European Commission

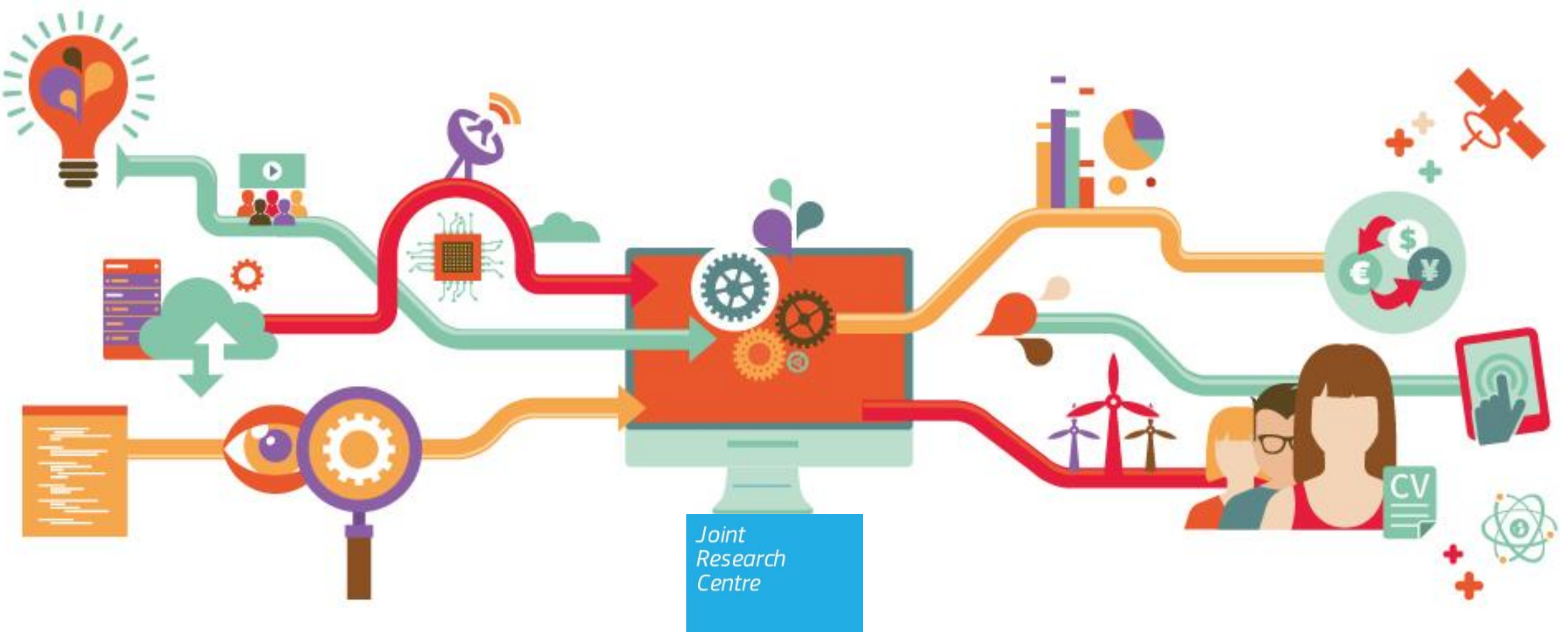
JRC TECHNICAL REPORTS

Persistent heterogeneity of R&D intensities within sectors: Evidence and policy implications

JRC Working Papers on Corporate R&D and Innovation No 04/2017

Alex Coad

2017



Joint Research Centre

Contact information
Antonio Vezzani
Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: mailto: jrc-b3-secretariat@ec.europa.eu
Tel.: +34 954488463
Fax: +34 954488316

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

JRC105670
ISSN 1831-9408 (online)

Seville, Spain: European Commission, 2017

How to cite: Coad, A. (2017). Persistent heterogeneity of R&D intensities within sectors: Evidence and policy implications. JRC Working Papers on Corporate R&D and Innovation, No 04/2017, Joint Research Centre.

The JRC Working Papers on Corporate R&D and Innovation addresses economic and policy issues related to industrial research and innovation and to the competitiveness of the European industry. Mainly addressed to policy analysts and the academic community, these are policy relevant early-stage scientific articles highlighting policy implications. These working papers are meant to communicate to a broad audience preliminary research findings, generate discussion and attract critical comments for further improvements. All papers have undergone a peer review process.

Persistent heterogeneity of R&D intensities within sectors: Evidence and policy implications

Alex Coad[1]

CENTRUM-Catolica Graduate Business School, Pontificia Universidad Catolica del Peru, Lima, Peru.

European Commission, Joint Research Centre, Seville, Spain

Abstract

Do firms in the same sector converge towards the same R&D intensities? Previous research has often assumed this to be true. A closer examination, using microdata from the EU Industrial R&D Investment Scoreboard for the years 2000-2015, shows a large amount of heterogeneity in R&D intensities among firms in the same sector, and that this heterogeneity persists over time. Statistical tests of convergence show that the variation in R&D intensities does not decrease over time (i.e. no σ-convergence), although firms with an R&D intensity below the industry average do seem to catch up with the leaders (i.e. evidence of β-convergence). Overall, firms in the same industry do not converge to a common R&D intensity. Policy implications are discussed.

Keywords: R&D investment, R&D intensity, convergence, benchmarking.

JEL Classification: O3, L2.

---

[1] I am grateful to Flavio Calvino, Matthias Deschryvere, Andrea Filippetti, Nicola Grassano, Daniel Halvarsson, Nanditha Mathew, Pietro Moncada-Paternò-Castello, Martin Srholec, Federico Tamagni, Alex Tuebke, Jan van den Biesen (Philips), Antonio Vezzani, and participants at the 6th IRIMA workshop (JRC-IPTS, European Commission, Brussels, Dec 2015) and EMAEE 2017 in Strasbourg for helpful comments. The usual caveat applies.

1. Introduction

Innovation plays a key role in the creation of high-productivity jobs in today's knowledge-based economies. Countries have therefore sought to encourage investments in innovation via the R&D spending of private firms. In Europe, the levels of R&D investments of EU member states lags behind its potential, as well as lagging behind other countries such as the USA, Japan and South Korea. For this reason, the Lisbon Strategy in 2000 has set an R&D intensity target of R&D expenditures as 3% of GDP. However, this 3% target has not been reached – instead R&D intensities in Europe have stagnated (van Pottelsberghe de la Potterie, 2008; Moncada-Paternò-Castello et al., 2011).

How can the gap to the 3% target be closed? Two different routes to 3% can be mentioned here, often dubbed the 'structural effect' and the 'intrinsic effect' (Cincera and Veugelers, 2013). The 'structural effect' emphasizes the role of industrial sector, and expects that high-R&D firms are to be found in high-tech sectors, while low-R&D firms are to be found in low-tech sectors. If firm-level R&D intensities are inert and determined by industrial sector, then increasing the R&D intensities of countries therefore involves increasing their presence in high-tech sectors. In contrast, the 'intrinsic effect' maintains that incumbent firms could be stimulated to increase their R&D investments within their existing sectors. Scholars have often emphasized the first route, e.g.:

> *"In order to achieve its 3 % target for R&D intensity and boost its competitiveness and job creation, the EU needs to adapt its industrial structure and increase economic activity in the high-R&D-intensive sectors."*
> Moncada-Paternò-Castello et al., (2016, p1).

Scholars have noticed that, within industries, European firms are not less R&D intensive than their US counterparts (Veugelers and Cincera, 2010; Moncada-Paternò-Castello et al., 2010). "Within any particular sector, the R&D intensity of individual firms in the EU is superior to that of similar firms in competitor economies." (Moncada-Paternò-Castello et al., 2016, p2, see also Mathieu and van Pottelsberghe de la Potterie, 2010). Instead, it appears that the European R&D intensity gap is due to having fewer firms in high-tech sectors, alongside a higher proportion of firms in low- or medium-tech sectors. Europe, in other words, does have large innovative firms, but they might be in the 'wrong' sectors. In particular, Europe has fewer young leading innovators ("yollies"; Veugelers and Cincera, 2010) that expand rapidly to become leaders of newly-created high-tech industries.

Much attention has been paid to the compositional structure of European industry (Mathieu and van Pottelsberghe de la Potterie, 2010, Reinstaller and Unterlass, 2012, Cincera and Veugelers, 2013). It has been emphasized that EU member states could pursue the 3% target by changing their industrial structure, in favour of high-tech industries, rather than boosting the R&D intensities of firms in low- or medium-tech

sectors. This hinges on the assumption that all firms in high-tech sectors have high R&D intensities, while firms in low- or medium-tech sectors have lower R&D intensities. The current paper argues that the 'second route to 3%' via increases in intrinsic R&D efforts provides scope for increasing countries' R&D intensities by encouraging firms to increase their R&D intensities whatever their sector of activity.

