different samples of firms (e.g. using data from Compustat) to identify possibly other emergent clubs and compare the results in light of the different features that non-leading firms have.

References

S. Amoroso, P. Moncada-Paternò-Castello, and A. Vezzani. R&D profitability: the role of risk and Knightian uncertainty. *Small Business Economics*, 48:331–343, 2017.

D. Archibugi. Pavitt's taxonomy sixteen years on: a review article. *Economics of Innovation and New Technologies*, 10:415–425, 2001.

D. Archibugi and M. Pianta. Aggregate covergence and sectoral specialization in innovation. *Journal of Evolutionary Economics*, 4:17–33, 1994.

R. J. Barro. Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106: 407–444, 1991.

R. J. Barro and X. Sala-i Martin. *Economic Growth*. McGraw-Hill, 1995.

M Bartkowska and A. Riedl. Regional convergence clubs in europe: identification and conditioning factors. *Economic Modelling*, 29:22–31, 2012.

M. Bianchi. Testing for convergence: evidence from non-parametric multimodality tests. *Journal of Applied Econometrics*, 12:393–409, 1997.

F. J. Blanco, F. Delgado and M. J. Presno. Fiscal decentralization policies in the EU: A comparative analysis through a club convergence analysis. *Journal of Comparative Policy Analysis: Research and Practice*, pages 1–24, 2019.

F. Bogliacino and M. Pianta. The Pavitt taxonomy, revisited: patterns of innovation in manufacturing and services. *Economia Politica*, 33:153–180, 2016.

M. T. Borsi and N. Metiu. The evolution of economic convergence in the European Union. *Empirical Economics*, 48:657–681, 2014.

G. Bottazzi and A. Secchi. Common properties and sectoral specificities in the dynamics of US manufacturing companies. *Review of Industrial Organization*, pages 217–232, 2003.

S. Breschi and F. Malerba. *Sectoral Innovation Systems: technological regimes, Schumpeterian dynamics, and spatial boundaries*, pages 130–156. Pinter, 1997.

S. Breschi, F. Malerba, and L. Orsenigo. Technological regimes and schumpeterian patterns of innovation. *The Economic Journal*, 110:388–410, 2000.

M. Camarero, A. J. Picazo-Tadeo, and C. Tamarit. Are the determinants of CO2 emissions converging among OECD countries? *Economics Letters*, 118:159–162, 2013.

F. Castellacci. Technology clubs, technology gaps and growth trajectories. *Structural Change and Economic Dynamics*, 19:301–314, 2008.

F. Castellacci and D. Archibugi. The technology clubs: the distribution of knowledge across nations. *Research Policy*, 37:1659–1673, 2008.

D. Castellani, M. Piva, T. Schubert, and M. Vivarelli. R&D and productivity in the us and the eu: sectoral specificities and differences in the crisis. *Technological Forecasting and Social Change*, page in press, 2018.

A. Coad. Persistent heterogeneity of R&D intensities within sectors: evidence and policy implications. *Research Policy*, 2018.

A. Colombelli, J. Krafft, and F. Quatraro. High-growth firms and technological knowledge: do gazelles follow exploration or exploitation strategies? *Industrial and Corporate Change*, 23:261–291, 2013.

M. Domns and T. Dunne. Capital adjustment patterns in manufacturing plants. *Review of economic dynamics*, 1:409–429, 1998.

G. Dosi. Sources, procedures, and microeconomic effects of innovation. *Journal of Economic Literature*, 26:1120–1171, 1988.

G. Dosi and M. Tranchero. The role of comparative advantage and endowments in structural transformation. *LEM Working Paper Series*, pages 1–37, 2018.

G. Dosi, G. Nelson, and S. Winter. *The Nature and Dynamics of Organizational Capabilities*. Oxford University Press, Oxford, 2008.

G. Dosi, S. Lechevalier, and A. Secchi. Introduction: interfirm heterogeneity - nature, sources and consequences for industrial dynamics. *Industrial and Corporate Change*, 19:1867–1890, 2010.

K. Du. Econometric convergence test and club clustering using stata. *The Stata Journal*, 17:882–900, 2017.

N. S. Durlauf and P. Johnson. Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics*, 10:365–384, 1995.

R. Evangelista. Sectoral patterns of technological change in services. *Economics of Innovation and New Technologies*, 9:183–221, 2000.

P Gkotsis, E. Pugliese, and A. Vezzani. A technology-based classification of firms: Can we learn something looking beyond industry classifications? *Entropy*, 20:887–901, 2018.

H. G. Graboskwi and N. D. Baxter. Rivalry in industrial research and development: an empirical study. *The Journal of Industrial Economics*, 21:209–235, 1973.

Z. Griliches. Issues in assessing the contribution of research and development to productivity growth. *The Bell Journal of Economics*, 10:92–1!6, 1979.

Z. Griliches, A. Pakes, and B. H. Hall. The value of patents as indicators of inventive activity. *NBER Working Paper*, 2083:1–43, 1986.

B. H. Hall. The stock market valuation of R&D investment during the 1980's. *American Economic Review*, 83:259–264, 1993.

B. H. Hall and F. Hayashi. Research and development as an investment. *NBER Working Paper*, 2973: 1–45, 1989.

H. Hernández, N. Grassano, A. Tübke, L. Potters, P. Gkotsis, and A. Vezzani. *The 2018 EU Industrial R&D investment scoreboard*. Publications Office of the European Union, Luxembourg, 2018.

G. Hoberg and G. Phillips. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124:1423–1465, 2016.

R. Hodrick and E. Prescott. Post war business cycles: an empirical investigation. *Journal of Money, Credit and Banking*, 29:1–16, 1997.

F. H. Knight. *Risk, Uncertainty, and Profit*. Hart, Schaffner and Marx, Houghton Mifflin, Boston, MA, 1921.

W. N. Leonard. Research and development in industrial growth. *Journal of Political Economy*, 79: 232–256, 1971.

R. E. Lucas. On the mechanics of economic development. *Journal of Monetary Economics*, 22:3–42, 1988.

D. G. Malcolm, J. H. Roseboom, C. E. Clark, and W. Fazar. Application of a technique for research and development program evaluation. *Operations Research*, 7:646–669, 1959.

F. Malerba. Sectoral systems of innovation and production. *Research Policy*, 31:247–264, 2002.

F. Malerba. Sectoral systems: How and why innovation differs across sectors. In J. Fagerberg, D. C. Mowery, and R. R. Nelson, editors, *The Oxford Handbook of Innovation*, pages 380–406. Oxford University Press, 2005.

N. Mankiw, D. Romer, and D. N. Weil. A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107:407–437, 1992.

E. Mansfield and R. Brandenburg. The allocation, characteristics, and outcome of the firm's research and development portfolio: A case study. *The Journal of Business*, 39:447–464, 1966.

M. Miozzo and L. Soete. Internationalization of services: a technological perspective. *Technological Forecasting and Social Change*, 67:159–185, 2001.

P. Moncada-Paternò-Castello, M. Vivarelli, and P. Voigt. Drivers and impacts in the globalization of corporate R&D: an introduction based on the European experience. *Industrial and Corporate Change*, 20:585–603, 2011.

P. Moncada-Paternò-Castello, F. Hervás, N. Grassano, A. Tübke, and A. Vezzani. EU corporate R&D intensity gap: structural features call for a better understanding of industrial dynamics. *JRC Policy Brief - European Commission*, pages 1–7, 2016.

M. Monfort, J. C. Cuestas, and J. Ordóñez. Real convergence in europe: A cluster analysis. *Economic Modelling*, 33:689–694, 2013.

F. Montobbio. Sectoral patterns of technological activity and export market share dynamics. *Cambridge Journal of Economics*, 27:523–545, 2003.