The emphasis on the structural effect, at the expense of the intrinsic effect, appears to us to be somewhat exaggerated in the literature. Two examples can be mentioned here. First, Mathieu and van Pottelsberghe de la Potterie (2010, p56) write: "For instance, country A might have two industries with higher R&D intensities than in country B, but display a lower aggregate R&D intensity because of a strong specialization in the low R&D intensity industry. One would wrongly conclude that country B is more R&D intensive than country A." We disagree – the latter conclusion would not be wrong, country B is indeed more R&D intensive than country A. When discussing unconditional R&D intensity targets, all that matters is aggregate R&D intensity – regardless the industrial structure. Second, the opening line of Reinstaller and Unterlass (2012) reads: "Direct comparisons of R&D expenditures relative to GDP are flawed as especially the Business R&D Expenditures (BERD) are heavily influenced by the industrial structure of each country." However, we disagree – if the goal is to make unconditional comparisons of R&D intensities across countries, then we are not interested in 'excuses' such as the industrial structure (or firm size, or firm age, or any other potential contributing factor); omitting the role of control variables is not a 'flaw' if the task is to compare unconditional aggregates.

We contribute new evidence to the debate regarding the importance of industry structure for R&D intensities, focusing on heterogeneity of R&D intensities within sectors. Cohen and Klepper (1992) use line-of-business-level data to investigate heterogeneity of R&D intensities within sectors, and Reinstaller and Unterlass (2012) use sector-level data to investigate different R&D intensities of sectors across countries, whereas we use a unique source of firm-level data covering the world's largest R&D investors. We show that there is significant heterogeneity in R&D intensities among firms in the same sector. Although previous work has shown evidence of heterogeneity of R&D intensities within sectors (e.g. the 'intrinsic' effect in Mathieu and van Pottelsberghe de la Potterie, 2010; Cincera and Veugelers, 2013), nevertheless, the emphasis has usually been placed on the 'structural' effect and the role of industrial sector. We also investigate the possibility of the dynamic phenomenon of convergence of R&D intensities among firms over time. To our knowledge, it is the first investigation of industry-level convergence of R&D intensities. We present new evidence from the world's largest R&D investors, for recent years, including after the 2008 crisis. We present non-parametric graphs, as well as parametric regressions, that show a coherent story – that there is considerable heterogeneity in R&D intensities between firms in the same sector, and that this heterogeneity is persistent.

Section 2 contains the theoretical background and hypotheses development. Section 3 describes the database. Section 4 contains the non-parametric and parametric analysis and Section 5 concludes.

2. Background and Hypotheses Development

The literature on R&D investment holds that firms are unable to calculate the optimal R&D investment level, because of uncertainty surrounding the future benefits of R&D. Instead of making R&D decisions on the basis of exact rational calculations, they instead follow rules of thumb (Thompson, 1999), such as aiming for a target R&D intensity (as measured by a fixed R&D / sales ratio):

> *"… firms tend to work with relatively general and event-independent routines (with rules of the kind "... spend x% of sales on R&D," "...distribute your research activity between basic research, risky projects, incremental innovations according to some routine shares..." and sometimes metarules of the kind "with high interest rates or low profits cut basic research," etc.). This finding is corroborated by ample managerial evidence and also by recent more rigorous econometric tests…"* (Dosi, 1988, p1134)

The following quote illustrates how R&D decisions in firms may be taken with reference to a targeted R&D intensity which operates as a rough rule of thumb:

> *"You have a product. The product is selling. That gives you a certain stream of revenue. You can take that stream of revenue and put some of it into R&D for the next round. Some of it has to be reserved for manufacturing, some of it for profits. Now, if you are on an upward swing and your product is succeeding, you have a flow back of money to invest in R&D; and if it isn't, you don't. And in my experience, and the experience of many other people, oddly enough, R&D is determined, more or less, as a percent of sales. It is not an independent variable. Let me say once more.* ***R&D is often a fixed percent of sales****. Now I exaggerate to make my point.* ***Ten percent is a very reasonable sort of number in a high-tech industry****… It may be that, in the correlation, which has often been remarked on, between R&D spending and industrial success, it is the industrial success which causes the R&D spending, not the other way around."*

> Ralph Gomory, former senior vice-president of IBM and former member of the US President's Council of Advisers on Science and Technology. Gomory (1992, p392), cited in Thompson (1999 p323), emphasis added.)

Note that the quote above suggests that the figure of 10% is a reasonable figure for a high-tech industry. Indeed, in the presence of uncertainty surrounding the optimal level of R&D investment, firms may follow a second rule of thumb, which is pursuing the same R&D intensity as that of its rivals in the same industry (Grabowski and Baxter, 1973). If there is imperfect information on firm behaviour and performance, firms in the same sector can be benchmarked against each other to ensure that their performance remains competitive (Shleifer, 1985). Firms themselves may benefit from a sector-level R&D intensity target if it reduces the uncertainty surrounding their decisions and simplifies their strategic decisions. In addition, investors may lack the information to make detailed decisions at a firm-level, and may instead choose whether to invest in individual firms on the basis of how they perform relative to sector-level indicators. This may result in a situation where investors put pressure on firms in the same industry to pursue the same sector-level performance targets (Aune et al., 2010), such as putting pressure on firms in the same industry to pursue the same R&D intensity target. Scherer (1965, 1967) and Grabowski and Baxter (1973) suggest that competitors are more likely to imitate the R&D expenditures of their rivals as industries become more R&D intensive, thus dampening the relative dispersion of R&D intensities within industries.

This possibility of sector-specific R&D intensity targets has been confirmed by R&D managers:

> *"R&D budget [sic] are driven by sector specificities.* ***R&D budgets are largely driven by sector specific needs and are broadly set as a percentage of sales.*** *According to Phillips, this is one of the ratios to which financial analysts pay attention for R&D intensive companies. Profits do not seem to be an indicator to fix R&D budgets as most times profits are not reinvested in R&D."* Hervas, Dosso and Vezzani (2015, p5, emphasis added)

We therefore hypothesise that:

Hypothesis 1: *firms in the same industry will have the same R&D intensity*

However, industries are characterised by incessant turmoil and creative destruction. The situation in Hypothesis 1 may only arise after some time, as heterogeneous firms, with different starting points, converge towards the industry average R&D intensity. Given that R&D intensity is defined as R&D / sales, we assume that firms that seek to increase their desired R&D intensity will not do so by cutting back on their sales, but by boosting R&D intensity. In other words, the gap between a firm's R&D intensity and the industry's R&D intensity will be positively related to R&D growth (i.e. if the gap is positive, the firm will increase its R&D investment; but if the gap is negative, the firm may reduce R&D).

Hypothesis 2: *firms in the same industry will converge to a common R&D intensity.*

More specifically, statistical tests for convergence refer to whether laggards in a population grow faster than leaders (known as mean reversion or β-convergence) or whether the variation of a population decreases over time (known as σ-convergence). We therefore test the following hypotheses:

Hypothesis 2a: *firms whose R&D intensity is below the industry average will 'catch up' and increase their R&D intensity faster than firms whose R&D intensity is above the industry average (β-convergence).*

However, β-convergence is a necessary but not sufficient condition for convergence (Lichtenberg, 1994; Escribano and Stucchi, 2014). Therefore, we will also pursue a more direct test of whether variation in R&D intensities decreases over time.

Hypothesis 2b: *the variation in R&D intensities among firms in the same sector will decrease over time (σ-convergence)*

3. Database description

Our data come from the EU Industrial R&D Investment Scoreboard, compiled by Bureau Van Dijk (Hernandez et al., 2016). This data source contains information on several thousand of the world's largest R&D investing companies, and together the firms included in the Scoreboard account for about 90% of the total expenditure on R&D by business firms worldwide (Hernandez et al., 2016). Previous analyses of Scoreboard data include Cincera and Ravet, 2010; Garcia-Manjona and Romero-Merino, 2012; and Montresor and Vezzani, 2015. The individual waves of the Scoreboard were merged using the Bureau Van Dijk company-level identifiers, to obtain an unbalanced panel dataset which, in the latest available version, covers the period 2000-2015 (Coad and Grassano, 2017).