D. Neef. The effect of knowledge on national economies. In D. Neef, A. Siesfield, and J. Cefola, editors, *The Economic Impact of Knowledge*, pages 1–16. Butterworth Heinemann, 1998.

R. R. Nelson and S. Winter. *An Evolutionary Theory of Economic Change*. Harvard University Press, Cambridge, 1982.

OECD. *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*. OECD Publishing, 2015.

OECD. Research and Development Statistics, 2018. data retrieved from OECD Statistics, `http://oe.cd/rds` on January, 21 2019.

E. Panopoulou and T. Pantelidis. Club convergence in carbon dioxide emissions. *Environmental and Resource Economics*, 44:47–70, 2009.

P. Patel and K. Pavitt. Uneven and divergent technological accumulation among advanced countries: evidence and a framework of explanation. *Industrial and Corporate Change*, 13:343–373, 1994.

P. Patel and K. Pavitt. The technological competencies of the world's largest firms: complex and path-dependent, but not much variety. *Research Policy*, 26:141–156, 1997.

K. Pavitt. R&D, patenting and innovative activities: A statistical exploration. *Research Policy*, 11:33–51, 1982.

K. Pavitt. Sectoral patterns of technical change: towards a taxonomy and a theory. *Research Policy*, 13: 343–373, 1984.

M. Peneder. Technological regimes and the variety of innovation behaviour: creating integrated taxonomies of firms and sectors. *Research Policy*, 39:323–334, 2010.

C. B. Phillips and D. Sul. Transition modeling and econometric convergence tests. *Econometrica*, 75: 1771–1855, 2007a.

C. B. Phillips and D. Sul. Some empirics on economic growth under heterogeneous technology. *Journal of Macroeconomics*, 29:455–469, 2007b.

C. B. Phillips and D. Sul. Economic transition and growth. *Journal of Applied Econometrics*, 24: 1153–1185, 2009.

D. T. Quah. Empirics for economic growth and convergence. *European Economic Review*, 40:1353–1375, 1996.

P. M. Romer. Increasing returns and long-run growth. *Journal of Political Economy*, 94:1002–1037, 1986.

F. M. Scherer. Research and development resource allocation under rivalry. *Quarterly Journal of Economics*, 81:359–394, 1967.

F. M. Scherer. Inter-industry technology flows in the US. *Research Policy*, 11:227–245, 1982.

J. A. Schumpeter. *The Theory of Economic Development.* Harvard University Press, Boston, MA, 1911.

J. A. Schumpeter. *Capitalism, Socialism and Democracy.* Harper, New York, NY, 1942.

H. A. Simon. Rational choice and the structure of the environment. *Psychological Review*, 63:129–138, 1956.

J. J. W. Smith and D. Creamer. R&D and small-company growth; a statistical review and company case studies. *National Industrial Board - Technical Report*, 1968.

R. Veugelers and M. Cincera. The impact of horizon 2020 on innovation in europe. *Intereconomics - Forum*, 50:2–9, 2015.

E. R. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Association*, 78:81–89, 1923.

M. Yildizoglu. Competing R&D strategies in an evolutionary industry model. *Computational Economics*, 19:51–65, 2002.

Appendix

Table 6: Log-t regression test results on the initial club classification for the whole, balanced sample using R&D intensity

Initial clubs	Coefficient	T-statistic
Club 1	0.033	0.953
Club 2	0.407	1.367
Club 3	0.382	5.216
Club 4	0.606	8.275
Club 5	0.996	6.899
Club 6	0.391	14.185
Club 7	1.658	11.513
Club 8	1.353	9.721
Club 9	0.687	8.044
Club 10	0.476	2.712

Number of observations: 472. Number of time periods: 18.

Table 7: Number of firms in each convergence clubs by Revised Pavitt sector

Revised Pavitt sector	R&D Specialists	STEMmers	Good Old Manufacturing	Scale & Energy
Information Network	0	5	5	4
Knowledge-intensive Business Services	4	64	1	1
Physical Network	0	3	1	14
Science-based	4	74	2	3
Supplier-dominated	0	28	16	3
Scale Intensive	1	61	28	33
Specialized Suppliers	2	82	21	12
Total	11	317	74	70

Table 8: Number of firms in each convergence clubs by Revised Pavitt sector - 17-year sub-sample - robustness

Revised Pavitt sector	R&D Specialists	STEMmers	Good Old Manufacturing	Scale & Energy
Information Network	0	5	11	2
Knowledge-intensive Business Services	8	60	1	1
Physical Network	0	4	5	9
Science-based	7	72	2	1
Supplier-dominated	0	29	16	2
Scale Intensive	2	69	44	9
Specialized Suppliers	0	83	29	5
Total	16	322	104	29

Table 9: Number of firms in each convergence clubs by Revised Pavitt sector - 90% random sub-sample - robustness

Revised Pavitt sector	R&D Specialists	STEMmers & Good Old Manufacturing	Scale & Energy
Information Network	0	7	4
Knowledge-intensive Business Services	4	56	2
Physical Network	0	3	12
Science-based	4	71	3
Supplier-dominated	0	40	5
Scale Intensive	1	68	39
Specialized Suppliers	2	88	16
Total	11	333	81

Table 10: Number of firms in each convergence clubs by Revised Pavitt sector - 80% random sub-sample - robustness

Revised Pavitt sector	Large R&D Specialists	Small R&D Specialists	STEMmers	Good Old Manufacturing	Scale & Energy
Information Network	0	2	3	3	1
Knowledge-intensive Business Services	0	44	11	0	0
Physical Network	0	1	2	5	5
Science-based	2	50	13	3	0
Supplier-dominated	0	8	21	10	1
Scale Intensive	0	18	42	30	9
Specialized Suppliers	0	33	39	13	4
Total	2	156	131	64	20

Table 11: Log-t regression test results on the robustness, subsamples for R&D intensity industry gap

Variable	Coefficient	SE	T-statistic
R&D intensity gap 3-digit - 17-year subsample	-5.5903	3.1726	-1.7621
R&D intensity gap 2-digit - 17-year subsample	0.3574	1.0360	0.3450
R&D intensity gap 3-digit - 90% subsample	0.4764	1.0541	0.4520
R&D intensity gap 2-digit - 90% subsample	2.3846	1.5350	1.5535
R&D intensity gap 3-digit - 80% subsample	2.0618	0.7549	2.7312
R&D intensity gap 2-digit - 80% subsample	-1.0875	1.4144	-0.7689

Table 12: Descriptive statistics for each final club - 17-year sub-sample - robustness

Final club	R&D intensity	R&D investment	Net Sales	Employment	Operating Profit	
1 -	.91	928	3,446	8,123	-.01	mean
R&D Specialists	.31	651	1,723	7,339	.12	median
	2	1,240	5,074	8,106	.38	sd
	16	16	16	16	16	n
2 -	.083	1,182	18,365	50,712	.13	mean
STEMmers	.061	353	6,503	21,755	.11	median
	.061	2,113	34,089	75,643	.12	sd
	322	322	322	318	321	n
3 -	.017	303	20,712	60,614	.1	mean
Good, old manufacturers	.017	137	10,238	30,100	.1	median
	.0064	415	33,576	82,853	.077	sd
	104	104	104	103	104	n
4 -	.0043	113	37,097	86,447	.11	mean
Scale&Energy	.0035	76	21,882	37,230	.081	median
	.003	87	38,695	136,687	.11	sd
	29	29	29	28	29	n

The statistics presented refer to the final year of the sample: 2017. All monetary values are expressed in thousand Euros at the end of year 2017.