While many papers in the literature focus on just one country at a time, using data from that country's statistical office, an advantage of our dataset is that we have data on many countries. Furthermore, due to the way our database is constructed, virtually all firms in our dataset have positive R&D. We therefore do not suffer from statistical problems due to a large number of zero values for R&D, and don't need to tailor our econometric approach to deal with many zeroes (e.g. by using a Tobit approach, as in Czarnitzki and Toole 2011 and Baum et al. 2016).

Of central interest to our paper is the reporting of a firm's annual total R&D expenditure for each available year t, which is reported by the company's headquarters. Other variables, relating to firm performance, are net sales, total employment, capital expenditures, market capitalization, and operating profits. Other controls – which are included in the regressions using full sets of dummy variables – include years, regions (for each of 7 global regions: Asian Tigers, BRIC, EU, Japan, Switzerland, the USA, and 'RoW'.), and Industry Classification Benchmark (ICB) industries, measured either at the ICB 4-digit level (119 sectors) or the ICB 3-digit level (82 sectors). R&D intensity is defined as the ratio of R&D to sales (e.g. Grabowski and Baxter, 1973; Cohen and Klepper 1992), i.e.: R&D intensityit = R&Dit / Salesit .

Data are cleaned to remove negative values of sales or R&D investment. Appendix 1 provides details on the definitions of the main variables.

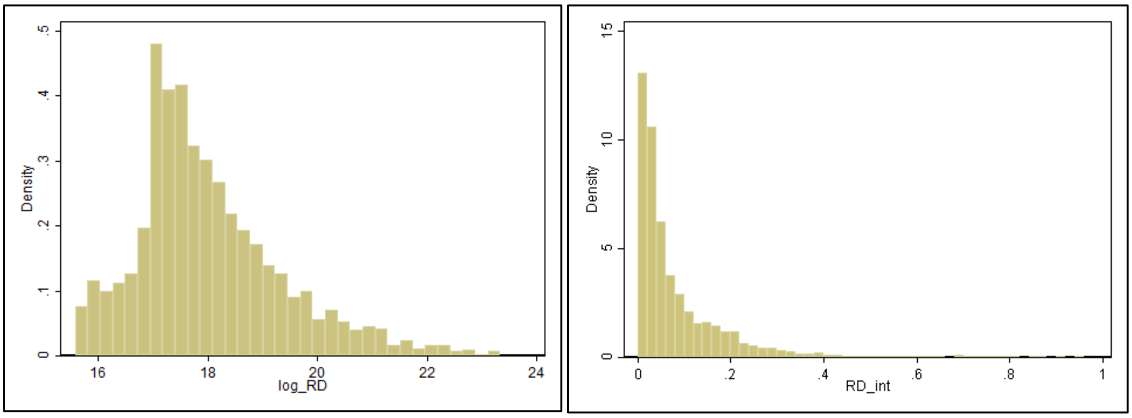
Before performing our regression estimations, we clean the data by removing observations of non-positive net sales, and removing observations where the R&D intensity is negative or is greater than 100% of sales.[2]

4. Analysis

4.1 Descriptives

The average R&D intensity is high in this sample, by construction. Figure 1 (left) shows that R&D investment amounts are smoothly distributed across a wide support: firms do not cluster around any particular 'optimal' amount of R&D investment, but instead there is considerable heterogeneity. Figure 1 (right) shows the histogram for the distribution of R&D intensities. We see a skewed distribution, as suggested at the aggregate level by Bound et al. (1984, their Figure 2.2), and at the industry level by Cohen and Klepper (1992) who analyse line-of-business data. There is lots of heterogeneity in R&D intensities, which can perhaps be explained by firms having heterogeneous innovation capabilities (Cohen and Klepper, 1992).

Figure 1 – Distribution of R&D investments (left) and R&D intensity (right) across firms



Note: Left: histogram of log R&D investment amounts of firms in 2015 (bin width = 0.228). Right: histogram of the distribution of R&D intensities of firms in 2015 (bin width = 0.02). Lower bound at 0.00.

Table 1 contains some summary statistics of R&D investment, R&D growth, R&D intensity, as well as a few other important firm-level variables. Table 2 provides a breakdown of the variance of R&D intensity, which is enabled by the longitudinal nature of the dataset. Table 2 shows that the between element of R&D intensity is larger than the within element. There is a small amount of variation within firms over time, but there is more variation between firms in terms of their R&D intensities. The

[2] This threshold of 100% is modified in subsequent robustness analysis.

within element shows that a firm's R&D intensity is not entirely inert but exhibits some variation over time.

Appendix 5 presents results from an ANOVA decomposition of variance into within-sector and between-sector components for individual years. There is overall more variation within sectors than between sectors. However, the F-statistics are highly statistically significant, which shows that sectors are a meaningful way of categorizing firms according to their R&D intensities, even if they can only explain a minority share of the variation in R&D intensity.

Table 2 –Summary statistics

Variable	Mean	Median	Sd	Skewness	Kurtosis	Min	Max	N
R&D investment	1.72E+08	31467104	6.22E+08	8.5173168	98.210018	1.74394	1.36E+10	45303
R&D growth	0.11497176	0.0721302	0.4347672	-1.365473	133.96978	-15.418806	8.8601933	40126
Net Sales	5.26E+09	9.02E+08	1.72E+10	10.56891	172.8816	0	4.32E+11	47101
R&D intensity	0.0890	0.0416	0.1288	3.2586	16.4716	8.713E-07	0.9968	43778
Capital expenditures	3.65E+08	38515843	1.56E+09	11.658787	192.95279	2.28103	4.41E+10	40413
Operating profit	4.81E+08	55144000	2.21E+09	7.3275903	205.88425	-8.86E+10	6.54E+10	47021
Employees	15323.745	2752	42408.222	6.4318206	60.063272	0	961000	50097
Market Capitalization	7.69E+09	1.32E+09	2.94E+10	23.807796	1030.8479	0	1.67E+12	31424

Note: level variables are reported in Euro.

Table 2: Summarizing the panel dataset: breakdown of the variance of R&D intensity into within-firm and between-firm components

Variable		Mean	Std. Dev.	Observations
RD_int	**overall**	0.088	0.129	N = 43778
	between		0.156	n = 4627
	within		0.053	T-bar = 9.46142

Convergence is investigated with reference to the industry average R&D intensity. For firms i=1..N in an industry, for year t, industry average R&D intensity for the industry (at the level of Industry Classification Benchmark (ICB) 4 digit codes) is calculated as:

$$\overline{RD\_intensity_{jt}} = \frac{\sum_{i=1}^{N} RD\_intensity_{it}}{N}$$

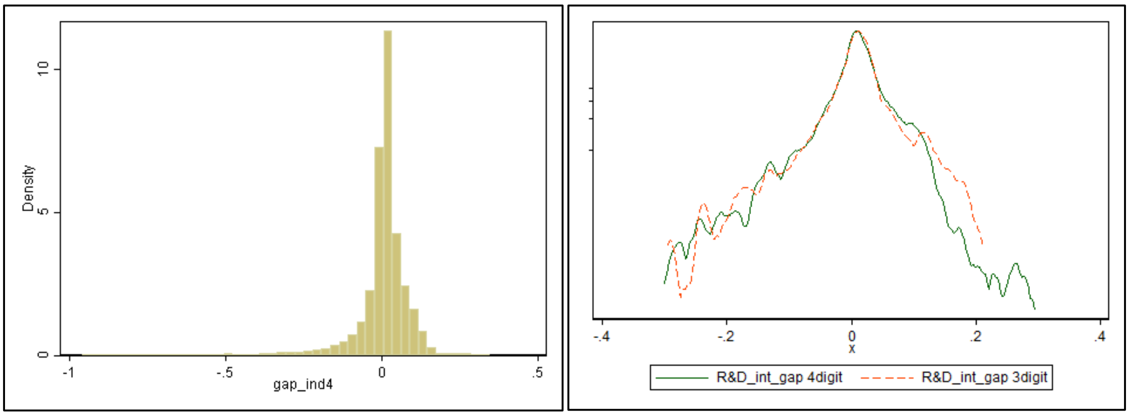
The gap with respect to the industry average (i.e. the variable R&D_intensity_gap) is calculated as:

$$RD\_intensity\_gap_{it} = \overline{RD\_intensity_{Jt}} - RD\_intensity_{it}$$

Hence, R&D_intensity_gap corresponds to the R&D intensity of a company, normalized by its sector of activity. The gap is positive if the firm's R&D intensity is lower than that of the industry. A test of convergence would be that if the gap is positive, then a firm will increase its R&D intensity, to close the gap.