Table 13: Descriptive statistics for each final club - 90% random sub-sample - robustness

Final club	R&D intensity	R&D investment	Net Sales	Employment	Operating Profit	
1 -	1.1	852	3,505	8,882	-.013	mean
R&D Specialists	.24	385	1,732	8,500	.15	median
	2.5	1,443	5,208	8,001	.47	sd
	11	11	11	11	11	n
2 -	.084	1,163	17,734	54,459	.12	mean
STEMmers &	.058	351	6,789	25,256	.11	median
Good, old manufacturers	.073	2,072	30,820	80,564	.12	sd
	333	333	333	330	332	n
3 -	.0086	182	32,859	64,374	.11	mean
Scale&Energy	.008	93	13,744	31,000	.092	median
	.005	251	49,690	102,677	.087	sd
	81	81	81	79	81	n

The statistics presented refer to the final year of the sample: 2017. All monetary values are expressed in thousand Euros at the end of year 2017.

Table 14: Descriptive statistics for each final club - 80% random sub-sample - robustness

Final club	R&D intensity	R&D investment	Net Sales	Employment	Operating Profit	
1 -	.97	100	222	349	-.32	mean
R&D Specialists	.97	100	222	349	-.32	median
(small)	1.1	20	220	33	.97	sd
	2	2	2	2	2	n
2 -	.13	1,486	13,960	41,685	.14	mean
R&D Specialists	.11	473	4,330	14,617	.13	median
(large)	.084	2,512	29,693	77,041	.14	sd
	156	156	156	154	156	n
3 -	.041	855	20,339	62,604	.11	mean
STEMmers	.036	320	9,929	34,459	.1	median
	.019	1,439	31,359	80,342	.097	sd
	131	131	131	130	130	n
4 -	.013	238	21,969	56,806	.12	mean
Good, old manufacturers	.013	114	10,911	30,000	.1	median
	.0056	331	38,726	80,915	.076	sd
	64	64	64	63	64	n
5 -	.0036	138	44,247	49,900	.12	mean
Scale&Energy	.0035	71	21,477	41,525	.072	median
	.0016	194	53,773	35,323	.12	sd
	20	20	20	19	20	n

The statistics presented refer to the final year of the sample: 2017. All monetary values are expressed in thousand Euros at the end of year 2017.

Table 15: Cross-correlation table for the classifications obtained with the main sample and the three, robustness subsamples

Sample	Original	17-year sub-sample	90% random sub-sample	80% random sub-sample
Original	1.000			
	(0.000)			
17-year sub-sample	0.838	1.000		
	(0.000)	(0.000)		
90% random sub-sample	0.885	0.727	1.000	
	(0.000)	(0.000)	(0.000)	
80% random sub-sample	0.832	0.772	0.755	1.000
	(0.000)	(0.000)	(0.000)	(0.000)

Number of observations: 372

Table 16: Revised Pavitt Taxonomy (2001)

Sector name	Abbreviation
Manufacturing	
Scale-intensive	SI
Supplier-dominated	SD
Science-based	SB
Specialized Suppliers	SS
Services	
Supplier-dominated Services	SDS
Physical Network Services	PNS
Information Network Services	INS
Knowledge Intensive Business Services	KIBS