Figure 2 (left) presents a histogram of the distribution of R&D_intensity_gap. There is lots of heterogeneity in R&D intensities, even between firms in the same industry. While most firms have an R&D intensity close to the industry average (i.e. R&D_intensity_gap is close to zero), nevertheless the large variation in the distribution of R&D_intensity_gap suggests that there is lots of variation in R&D intensities even within industries. Figure 2 (right) shows the same distribution with a logarithmic y-axis. (The familiar 'tent-shaped' distribution, when plotted with a logarithmic y-axis, suggests that the distribution could be approximately Laplace).

Figure 2 – R&D investment gap (left) and its kernel density (right)



Note: Left: histogram of the variable R&D_intensity_gap (i.e. the gap between a company's R&D intensity and the average R&D intensity of its ICB 4 digit sector). Right: kernel density of the variable R&D_intensity_gap, where sector is defined at the 3-digit or 4-digit sector. Note the log scale on the y-axis.

4.2 Line plots of R&D intensity of leading firms in the same sector

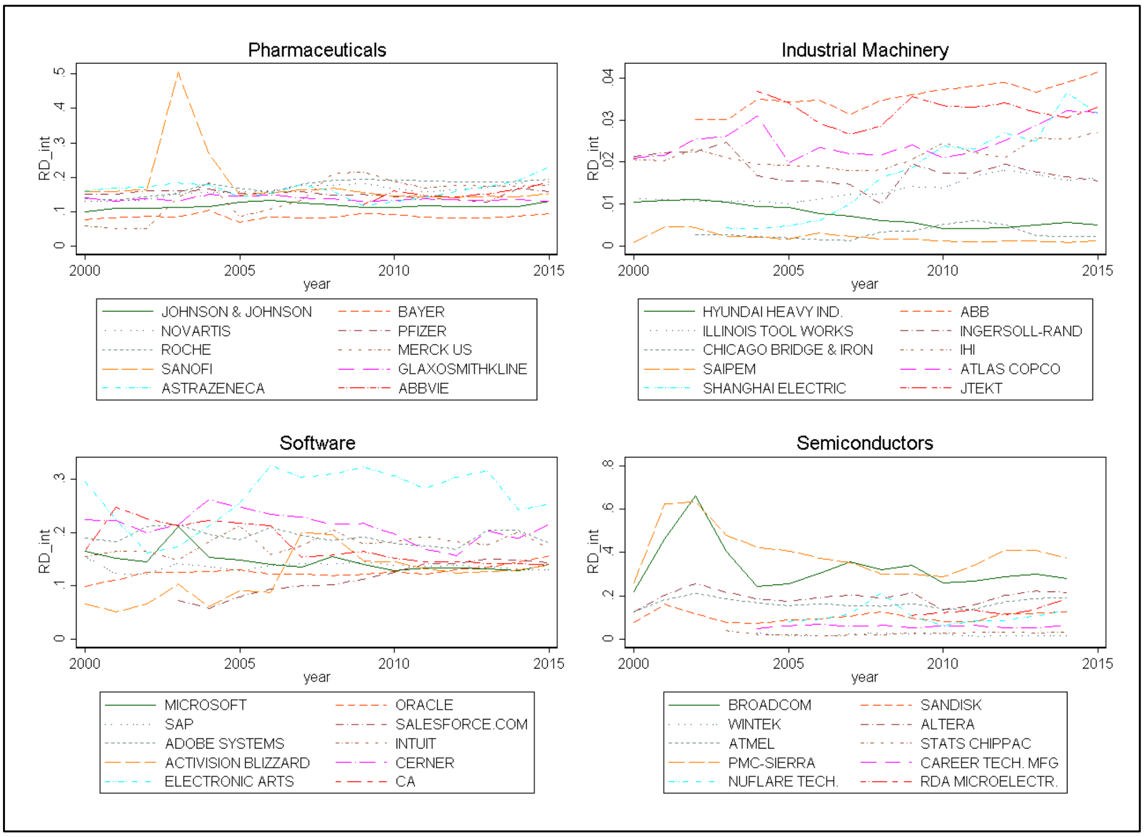
In this subsection, we focus on specific sectors, and see if the leading firms in each sector show any convergence to a common R&D intensity over the available years. More specifically, we focus on the sectors that are both well-known to innovation scholars, and also among the most highly represented (in terms of numbers of companies) in the year 2015. These sectors are: Industrial machinery (169 firms with non-missing R&D intensities in 2015), Software (167 firms in 2015), and Pharmaceuticals (140 firms in 2015). In addition, we include the semiconductors

sector, because this sector is often considered to be an R&D-intensive industry, where firms are assumed to share a similar R&D intensity.[3]

By "leading firms", we choose firms that have the largest total sales. We focus on the 10 largest firms in each sector to have a reasonably large number of leaders in order to view interactions and interdependencies between them, allowing a clear picture of the overall trends even if some particular firms may appear erratic.[4]

The plots in Figure 3 show that the top-10 leading R&D investors in each sector do not clearly show convergence to a common R&D intensity. Instead they display persistent heterogeneity. This gives some early doubts against Hypothesis 1 and Hypotheses 2a and 2b.

Figure 3: Evolution of R&D intensities of the 10 largest firms (in terms of sales in 2015) in four sectors.



Note: For the Semiconductors sector, there are few observations for 2015, so firms are instead ranked in terms of sales in 2014.

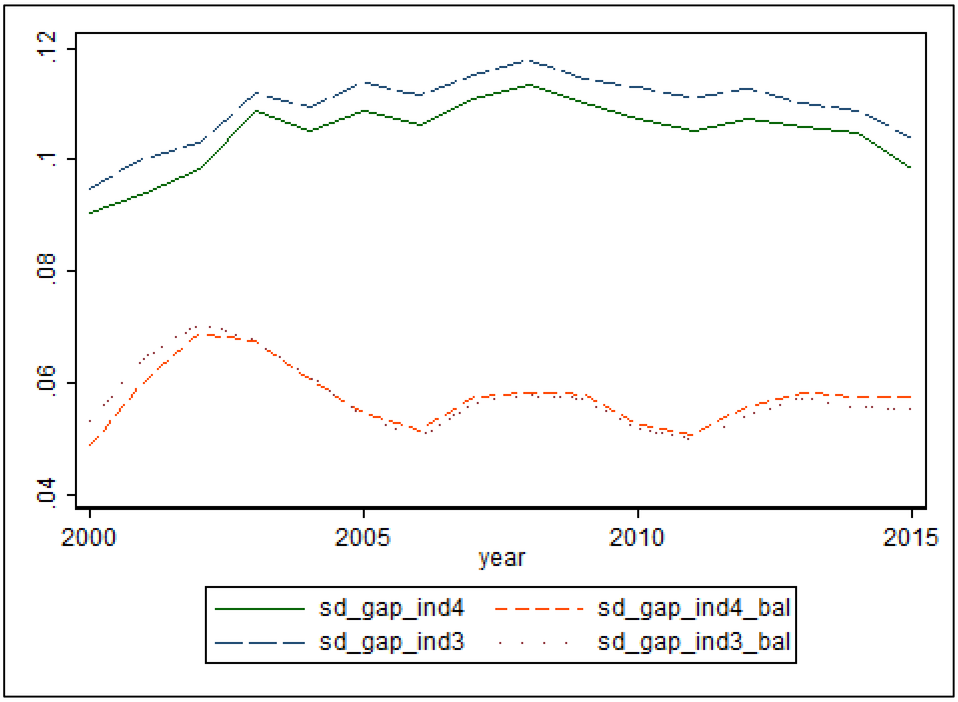
For the industrial machinery sector, the R&D intensities seem to fan out, and diverge over time, with some firms overtaking others. The opposite seems to occur for

[3] We are grateful to Jan van den Biesen (Philips) for this helpful suggestion.
[4] Focusing instead on a smaller number of leaders (e.g. the 2 largest firms; Sutton 2007 AER) would be less appropriate here, because there would be fewer cases.

software - the variation appears to decrease slightly over the years. The pharmaceuticals sector appears to show persistent heterogeneity without convergence (ignoring the blip in 2003 for Sanofi), because the variation between companies in 2015 appears similar to the variation in 2000. The R&D intensity of AstraZeneca remains about double the R&D intensity of Bayer (about 16% vs about 8%), during the period studied. For semiconductors, if we ignore the turbulence in 2000-2003, there is sustained heterogeneity in R&D intensities. In all four cases, there are some large firms in each sector that have very low R&D intensities, co-existing in the same sector with other large firms that have higher R&D intensities. Overall, these figures present some early hints that the 10 largest firms in the same sector have heterogeneous R&D intensities, and that the hypothesis of convergence in R&D intensities is not clearly supported.

Figure 3 also helps to alleviate concerns that the industry categories used in this study (ICB3, ICB4) are too broad to capture competitive groups in a meaningful way. Figure 3 highlights some well-known firms in each sector, and shows that these prominent rivals do not closely mimic each other but appear to follow their own R&D intensity paths. Figure 4 shows the evolution of the standard deviation of “R&D_intensity_gap” over years. Measuring the R&D intensity gap with respect to 3-digit or 4-digit sectors changes little the results. What matters more is whether we focus on an unbalanced or a balanced panel – an unbalanced panel contains more observations, but could be affected by the entry and exit of short-lived firms, as well as a shifting sample composition over time, perhaps with an increasing share of small young firms as data coverage improves over time.

Figure 4: evolution of the standard deviation of “R&D_intensity_gap” for panels that are unbalanced or balanced, at the 4-digit and 3-digit levels.



There is no clear trend of either an increase or decrease in the variation of R&D intensities over time. In other words, there is no clear evidence for either convergence or divergence. One hypothesis is that recessions are periods of intense competition, that tends to reduce the diversity in firms' productivity levels (Escribano and Stucchi, 2014) – but Figure 4 does not suggest that within-industry differences in R&D intensities become more or less dispersed over the business cycle (or at least not in our data on large R&D investors).

Figure 4 shows the evolution of the standard deviation of "R&D_intensity_gap" over years. Measuring the R&D intensity gap with respect to 3-digit or 4-digit sectors changes little the results. What matters more is whether we focus on an unbalanced or a balanced panel – an unbalanced panel contains more observations, but could be affected by the entry and exit of short-lived firms, as well as a shifting sample composition over time, perhaps with an increasing share of small young firms as data coverage improves over time. There is no clear trend of either an increase or decrease in the variation of R&D intensities over time. In other words, there is no clear evidence for either convergence or divergence. One hypothesis is that recessions are periods of intense competition, that tends to reduce the diversity in firms' productivity levels (Escribano and Stucchi, 2014) – but Figure 4 does not suggest that within-industry differences in R&D intensities become more or less dispersed over the business cycle (or at least not in our data on large R&D investors).

Our initial observation of the line plots in Figure 3 gives us an initial idea of the dynamics of firms' R&D intensities within specific sectors, while Figure 4 shows that R&D_intensity_gap does not decrease over time, even for the balanced panel (which focuses on changes in surviving firms only). These graphs can be valuable for visualizing time trends and firm heterogeneity, although in the rest of the paper we will complement these initial graphical analysis with formal statistical tests of convergence.

4.3 Parametric Analysis

4.3.1 How well do industry dummies explain the variation in R&D intensity?

How much of the variation in R&D intensity can be explained by a firm's industrial sector, compared to a set of firm-specific performance variables? Table 3 compares the R^2 statistics that emerge when different variables are included in a regression model of R&D intensity.

Table 3 shows that sector dummies, either at the 3-digit level or the 4-digit level, explain much more than country dummies (compare 30% with 8%). If we had more detailed industry classification codes, then we would expect the R^2 to become even higher. However, (logs of) sales or employment are also capable of explaining a large part of the variation in R&D intensities – where just one variable has an R^2 of around

25%. This could be due to sample selection bias, however.[5] Nevertheless, operating profit, which probably suffers less from sample selection bias (because inclusion in the Scoreboard data does not necessarily depend on operating profit), explains by itself a fair share of the variation in R&D intensities (around 12%). We therefore echo the earlier findings of Scott (1984: p233): "company effects as well as industry effects explain a substantial proportion of the variance in R&D intensity."

Table 3: R2 statistics from different OLS regressions of R&D intensity

VARIABLES	(1) RD_int	(2) RD_int	(3) RD_int	(4) RD_int	(5) RD_int	(6) RD_int	(7) RD_int
ICB 3-digit codes	yes						
82 dummies							
ICB 4-digit codes		yes					
119 dummies							
Country codes			yes				
53 dummies							
log_NS				-0.0339***			
				(0.00107)			
log_EMP					-0.0352***		
					(0.00101)		
log_OP						-0.0144***	
						(0.000709)	
log_labprod							-0.0141***
							(0.00200)
Constant	0.0391	0.131	0.0332	0.790***	0.382***	0.329***	0.262***
				(0.0227)	(0.00918)	(0.0134)	(0.0250)
Observations	43,000	43,778	43,778	43,778	39,371	35,506	39,371
R-squared	0.261	0.318	0.088	0.271	0.254	0.122	0.009
ll	33746	35965	29607	34527	31241	40190	25658
ll_0	27246	27598	27598	27598	25485	37885	25485

Note: Robust standard errors in parentheses. Key to significance levels: *** p<0.01, ** p<0.05, * p<0.1.

4.3.1 Statistical tests of convergence

Barro and Sala-i-Martin (1992) suggested two tests for convergence, which remain the basis for the current state of the art (Rodrik, 2013, Escribano and Stucchi, 2014). The first, β-convergence, seeks a negative correlation between the growth rate of R&D intensity of a firm and its initial level. That is, firms with low R&D intensities to start with will have a more positive growth of R&D intensity than those firms that start with

[5] Firms are included in the Scoreboard sample if they have high levels of R&D investment. Small firms, therefore, must have high R&D intensities if they are to be included in the Scoreboard. Large firms, however, can still enter the Scoreboard sample even if they have low R&D intensities – as long as they are sufficiently large. Hence, R&D intensity might vary with size because of the sample construction criteria.

higher R&D intensities. β-convergence has the disadvantage that it is a necessary, but not sufficient, condition for a reduction in the variance (Lichtenberg, 1994). This has led some authors to suggest focusing on σ-convergence instead of β-convergence (Lichtenberg, 1994; Friedman, 1992; Carree and Klomp, 1997; Escribano and Stucchi, 2014).

The second test is σ-convergence, which directly tests for a reduction in variance of firms' R&D intensities. This is done by comparing the variance of R&D_intensity_gap for the initial period σ^2(t=0) with the variance of R&D_intensity_gap for the final period σ^2(t=T). The hypothesis of convergence is not rejected if $\hat{\sigma}_T^2$ is significantly smaller than $\hat{\sigma}_0^2$.

One issue when testing for statistical significance is that the variance at t=T depends on on the variance at t=0, which is problematic for the statistical tests that rely on an F-distribution (Carree and Klomp, 1997).

To be precise, precise inference regarding σ-convergence requires a test for a reduction in variance in the T2 and T3 statistics developed by Carree and Klomp (1997):

$$T_2 = (N - 2.5)ln\left[1 + \frac{1}{4}\frac{(\hat{\sigma}_1^2 - \hat{\sigma}_T^2)^2}{\hat{\sigma}_1^2\hat{\sigma}_T^2 - \hat{\sigma}_{1T}^2}\right]$$

$$T_3 = \frac{\sqrt{N}\,(\hat{\sigma}_1^2 \;/\; \hat{\sigma}_T^2 \; - 1)}{2\sqrt{1 - \hat{\pi}^2}}$$

In our analysis, we present results for both β-convergence and σ-convergence.

β-convergence

In our regressions, we test for β-convergence by examining if there is any effect of a firm's relative R&D intensity and its subsequent changes in R&D intensity. Our dependent variable is therefore the growth of R&D intensity.

An alternative dependent variable, that we use for robustness analysis, is growth of R&D expenditures, which is calculated by taking log-differences (Tornqvist et al., 1985; Coad, 2009). The reasoning here is that a firm that wants to increase its R&D/Sales ratio (i.e. R&D intensity) will not do this by decreasing its sales, but will instead seek to increase its R&D investments. Hence, we assume that firms will manipulate their R&D/Sales ratio through the channel of R&D growth.

Our main explanatory variable is R&D_intensity_gap (t-1), because we are interested in seeing if lagging R&D investors will catch up and converge with the industry average R&D intensity. R&D_intensity_gap is lagged, in order to correspond to the conditions at the beginning of the period of R&D growth (t-1:t). We also control for a number of other possible influences, using the data available to us, such as firm-level information on net sales, capital expenditures, operating profits, number of employees, as well as region and year dummies.