Table 17: Number of firms in each convergence clubs by ICB 3-digit sector

ICB name	ICB code	R&D Specialists	STEMmers	Good Old Manufacturing	Scale & Energy
53	Oil & Gas Producers	0	0	0	10
57	Oil Equipment, Services & Distribution	0	1	2	0
135	Chemicals	1	29	13	5
173	Forestry & Paper	0	0	0	2
175	Industrial Metals & Mining	0	0	2	7
177	Mining	0	0	0	3
235	Construction & Materials	1	4	1	8
271	Aerospace & Defense	0	12	1	2
272	General Industrials	0	13	5	1
273	Electronic & Electrical equipment	0	36	6	0
275	Industrial Engineering	1	26	13	3
277	Industrial Transportation	0	0	0	1
279	Support Services	0	1	1	1
335	Automobiles & Parts	0	27	4	2
353	Beverages	0	1	0	1
357	Food Producers	0	2	6	2
372	Household Goods & Home Construction	0	3	6	0
374	Leisure Goods	0	10	1	0
376	Personal Goods	0	2	4	2
378	Tobacco	0	1	1	1
453	Health Care Equipment & Services	0	16	1	1
457	Pharmaceuticals & Biotechnology	4	42	0	0
537	General Retailers	0	1	0	0
555	Media	0	3	1	0
575	Travel & Leisure	0	1	0	1
653	Fixed Line Communication	0	1	4	3
657	Mobile Communications	0	0	0	1
753	Electricity	0	0	0	8
757	Gas, Water & Multi-utilities	0	0	0	3
877	Financial Services	0	1	0	0
953	Software & Computer Services	2	24	1	1
957	Technology Hardware & Equipment	2	60	1	1
	Total	11	317	74	70

Table 18: Correspondence table between ICB and Pavitt-Miozzo-Soete classifications

Industry	Super-sector	Sector	Sub-sector	Pavitt
0001 - Oil & Gas	0500 - Oil & Gas	0530 - Oil & Gas Producers	0533 - Exploration & Production	SI
			0537 - Integrated Oil & Gas	SI
		0570 - Oil Equipment, Services & Distribution	0573 - Oil Equipment & Services	SI
			0577 - Pipelines	SI
		0580 - Alternative Energy	0583 - Renewable Energy Equipment	SI
			0587 - Alternative Fuels	SI
1000 - Basic Materials	1300 - Chemicals	1350 - Chemicals	1353 - Commodity Chemicals	SI
			1357 - Specialty Chemicals	SI
	1700 - Basic Resources	1730 - Forestry & Paper	1733 - Forestry	SI
			1737 - Paper	SI
		1750 - Industrial Metals & Mining	1753 - Aluminum	SI
			1755 - Nonferrous metals	SI
			1757 - Iron & Steel	SI
		1770 - Mining	1771 - Coal	SI
			1773 - Diamonds	SI
			1775 - General Mining	SI
			1777 - Gold Mining	SI
			1779 - Platinum & Precious Metals	SI

	2300 - Construction & Materials	2350 - Construction & Materials	2353 - Building Materials & Fixtures	SS
			2357 - Heavy Construction	SS
2000 - Industrials	2700 - Industrial Good & Services	2710 - Aerospace & Defense	2713 - Aerospace	SB
			2717 - Defense	SB
		2720 - General Industrials	2723 - Containers & Packaging	SD
			2727 - Diversified Industrials	SD
		2730 - Electronic & Electrical Equipment	2733 - Electrical Components & Equipment	SS
			2737 - Electronic Equipment	SS
		2750 - Industrial Engineering	2753 - Commercial Vehicles & Trucks	SS
			2757 - Industrial Machinery	SS
		2770 - Industrial Transportation	2771 - Delivery Services	PN
			2773 - Marine Transportation	PN
			2775 - Railroads	PN
			2777 - Transportation Services	PN
			2779 - Trucking	PN
		2790 - Support Services	2791 - Business Support Services	PN
			2793 - Business Training & Employment Agencies	PN
			2795 - Financial	PN
			2797 - Industrial Suppliers	PN
			2799 - Waste & Disposal Services	PN
	3300 - Automobiles & Parts	3350 - Automobiles & Parts	3353 - Automobiles	SI
			3355 - Parts	SI
			3357 - Tires	SI