We first present OLS (Ordinary Least Squares) regressions, pooling together observations without adjusting for firm-specific time-invariant components of the dependent variable (R&D intensity growth or R&D growth). Pooling together observations in this way would be acceptable if the R&D intensity growth or R&D growth of a firm in a given year is statistically independent from its value in the following year. In case this assumption is not true, we also present results from the Fixed-Effects or 'within' estimator, which allow us to take into consideration changes within firms over time, and also the effects of changes between firms. For our OLS and Fixed-effects estimates, we exploit the panel dimension of our data by clustering standard errors at the firm-level.

Table 4 shows our baseline results. In all cases, there is a strong positive association between lagged R&D intensity gap and growth of R&D intensity. This is also visible when taking an alternative indicator of R&D intensity gap (measured with respect to 3-digit sectors) as well as an alternative dependent variable (growth of R&D in Appendix 2 and change of R&D in Appendix 3). The positive coefficients are consistent with the hypothesis of β-convergence or 'mean reversion'. The estimated coefficients are higher in the fixed effects specification (columns (9) and (10)) compared to the OLS or between effects specifications, highlighting that the largest effects of β-convergence are to be found within firms over time.

Further robustness analysis can be found in Appendix 4. Table A4.1 shows OLS regressions on subsamples of low-R&D vs high-R&D firms (i.e. whether R&D_intensity_gap is positive or negative), time periods before or after the crisis year 2008, samples delimited by their total sales (thresholds of 1 billion or 10 billion in sales), and with outliers removed (outliers are dropped if R&D intensity is greater than 50%, 30% or 20% respectively). We observe some interesting heterogeneity across subsamples. The coefficient is particularly positive (i.e. high β -convergence of R&D intensities) for the subsample of low-R&D firms; and the subsamples of larger firms. The evidence for β -convergence becomes stronger as outliers are removed.

Table A4.2 repeats the analysis in Table A4.1, but with an alternative dependent variable: growth of R&D, instead of growth of R&D intensity, and similar results are obtained.

Overall, we obtain robust support for Hypothesis 2a.

Table 4: baseline estimates from least squares regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	ols	ols	ols	ols	ols	ols	ols	ols	fe	fe	be	be
	gr_RD_int	gr_RD_int	gr_RD_int	gr_RD_int	gr_RD_int	gr_RD_int	gr_RD_int	gr_RD_int	gr_RD_int	gr_RD_int	gr_RD_int	gr_RD_int
L.gap_ind4	0.802***		0.806***		0.886***		0.884***		2.419***		0.520***	
	(0.0420)		(0.0430)		(0.0450)		(0.0493)		(0.113)		(0.0364)	
L.gap_ind3		0.760***		0.807***		0.884***		0.881***		2.419***		0.520***
		(0.0386)		(0.0429)		(0.0447)		(0.0489)		(0.113)		(0.0361)
L.log_NS					-0.0311*	-0.0318*	-0.0305*	-0.0313*	-0.289***	-0.289***	-0.0294	-0.0305
					(0.0173)	(0.0172)	(0.0179)	(0.0178)	(0.0506)	(0.0506)	(0.0194)	(0.0194)
L.log_NS_sq					0.000477	0.000503	0.000456	0.000486	0.00984***	0.00984***	0.000304	0.000341
					(0.000408)	(0.000406)	(0.000421)	(0.000420)	(0.00137)	(0.00137)	(0.000480)	(0.000481)
L.OPNS							0.00269	0.00282	-0.0115	-0.0115	0.00294	0.00303
							(0.00622)	(0.00622)	(0.0102)	(0.0102)	(0.00219)	(0.00219)
Sector dummies	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year dummies	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Region dummies	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Constant	0.0253***	0.0251***	0.0161	0.0178	0.462**	0.465**	0.379**	0.383**	1.704***	1.703***	0.439	0.451
	(0.00184)	(0.00182)	(0.0160)	(0.0160)	(0.185)	(0.184)	(0.191)	(0.190)	(0.487)	(0.487)	(0.311)	(0.311)
Observations	38,897	38,897	38,897	38,897	38,897	38,897	37,984	37,984	37,984	37,984	37,984	37,984
R-squared	0.038	0.037	0.056	0.056	0.058	0.058	0.058	0.058	0.122	0.122	0.140	0.141
Number of panelid									4,492	4,492	4,492	4,492

Note: Standard errors clustered at the firm level (although Between-Effects regressions report conventional standard errors). Key to significance levels: *** p<0.01, ** p<0.05, * p<0.1

σ-convergence

Table 5 below shows that the variance of the variable R&D_intensity_gap actually increases over the period 2000-2015; i.e. that $\hat{\sigma}_{2015}^2 > \hat{\sigma}_{2000}^2$. This is observed in all cases: for both the unbalanced and balanced panels, and whether we take 3-digit or 4-digit ICB industry codes. This is in line with the evidence in Figure 4 above. We therefore conclude that there is no σ-convergence in terms of R&D intensities for firms in our sample (i.e. no support for Hypothesis 2b).

Table 5: testing for σ-convergence of R&D_intensity_gap between 2000 and 2015.

		$\hat{\sigma}^2_{2000}$	$\hat{\sigma}^2_{2015}$	N (2000)	N (2015)
4 digit level	**unbalanced**	0.00816	0.00970	833	2409
	balanced	0.00236	0.00328	535	535
3 digit level	**unbalanced**	0.00900	0.01075	833	2409
	balanced	0.00281	0.00305	535	535

5. Conclusions

Do firms in the same sector converge towards the same R&D intensities? We shed light on this issue by analysing panel data on the world's largest R&D investors over the period 2000-2015. We observe considerable heterogeneity in R&D intensities, even among firms in the same sector. Moreover, these differences do not disappear with time. There is no evidence of convergence in R&D intensities over time, for firms in the same sector. In fact, the evidence suggests a mild divergence of R&D intensities during the period of study. Instead of firms having the same behaviour as their industry rivals, instead persistent heterogeneity seems to be the main theme (Dosi and Marengo, 2007).

Our results have implications for the debate about sectoral systems of innovation. While some scholars have suggested that firms in the same sector have similar patterns of innovative activity (Pavitt, 1984; Malerba, 2004), some empirical contributions have found considerable heterogeneity with sectors, and have suggested that the sector of activity has a limited ability to explain differences between firms' innovative behaviours (Leiponen and Drejer, 2007; Srholec and Verspagen, 2012). Our paper provides some modest support for the latter camp, because we also observe persistent heterogeneity in R&D intensities among firms in the same sectors.

Our findings have implications for policy. Previous work has suggested that national R&D intensity targets should be reached by adjusting the industry structure. We present clear evidence of variation in R&D intensities, even for firms in the same sectors, that persists over time. Some firms in low-tech sectors have relatively high R&D intensities. The flipside of the coin is that some firms that are present in high-

tech sectors may have relatively low R&D intensities. We therefore suggest that there are two routes to increasing a country's R&D intensity: policymakers that seek to raise aggregate R&D investment should not focus exclusively on stimulating new firm entry in high-tech sectors, but should also encourage incumbent firms to increase their R&D within their sectors.

References

Aune, F. R., Mohn, K., Osmundsen, P., & Rosendahl, K. E. (2010). Financial market pressure, tacit collusion and oil price formation. Energy Economics, 32(2), 389-398.

Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. Journal of Political Economy, 100(2), 223-251.

Baum, C. F., Caglayan, M., & Talavera, O. (2016). R&D expenditures and geographical sales diversification. The Manchester School, 84(2), 197-221.

Bound J., Cummins C., Griliches Z., Hall B.H., Jaffe A. (1984). Who does R&D and who patents? pp21-54 in: Griliches, Z., (Ed.), R&D, patents, and productivity. Chicago University Press, Chicago.

Carree, M., & Klomp, L. (1997). Testing the convergence hypothesis: a comment. Review of Economics and Statistics, 79(4), 683-686.

Cincera, M., Ravet, J., 2010. Financing constraints and R&D investments of large corporations in Europe and the US. Science and Public Policy 37 (6), 455–466.

Cincera, M., & Veugelers, R. (2013). Young leading innovators and the EU's R&D intensity gap. Economics of Innovation and New Technology, 22(2), 177-198.

Coad A., 2009. The Growth of Firms: a Survey of Theories and Empirical Evidence. Edward Elgar, Cheltenham, UK.

Coad A., Grassano N., (2017). Firm growth and R&D investment: SVAR evidence from the world's top R&D investors. Mimeo.

Cohen, W. M., & Klepper, S. (1992). The anatomy of industry R&D intensity distributions. American Economic Review, 773-799.

Czarnitzki, D., & Toole, A. A. (2011). Patent protection, market uncertainty, and R&D investment. The Review of Economics and Statistics, 93(1), 147-159.

Dosi G., (1988). Sources, Procedures, and Microeconomic Effects of Innovation. Journal of Economic Literature 26 (3), 1120-1171.

Dosi G, Marengo L (2007). On the Evolutionary and Behavioral Theories of Organizations: A tentative roadmap. Organization Science 18 (3), May-June, 491-502.

EPSC (2015). Integration of products and services: Taking the single market into the 21st Century. EPSC strategic notes. Issue 7, 06 October.

Escribano, Á., & Stucchi, R. (2014). Does recession drive convergence in firms' productivity? Evidence from Spanish manufacturing firms. Journal of Productivity Analysis, 41(3), 339-349.

Friedman, M. (1992). Do old fallacies ever die? Journal of Economic Literature, 30 (4), 2129-2132

García-Manjóna, J.V., Romero-Merino, M.E., 2012. Research, development, and firm growth. Empirical evidence from European top R&D spending firms. Research Policy 41 (6), 1084–1090.

Grabowski, H. G., & Baxter, N. D. (1973). Rivalry in industrial research and development: An empirical study. Journal of Industrial Economics, 209-235.

Hall B.H., Jaffe A., Trajtenberg M., (2005). Market Value and Patent Citations. Rand Journal of Economics, 36 (1), 16-38.

Hernandez, H., Hervás Soriano, F., Tubke, A., Vezzani, A., Dosso M., Amoroso A., Grassano N., Coad A., and Gkotsis P. (2016). The 2016 EU industrial R&D investment Scoreboard. Publications Office of the European Union, Luxembourg.

Hervás F., Dosso M., Vezzani A., (2015). Summary report: Fourth IRIMA workshop on "Leading R&D Investors and the European Manufacturing Industry". JRC-IPTS, European Commission. Available from:

http://iri.jrc.ec.europa.eu/documents/10180/414641/4th%20IRIMA%20WS_Summary%20Report

Leiponen A., Drejer I., (2007). What exactly are technological regimes? Intra-industry heterogeneity in the organization of innovation activities. Research Policy 36, 1221-1238.

Lichtenberg, F. R. (1994). Testing the convergence hypothesis. Review of Economics and Statistics 76 (3), 576-579.

Malerba, F. (2004), 'Sectoral systems: how and why innovation differs across sectors,' chapter 14

in J. Fagerberg, D. Mowery and R. R. Nelson (eds), The Oxford Handbook of Innovation. Oxford University Press: Oxford, pp. 380–406.

Mathieu A., van Pottelsberghe de la Potterie B., (2010). A Note on the Drivers of R&D Intensity. Research in World Economy 1(1): 56-65.

Moncada-Paterno-Castello P., Ciupagea C., Smith K., Tubke A., Tubbs M., (2010). Does Europe perform too little corporate R&D? A comparison of EU and non-EU corporate R&D performance. Research Policy 39, 523-536.

Moncada-釐aternò-Castello, P., Vivarelli, M., & Voigt, P. (2011). Drivers and impacts in the globalization of corporate R&D: an introduction based on the European experience. Industrial and Corporate Change, 20(2), 585-603.

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

Montresor, S., & Vezzani, A. (2015). The production function of top R&D investors: Accounting for size and sector heterogeneity with quantile estimations. Research Policy, 44(2), 381-393.

Pavitt, K., (1984). Sectoral patterns of technical change: Towards a taxonomy and a theory. Research Policy, 13, 343-373

Reinstaller, A., & Unterlass, F. (2012). Comparing business R&D across countries over time: a decomposition exercise using data for the EU 27. Applied Economics Letters, 19(12), 1143-1148.

Rodrik, D. (2013). Unconditional convergence in manufacturing. Quarterly Journal of Economics, 165-204.

Scherer, F. M. (1965). Firm size, market structure, opportunity, and the output of patented inventions. American Economic Review, 55(5), 1097-1125.

Scherer F.M., (1967). Research and Development Resource Allocation Under Rivalry. Quarterly Journal of Economics 81 (3), 359-394.

Scott, J. (1984). Firm versus industry variability in R&D intensity. In R&D, patents, and productivity (pp. 233-248). University of Chicago Press.

Shleifer A., (1985). A Theory of Yardstick Competition. RAND Journal of Economics 16 (3), 319-327.

Srholec M, Verspagen B (2012). The Voyage of the Beagle into innovation: explorations on heterogeneity, selection, and sectors. Industrial and Corporate Change 21(5): 1221-1253.

Thompson P., (1999). Rationality, Rules of Thumb, and R&D. Structural Change and Economic Dynamics 10, 321-340

Tornqvist L., Vartia P., Vartia Y.O. (1985). How Should Relative Changes Be Measured? American Statistician 39(1), 43-46.

van Pottelsberghe de la Potterie, B. (2008). Europe's R&D: Missing the Wrong Targets?. Bruegel Policy Brief, 2008/03, February.

Veugelers R., Cincera M., (2010). Europe's missing Yollies. Bruegel Policy Brief 2010/06, August.

Appendix 1: Variables definitions (based closely on Hernandez et al., 2016)

Research and Development (R&D) investment in our dataset is the cash investment funded by the companies themselves. It excludes R&D undertaken under contract for customers such as governments or other companies. It also excludes the companies' share of any associated company or joint venture R&D investment. Being that disclosed in the annual report and accounts, it is subject to the accounting definitions of R&D. For example, a definition is set out in International Accounting Standard (IAS) 38 "Intangible assets" and is based on the "Frascati" manual of the OECD. Research is defined as original and planned investigation undertaken with the prospect of gaining new scientific or technical knowledge and understanding. Expenditure on research is recognised as an expense when it is incurred. Development is the application of research findings or other knowledge to a plan or design for the production of new or substantially improved materials, devices, products, processes, systems or services before the start of commercial production or use. Development costs are capitalised when they meet certain criteria and when it can be demonstrated that the asset will generate probable future economic benefits. Where part or all of R&D costs have been capitalised, the additions to the appropriate intangible assets are included to calculate the cash investment and any amortisation eliminated.

Net sales follow the usual accounting definition of sales, excluding sales taxes and shares of sales of joint ventures & associates. For banks, sales are defined as the "Total (operating) income" plus any insurance income. For insurance companies, sales are defined as "Gross premiums written" plus any banking income.

Operating profit is calculated as profit (or loss) before taxation, plus net interest cost (or minus net interest income) minus government grants, less gains (or plus losses) arising from the sale/disposal of businesses or fixed assets.

Number of employees is the total consolidated average employees or year-end employees if average not stated.

Appendix 2: robustness analysis using an alternative dependent variable: growth of R&D

Table A2.1: Estimates from least squares regressions

VARIABLES	(13) ols gr_RD	(14) ols gr_RD	(15) ols gr_RD	(16) ols gr_RD	(17) fe gr_RD	(18) fe gr_RD	(19) be gr_RD	(20) be gr_RD
L.gap_ind4	0.330***		0.626***		2.112***		0.287***	
	(0.0307)		(0.0445)		(0.110)		(0.0406)	
L.gap_ind3		0.330***		0.617***		2.112***		0.296***
		(0.0306)		(0.0439)		(0.110)		(0.0402)
L.log_NS			-0.134***	-0.134***	-0.909***	-0.909***	0.0572***	0.0557***
			(0.0510)	(0.0507)	(0.293)	(0.293)	(0.0216)	(0.0216)
L.log_NS_sq			0.00239**	0.00240**	0.0181**	0.0181**	-0.00204***	-0.00200***
			(0.00119)	(0.00119)	(0.00719)	(0.00719)	(0.000534)	(0.000534)
L.OPNS			-0.0181*	-0.0179*	-0.00416	-0.00416	-0.0144***	-0.0145***
			(0.0103)	(0.0103)	(0.0147)	(0.0147)	(0.00244)	(0.00244)
Sector dummies	yes	yes	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes	yes	yes	yes
Constant	0.116***	0.116***	1.790***	1.784***	11.01***	11.01***	0.0588	0.0769
	(0.0178)	(0.0178)	(0.550)	(0.546)	(2.957)	(2.957)	(0.346)	(0.346)
Observations	39,214	39,214	38,300	38,300	38,300	38,300	38,300	38,300
R-squared	0.045	0.045	0.065	0.065	0.111	0.111	0.182	0.182
Number of panelid					4,503	4,503	4,503	4,503

Note: Standard errors clustered at the firm level (although Between-Effects regressions report conventional standard errors). Key to significance levels: *** p<0.01, ** p<0.05, * p<0.1

Appendix 3: Robustness analysis using an alternative dependent variable: change in R&D

Table A3.1: Estimates from least squares regressions. Standard errors clustered at the firm level

VARIABLES	(21) ols ch_RD	(22) ols ch_RD	(23) ols ch_RD	(24) ols ch_RD	(25) fe ch_RD	(26) fe ch_RD	(27) be ch_RD	(28) be ch_RD
L.gap_ind4	2.676e+07***		-1.375e+07		1.124e+08***		-3.718e+07***	
	(5.549e+06)		(8.679e+06)		(1.713e+07)		(6.169e+06)	
L.gap_ind3		2.607e+07***		-1.280e+07		1.124e+08***		-3.556e+07***
		(5.512e+06)		(8.507e+06)		(1.713e+07)		(6.125e+06)
L.log_NS			-4.181e+07***	-4.191e+07***	-4.585e+07**	-4.585e+07**	-2.488e+07***	-2.498e+07***
			(1.045e+07)	(1.048e+07)	(2.320e+07)	(2.320e+07)	(3.285e+06)	(3.288e+06)
L.log_NS_sq			1.232e+06***	1.234e+06***	1.302e+06**	1.302e+06**	804,536***	805,239***
			(276,068)	(276,698)	(659,219)	(659,219)	(81,214)	(81,286)
L.OPNS			1.666e+06**	1.651e+06**	909,299	909,299	383,593	362,018
			(848,104)	(840,893)	(800,040)	(800,040)	(371,614)	(371,489)
Sector dummies	yes	yes	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes	yes	yes	yes
Constant	1.281e+07*	1.291e+07*	3.232e+08***	3.244e+08***	3.969e+08**	3.969e+08**	1.702e+08***	1.716e+08***
	(6.631e+06)	(6.629e+06)	(9.935e+07)	(9.968e+07)	(1.977e+08)	(1.977e+08)	(5.262e+07)	(5.265e+07)
Observations	39,214	39,214	38,300	38,300	38,300	38,300	38,300	38,300
R-squared	0.014	0.014	0.039	0.039	0.010	0.010	0.132	0.131
Number of panelid					4,503	4,503	4,503	4,503

Note: Standard errors clustered at the firm level (although Between-Effects regressions report conventional standard errors). Key to significance levels: *** p<0.01, ** p<0.05, * p<0.1

Appendix 4: Robustness analysis

Table A4.1: OLS regressions on subsamples of low-R&D vs high-R&D firms, time periods before or after 2008, samples delimited by their total sales, and with outliers removed (if R&D intensity is greater than 50%, 30% or 20% respectively)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	**OLS low-RD**	**OLS hi-RD**	**OLS pre-08**	**OLS post-08**	**OLS lt1bn**	**OLS gt1bn**	**OLS gt10bn**	**OLS noout 50%**	**OLS noout 30%**	**OLS noout 20%**
	gr_RD_int	**gr_RD_int**	**gr_RD_int**	**gr_RD_int**	**gr_RD_int**	**gr_RD_int**	**gr_RD_int**	**gr_RD_int**	**gr_RD_int**	**gr_RD_int**
L.gap_ind4	3.526***	1.252***	1.005***	0.806***	0.811***	1.467***	1.722***	1.122***	1.445***	1.935***
	(0.194)	(0.0578)	(0.0708)	(0.0645)	(0.0473)	(0.164)	(0.395)	(0.0491)	(0.0591)	(0.0776)
L.log_NS	-0.00191	-0.0402	-0.0651*	-0.00717	-0.0495	0.769***	1.994**	-0.0277*	-0.0193	-0.0230
	(0.0140)	(0.0364)	(0.0383)	(0.0185)	(0.0373)	(0.285)	(0.945)	(0.0168)	(0.0150)	(0.0162)
L.log_NS_sq	-6.97e-05	0.000622	0.00130	-0.000116	0.00138	-0.0168***	-0.0405**	0.000367	0.000179	0.000246
	(0.000332)	(0.000875)	(0.000911)	(0.000435)	(0.00110)	(0.00625)	(0.0194)	(0.000396)	(0.000355)	(0.000383)
L.OPNS	0.0333	0.00243	0.0167	-0.00130	0.00655*	-0.0285	0.0717	0.0768***	0.0980***	0.106***
	(0.0468)	(0.00398)	(0.0124)	(0.00791)	(0.00387)	(0.0663)	(0.0456)	(0.0112)	(0.0114)	(0.0122)
Sector dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Constant	-0.152	0.660*	0.756*	0.0683	0.516*	-8.820***	-24.21**	0.400**	0.305*	0.344**
	(0.152)	(0.383)	(0.405)	(0.194)	(0.312)	(3.222)	(11.46)	(0.179)	(0.159)	(0.174)
Observations	25,764	12,220	13,731	21,088	18,725	19,258	4,787	37,087	35,793	33,251
R-squared	0.142	0.190	0.060	0.073	0.078	0.060	0.057	0.056	0.055	0.056

Note: Standard errors clustered at the firm level. Key to significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table A4.2: OLS regressions (with R&D growth as dependent variable) on subsamples of low-R&D vs high-R&D firms, time periods before or after 2008, samples delimited by their total sales, and with outliers removed (if R&D intensity is greater than 50%, 30% or 20% respectively)

VARIABLES	(1) OLS low-RD gr_RD	(2) OLS hi-RD gr_RD	(3) OLS pre-08 gr_RD	(4) OLS post-08 gr_RD	(5) OLS lt1bn gr_RD	(6) OLS gt1bn gr_RD	(7) OLS gt10bn gr_RD
L.gap_ind4	2.119***	0.799***	0.827***	0.521***	0.805***	0.821***	1.216***
	(0.131)	(0.0604)	(0.0748)	(0.0467)	(0.0453)	(0.0921)	(0.288)
L.log_NS	-0.0895*	-0.237***	-0.362***	-0.0630	0.142***	-0.495***	-0.811*
	(0.0479)	(0.0900)	(0.0837)	(0.0424)	(0.0449)	(0.0851)	(0.460)
L.log_NS_sq	0.00140	0.00477**	0.00780***	0.000702	-0.00638***	0.0104***	0.0161*
	(0.00111)	(0.00214)	(0.00196)	(0.000992)	(0.00122)	(0.00187)	(0.00943)
L.OPNS	0.0875***	-0.0236***	0.00661	-0.0220*	-0.0202*	0.0535**	0.0322
	(0.0111)	(0.00908)	(0.0188)	(0.0115)	(0.0108)	(0.0218)	(0.0979)
Sector dummies	yes	yes	yes	yes	yes	yes	yes
Year dummies	yes	yes	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes	yes	yes
Constant	1.089**	2.973***	4.164***	1.004**	-0.287	5.871***	10.53*
	(0.527)	(0.945)	(0.903)	(0.456)	(0.420)	(0.964)	(5.581)
Observations	25,764	12,220	13,735	21,396	18,725	19,258	4,787
R-squared	0.084	0.156	0.065	0.094	0.087	0.062	0.065

Note: Standard errors clustered at the firm level. Key to significance levels: *** p<0.01, ** p<0.05, * p<0.1

Appendix 5: ANOVA results

Table A5.1: Oneway anova for individual years to decompose the variance of R&D intensity into within-sector and between-sector components

year = 2015						
ICB 4-digit sectors	Analysis of variance					
Source	Sum of squares	df	MS	F	Prob > F	Bartlett p-value
Between groups	12.92	97	0.13	13.69	0	0
Within groups	22.47	2311	0.01			
Total	35.39	2408	0.01			
ICB 3-digit sectors	Analysis of variance					
Source	Sum of squares	df	MS	F	Prob > F	Bartlett p-value
Between groups	9.8	38	0.26	23.89	0	0
Within groups	25.59	2370	0.01			
Total	35.39	2408	0.01			
year = 2004						
ICB 4-digit sectors	Analysis of variance					
Source	Sum of squares	df	MS	F	Prob > F	Bartlett p-value
Between groups	15.36	98	0.16	13.97	0	0
Within groups	31.15	2777	0.01			
Total	46.51	2875	0.02			
ICB 3-digit sectors	Analysis of variance					
Source	Sum of squares	df	MS	F	Prob > F	Bartlett p-value
Between groups	12.29	76	0.16	13.62	0	0
Within groups	32.34	2725	0.01			
Total	44.63	2801	0.02			

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

Joint Research Centre

EU Science Hub