	3500 - Food & Beverages	3530 - Beverages	3533 - Brewers	SI
			3535 - Distillers & Vintners	SI
			3537 - Soft Drinks	SI
		3570 - Food Producers	3573 - Farming & Fishing	SI
			3577 - Food Products	SI
	3700 - Personal & Household Goods	3720 - Household Goods & Home Construction	3722 - Durable Household Products	SD
			3724 - Non-Durable Household Products	SD
			3726 - Furnishing	SD
			3728 - Home Construction	SD
		3740 - Leisure Goods	3743 - Consumer Electronics	SD
			3745 - Recreational Products	SD
			3747 - Toys	SD
		3760 - Personal Goods	3763 - Clothing & Accessories	SD
			3765 - Footwear	SD
			3767 - Personal Products	SD
		3780 - Tobacco	3785 - Tobacco	SI
4000 - Health Care	4500 - Health Care	4530 - Health Care Equipment & Services	4533 - Health Care Providers	SDS
			4535 - Medical Equipment	SS
			4537 - Medical Supplies	SS
		4570 - Pharmaceuticals & Biotechnology	4573 - Biotechnology	SB
			4577 - Pharmaceuticals	SB

5000 - Consumer Services	5300 - Retail	5330 - Food & Drug Retailers	5333 - Drug Retailers	PN
			5337 - Food Retailers & Wholesalers	PN
		5370 - General Retailers	5371 - Apparel Retailers	PN
			5373 - Broadline Retailers	PN
			5375 - Home Improvement Retailers	PN
			5377 - Specialized Consumer Services	PN
			5379 - Specialty Retailers	PN
	5500 - Media	5550 - Media	5553 - Broadcasting & Entertainment	IN
			5555 - Media Agencies	IN
			5557 - Publishing	SD
	5700 - Travel & Leisure	5750 - Travel & Leisure	5751 - Airlines	PN
			5752 - Gambling	PN
			5753 - Hotels	SDS
			5755 - Recreational Services	PN
			5757 - Restaurant & Bars	SDS
			5759 - Travel & Tourism	PN
6000 - Telecommunications	6500 - Telecommunications	6530 - Fixed Line Communications	6533 - Fixed Line Communications	IN
		6570 - Mobile Communications	6575 - Mobile Communications	IN
7000 - Utilities	7500 - Utilities	7530 - Electricity	7535 - Conventional Electricity	PN
			7537 - Alternative Electricity	PN
		7570 - Gas, Water & Multi-utilities	7573 - Gas Distribution	PN
			7575 - Multi-utilities	PN

			7577 - Water	PN
8000 - Financials	8300 - Banks	8350 - Banks	8355 - Banks	IN
	8500 - Insurance	8530 - Non-life Insurance	8532 - Full Line Insurance	IN
			8534 - Insurance Brokers	IN
			8536 - Property & Casualty Insurance	IN
			8538 - Reinsurance	IN
		8570 - Life Insurance	8575 - Life Insurance	IN
	8600 - Real Estate	8630 - Real Estate Investment & Services	8633 - Real Estate Holding & Development	IN
			8637 - Real Estate Services	IN
		8670 - Real Estate Investment Trusts	8671 - Industrial & Office REITs	IN
			8672 - Retail REITs	IN
			8673 - Residential REITs	IN
			8674 - Diversified REITs	IN
			8675 - Specialty REITs	IN
			8676 - Mortgage REITs	IN
			8677 - Hotel & Lodging REITs	IN
	8700 - Financial Services	8770 - Financial Services	8771 - Asset Managers	IN
			8773 - Consumer Finance	IN
			8775 - Specialty Finance	IN
			8777 - Investment Services	IN
			8779 - Mortgage Finance	IN
	8900 - Investment Instruments	8980 - Equity Investment Instruments	8985 - Equity Investment Instruments	IN

		8990 - Non-Equity Investment Instruments	8995 - Non-Equity Investment Instruments	IN
9000 - Technology	9500 - Technology	9530 - Software & Computer Services	9533 - Computer Services	KIBS
			9535 - Internet	KIBS
			9537 - Software	KIBS
		9570 - Technology Hardware & Equipment	9572 - Computer Hardware	SB
			9574 - Electronic Office Equipment	SS
			9576 - Semiconductors	SB
			9578 - Telecommunication Equipment	SS

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub