UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO DI SCIENZE ECONOMICHE E SOCIALI

# PERSISTENCE OF R&D INTENSITIES IN THE WORLD'S TOP INVESTORS IN R&D

CLAUDIA PIGINI   ALESSANDRO STERLACCHINI   FRANCESCO VALENTINI

QUADERNO DI RICERCA n. 480

ISSN: 2279-9575

*June 2023*

## Abstract

There is extensive empirical evidence of a within-sector heterogeneity in terms of firms' R&D intensity (share of expenditures on sales) which, moreover, does not converge to a common level over time. Using a balanced panel of the world's top R&D investors, we first investigate whether there is a different degree of time persistence along the R&D intensity distribution. Secondly, we analyse whether the persistence in and the transition to different levels are heterogeneous between four R&D-intensive sectors. As a general result, we find that companies starting with low R&D intensities are more likely to move towards the sector medium levels than those exerting a high innovative effort, which persist in the right tail of the distribution. With the exception of the Pharmaceutical sector, company size affects negatively (positively) the persistence and the entry rate into the top (bottom) 20% of the R&D intensity distribution. Differences across sectors emerge with respect to the impact of other company characteristics (profitability, capital investment, and location).



**JEL Class:** O3, L2

**Keywords:** Innovation persistence, R&D intensity, High-tech sectors, Large companies.



**Address:** Alessandro Sterlacchini, corresponding author. Università Politecnica delle Marche, Department of Economics and Social Sciences, Piazzale Martelli, 8, 60121, Ancona, Italy. Email address: a.sterlacchini@staff.univpm.it.

# Persistence of R&D intensities in the world's top investors in R&D

Claudia Pigini, Alessandro Steralcchini and Francesco Valentini

## 1 Introduction

A stylized fact that emerges from the empirical literature is the presence of within-sector heterogeneity in terms of firms' R&D intensity (Cohen and Klepper, 1992), which is characterised by a very similar distribution across different industries. Another relevant feature stressed by more recent studies is that such heterogeneity is persistent over time, i.e., in each sector the R&D intensity does not converge to a common level. Klette and Johansen (2000) show that firms whose innovative effort is the lowest or highest in their sample tend to keep on their R&D intensity at the same level. Coad (2019) provides further evidence of persistent within-sector heterogeneity highlighting that, in the same sector, the variation of firms' R&D expenditures on sales does not decrease over time.

In this paper, we take the analysis further in two directions: first, we investigate whether there is a different degree of time persistence along the R&D intensity distribution, that is firms whose intensity of innovative effort is high (low) might be more or less likely to maintain the same level over time than firms characterized by a low (high) intensity; secondly, we analyse whether the possible R&D intensity persistence is heterogeneous across four R&D-intensive sectors, namely Electronic & Electrical Equipment, Pharmaceuticals & Biotechnology, Software & Computer Services and Technology Hardware & Equipment.

Using a balanced panel of large companies included in the 2014-2020 waves of the EU Industrial R&D Investment Scoreboard, we estimate a dynamic ordered probit model in order to investigate the evolution of the R&D intensity from an initial to a steady state distribution for each sector: the comparison between these two states is informative of whether R&D intensity has converged towards a within-sector common level or not. We further examine the presence of within-sector heterogeneity in the transitions in and out of the top and bottom 20% of the R&D intensity distribution by means of sector-specific transition probability models.

From the first part of the analysis, we find that the overall observed persistence (or the absence of convergence) is mainly due to the fact that companies with the highest intensities of R&D tend to maintain this behaviour over time. This finding is discussed in the light of the main hypotheses adopted in the literature to explain the persistence of innovative activities. With respect to the transition probability model, some interesting differences across sectors emerge. In general, the probability to persist and to enter in the top (bottom) 20% of the R&D intensity distribution is negatively (positively) affected by the companies' sales. Due to the specificity of its research and innovation process, the only sector in which the firm size does not play a significant role is that of Pharmaceuticals & Biotechnology. Other companies' characteristics (profitability, capital investment, and geographical location of companies' headquarters) play a minor role and only in particular sectors.

The structure of paper is the following. Section 2 reviews the empirical literature dealing with the issue of innovative persistence and highlights the main research questions addressed with our work. Section 3 describes the data and the empirical strategy is illustrated in Section 4. Section 5 presents the estimation results which are discussed in Section 6.

## 2 Literature review and research questions

In examining the innovation persistence at firm level, the literature has stressed three potential explanations (Le Bas and Scellato, 2014; Arroyabe and Schumann, 2022). A first one is based on the hypothesis that "success breeds success": successful innovators continue innovating thanks to greater market power and profits which reduce financial constraints and increase the capability to invest in risky innovative activities and, hence, to exploit a wider range of technological opportunities. A second explanation, mainly stressed by evolutionary scholars, relies upon the knowledge accumulation hypothesis. Innovation experience is associated with dynamic increasing returns due to learning by doing and learning to learn phenomena. For instance, after the launch of an entirely new product (radical innovation) a company can introduce incremental improvements which make the product more attractive to consumers. On the other hand, by investing in R&D activities a firm increases its stock of knowledge and, thus, its capability to discover and/or absorb new pieces of knowledge from external sources. A third argument refers to the fact that in performing R&D activities companies incur in sunk costs, so that the decision to engage in R&D today constraints the firm to continue to do so in the future (Manez et al., 2009). Along with entry barriers, sunk costs in R&D give rise to barriers to exit and, thus, innovation persistence.

From the empirical point of view, a first strand of studies has examined the innovation persistence at the firm level by using patent data. In this case, a low degree of patenting persistence emerged since only a small group of large firms characterised by a higher propensity to patent were found to be persistently innovative (Geroski et al., 1997; Cefis, 2003). A second group of studies have used survey data on innovations (such as the Community Innovation Survey). In this regard, the most common result is that there is persistence of innovation by firms especially if they belong to high-tech industries (Raymond et al., 2010) and introduce product rather than process innovations (Le Bas and Scellato, 2014; Tavassoli and Karlsson, 2015). Antonelli et al. (2012) stress the important role played by the sunk-cost nature of R&D expenditures by finding that the highest level of persistence in product innovations arises for R&D performing firms. Similar findings are achieved by Clausen et al. (2012) who observe that, as far as R&D-intensive strategies yield more radical innovations, a higher persistence could also be due to success-breads-success dynamics and learning effects.

Accordingly, although the hypothesis of "success breeds success" appears to be more outcome-oriented than the other two, R&D investment (i.e. an input of the innovation process) can be viewed as a crucial element for explaining innovation persistence. Thus, a third group of studies has focussed on the persistence of R&D activities. Peters (2009) analyses both a balanced and unbalanced panel of German firms. By using a dummy variable for the engagement in R&D and a dynamic random effect discrete choice model, her results support the hypothesis of true state dependence and, hence, R&D persistence. For a balanced panel of Spanish firms Manez et al. (2009) consider a set of binary variables for the decision to undertake R&D by firms of different size and belonging to low-, medium- and high-tech industries. By means of a dynamic multivariate probit model, they find that R&D persistence is higher for large firms and/or firms in high-tech industries. Again with panel data for Spanish firms, Arqué-Castells (2013) and Triguero and Córcoles (2013) confirm the presence of state dependence in the R&D choice. Moreover, the latter finds that the persistence in R&D is higher than that in innovation output and, with respect to firm specific characteristics, only size and the outsourcing of production stages exert a positive impact in both cases.

Woerter (2014) considers the share of R&D expenditures on sales for an unbalanced panel of Swiss firms observed in five periods, that is every three years starting from 1996 up to 2008. The continuous dependent variable only refers to R&D performing firms and, controlling for the selection in doing R&D or not, the dynamic model is estimated with a Generalized Least Square method with random effects. The results support the state dependence of the R&D intensity and also show that its persistence is higher for firms competing with a few main rivals. Using data taken from the EU Industrial R&D Investment Scoreboard over the period 2000-2015, Coad (2019) estimates a growth model for the R&D intensity of companies and provides evidence of persistent heterogeneity within sectors. In particular, his empirical analysis highlights the absence of $\sigma$-convergence, meaning that the between-firms variation of R&D intensity does not decrease over time.

To our knowledge, only these two studies assess the presence of R&D intensity persistence by means econometric analyses. This is not without consequence since here the focus is not on the firms' decision to invest in R&D but on how much to invest with respect to their sales.

Regarding the latter choice, a stylised fact arising from empirical studies is that "Within industries, among R&D performers, R&D expenditures rise monotonically with firm size across all firm size ranges, with firm size typically explaining well over half the intraindustry variation in R&D " (Cohen and Klepper, 1996, p. 929). Accordingly, firm size does not significantly affect R&D intensity (Cohen et al., 1987; Sterlacchini, 1994): with the partial exception of small firms, the impact of size becomes insignificant or even negative moving to medium-sized and large firms (Barge-Gil and López, 2014; Galaasen and Irarrazabal, 2021).

A more significant role for explaining the heterogeneity of R&D intensity across firms has been ascribed to industry differences in market structure, demand changes, and, especially, technological opportunities and appropriability conditions (Levin et al., 1985). The latter variables are usually captured by some proxies. Technological opportunities increases when the firms' research activities are closer to scientific advancements and exploit external sources of knowledge. Appropriability, instead, mainly refers to the mechanisms used by firms to reduce the scope for imitation (such as patents) and, hence, internalize the benefits of innovations. However, these variables have been effective in explaining differences in R&D intensity between rather than within industries.

Hence, with respect to within-industry heterogeneity there have been some attempts to identify firm-level determinants of R&D intensity such as profitability, degree of diversification, ownership and managerial characteristics, risk-taking propensities and corporate governance practices but the scattered empirical evidence in this regard has been rather inconclusive. For instance, only in some cases a positive relationship between profitability and innovative efforts has been found. Opposite results have emerged from other studies, stressing that firms increase R&D investment when profitability falls below the industry average (Antonelli, 1989). By considering corporate governance practices, such as limitations of voting rights restrictions and managers' remuneration based on financial performance, Honoré et al. (2015) find that they are negatively correlated with the R&D intensities of large European corporations. However, as these same authors acknowledge in reviewing the empirical literature, previuos studies have found contrasting results.

The inconsistency of empirical evidence has prompted some scholars to take an alternative route. Cohen and Klepper (1992) showed that the industry distributions of firm R&D intensities are unimodal and skewed with a long tail to the right: most of the values are concentrated at low levels, but there are few firms with large shares of R&D expenditures with respect to their size (a regularity found by several subsequent studies: see, among others, Sterlacchini, 1994; Klette and Johansen, 2000; Galaasen and Irarrazabal, 2021). Cohen and Klepper (1992) developed a model in which unobservable R&D-related capabilities are allocated probabilistically among firms within an industry. Then, they showed that a binomial distribution of these capabilities could account for the distributional regularities of firm R&D intensities.

According to some "static" assumptions, this probabilistic model was applied to single cross sections of firms. Hence, the model is less suitable for industries that, due to the emergence of new technological opportunities, may experience significant changes in their R&D intensity distribution over time. More specifically, an issue that Cohen and Klepper's model neglects is whether the firms with the lowest or highest ratios of R&D expenditures on sales are the same or do change over time.

To put it another way, is there persistence of the firms placed in the left and right parts of the R&D intensity distribution? Are there other firm characteristics affecting the likelihood of keeping low or high shares of R&D expenditures on sales? These are the main research questions that this paper addresses.

Klette and Johansen (2000) argue that, according to the usual framework for computing the knowledge (or R&D) capital of a company, coupled with the standard model in which the increase of this capital affects Total Factor Productivity,[1] profit maximising firms in the same industry should converge to the same R&D intensity (see also Nelson, 1988). A further, related, argument in favour of such a convergence is that firms producing similar goods or services rely on the same knowledge base and face the same technological opportunities. Moreover, and especially in R&D-intensive industries, a convergence process could occur if, for choosing the intensity of R&D efforts, firms observe the behaviour of their competitors and adjust their efforts towards the industry mean (Coad, 2019).

This prediction is at odds with "the widely observed pattern that the same firms tend to persistently carry out above (or below) average amounts of R&D, say, relative to their sales" (Klette and Johansen, 2000) p. 392). Consistently with this evidence, the authors introduce a different model of knowledge accumulation in which past knowledge capital

[1]We refer to the Cobb-Douglas production function augmented with R&D capital, that is $Q_{it} = A_{it}C_{it}^{\alpha}L_{it}^{\beta}K_{it}^{\gamma}$ where for firm $i$ in year $t$, $Q$ is output, $A$ is the TFP term, $C$ is physical capital, $L$ the labour input and $K$ the knowledge capital. By employing the perpetual inventory method (as for physical capital), the beginning-of-period knowledge capital is computed as $K_{it+1} = K_{it}(1-\delta) + RD_{it}$ where $\delta$ is the depreciation rate and $RD$ is the R&D investment of firm $i$ in year $t$.

makes the current R&D effort "more productive" in the generation of additional capital.[2] With respect to the three main explanations for innovation persistence (summarised at the beginning of this section), it can be said that Klette and Johansen's approach jointly relies upon the hypotheses of "knowledge accumulation" and "success breeds success". The authors also test their model with a panel of Norwegian firms examined over the period 1980-1992. Together with the skewed distributions of R&D intensities across industries (see above), they show that within industries the same firms persistently invest more in R&D with respect to their sales: more than 60% of the firms in the highest quartile of R&D intensity remain in the same quartile two years later.

Along with the already mentioned explanations, persistent heterogeneity in R&D intensities may emerge for further reasons (Coad, 2019). Firms do not have identical capabilities to successfully commercialise the outcomes of innovative activities. Large and especially multiproduct firms can be more able than their smaller counterparts to translate R&D activities into new products and services (Baysinger and Hoskisson, 1989), to exploit positive demand shocks (Galaasen and Irarrazabal, 2021) and, hence, to spread the R&D costs over greater sales (Cohen and Klepper, 1996). More generally, firms adopt different strategies and R&D choices according to their level of product diversification, market shares and age. For instance, as opposed to long-standing incumbents, younger firms should invest more in R&D with respect to their sales for several years if they want to become part of the stable core of the industry.

# 3 Data and variables

## 3.1 Description and summary statistics

The data and variables used for our empirical analysis are taken from the 2014-2020 waves of the EU Industrial R&D Investment Scoreboard (henceforth *Scoreboard*; for the last two editions see Hernández et al., 2019 and Grassano et al., 2020). Since 2014 the *Scoreboard* has been providing data for the top 2500 R&D investors in the world.[3] Company data are available for R&D and capital investments, sales, operating profits, and employment, and refer to previous years (i.e. from 2013 to 2019). For different purposes and using different time frames, these data have been used in several previous studies. Along with Coad (2019), who inspects the persistent heterogeneity of R&D intensities as in our study, see, among others, Cincera and Veugelers (2013), Cincera and Ravet (2014), Honoré et al.

[2]The equation for the accumulation of knowledge capital is $K_{it+1} = K_{it}^{(\rho-\nu)} RD_{it}^{\nu}$ where $\nu$ captures the productivity of R&D in generating new knowledge and $\rho$ is a parameter reflecting increasing returns in knowledge production.

[3]Previous editions, from 2006 to 2013, included data for 2000 companies.

(2015), Montresor and Vezzani (2015), Ortega-Argilés et al. (2015), Moncada-Paternò-Castello (2022).

The amount of R&D by these global companies accounts for more than 60% of the total expenditure on R&D worldwide and about 90% of the world's business-funded R&D (Grassano et al., 2020). Regarding the geographical distribution, the reference countries are those in which the headquarters of these multinationals are located and include the R&D performed abroad by their subsidiaries. For the sectoral distribution, instead, the *Scoreboard* refers to the main sector of activity indicated by the same companies in their annual reports.[4]

Being the persistence of R&D intensity the topic of our study, we focus upon the sectors characterized by the highest percentages of R&D expenditures on sales: Pharmaceuticals & Biotechnology (PHB, hereafter, with an R&D intensity of 15.5% in 2019), Software & Computer Services (SCS, 11.8%), Technology Hardware & Equipment (THE, 9%) and Electronic & Electrical Equipment (EEE, 5.1%). All together these four sectors account for more than 57% of the total R&D expenditures of the *Scoreboard* companies and, hence, more than half of the business R&D worldwide. A secondary reason to concentrate the analysis on these sectors is the high number of companies (observations) for each of them.[5]

Using the seven (2014-2020) waves of the *Scoreboard*, a panel dataset can be built by linking companies over time through their registered name. Yet the simple procedure of appending waves leads to a mismatch of companies whose names have changed from one year to the next. For instance, if we just used the company name as a panel identifier, `DESCARTES SYSTEMS` and `THE DESCARTES SYSTEMS` would result into two different companies when in fact the mismatch is just due to misreporting. In other cases, we may have companies that change names because of mergers occurring over the years. In order to properly handle such cases, we pre-processed company names by performing fuzzy string matching (Zobel and Dart, 1995; Christen, 2006).[6] The approximate string matching algorithm we employ provides a similarity score between pairs of strings ranging from 0 to 1, which is achieved in case of perfect similarity. We individually checked the instances in

[4]Since most of the companies included in the *Scoreboard* consist of diversified conglomerates, the imputation of a unique sector of activity represents a clear limitation of the database and, a fortiori, of our study. For instance, the sector ascribed to some companies could change over time: this occurred to only 15 companies included in our balanced panel for 2013-2019; in these cases we imputed to them the prevalent (modal) sector.

[5]In terms of the absolute amount of R&D investment another relevant sector is that of "Automobile & Parts" (together with the four mentioned sectors it accounts for about 72% of the total R&D reported in the *Scoreboard*). However, with a share of R&D expenses on sales lower than 5%, this sector is not classified as R&D-intensive. Moreover, as compared with the selected sectors, the number of included companies is lower.

[6]Fuzzy string matching is implemented using the Stata module `matchit` provided by Raffo (2015). Strings are parsed into single and collocated tokens.

which the score was greater than 0.5 and strictly smaller than 1 for potential misreporting and, then, replaced the strings.

To achieve the final sample, we removed the companies reporting an intensity of R&D expenditures on sales equal or greater than 100%, a procedure also adopted by Cincera and Ravet (2014), Coad (2019) and Galaasen and Irarrazabal (2021). The companies with abnormally high R&D intensities are particularly concentrated in PHB, a sector that requires firms to invest heavily in R&D but in which sales may be low for several years until new products (e.g. drugs) can be successfully introduced (this peculiar feature will be further stressed in Section 6).

It should be added that, over the 2014-2010 waves of the *Scoreboard*, most of these companies are present in a few years only. Indeed, by including the 2500 top R&D investors of the world for each year, in the fringe of the ranking there are companies that entered in the *Scoreboard* and exited after one or a few years.[7] In order to estimate whether there is persistent heterogeneity or convergence in terms of R&D intensity these companies cannot be taken into account. By removing them we ended up with a balanced panel of large transnational corporations for the years 2013-2019. Being involved in global competition processes, these companies, permanently present in the *Scoreboard*, should be more likely to check the behaviour of direct competitors and, then, to adjust their intensity of R&D investment toward the sectoral mean or mode (cf. Section 2): hence, it should be more likely to observe a process of convergence within different sectors.

The first two rows of Table 1 report the number of companies and observations for each sector considered. For these companies, it is available a complete series of data concerned with: R&D investment (M €); Sales (M €); R&D intensity over sales (percentage); Profitability (percentage of operating profits over sales); Capex intensity (percentage of expenditures for tangible capital over sales) and geographical location distinguished in five groups of countries/regions: USA (US), Europe, China, Other Eastern Asian countries (East Asia), Rest of the World (RoW).[8] Due to many missing values, employment data are not used to preserve an adequate number of observations for each sector. Hence, at this stage, the analysis is based upon a limited set of variables for which descriptive statistics are reported in Table 1.

The sector with the highest average R&D intensity is SCS, followed by PHB and THE with the first two sectors in the *Scoreboard* ranking having maximum values close to 100%.[9] With the exception of EEE, all the sectors show an intensity of R&D much

[7]Regarding the companies that exited from the *Scoreboard*, another important reason is that other companies acquired them or, in any case, they were no more independent (Grassano et al., 2020). With respect to the role of Mergers & and Acquistions in our context see the discussion in Section 6.

[8]A detailed list of the countries included in each area is reported in Table 10 in Appendix A

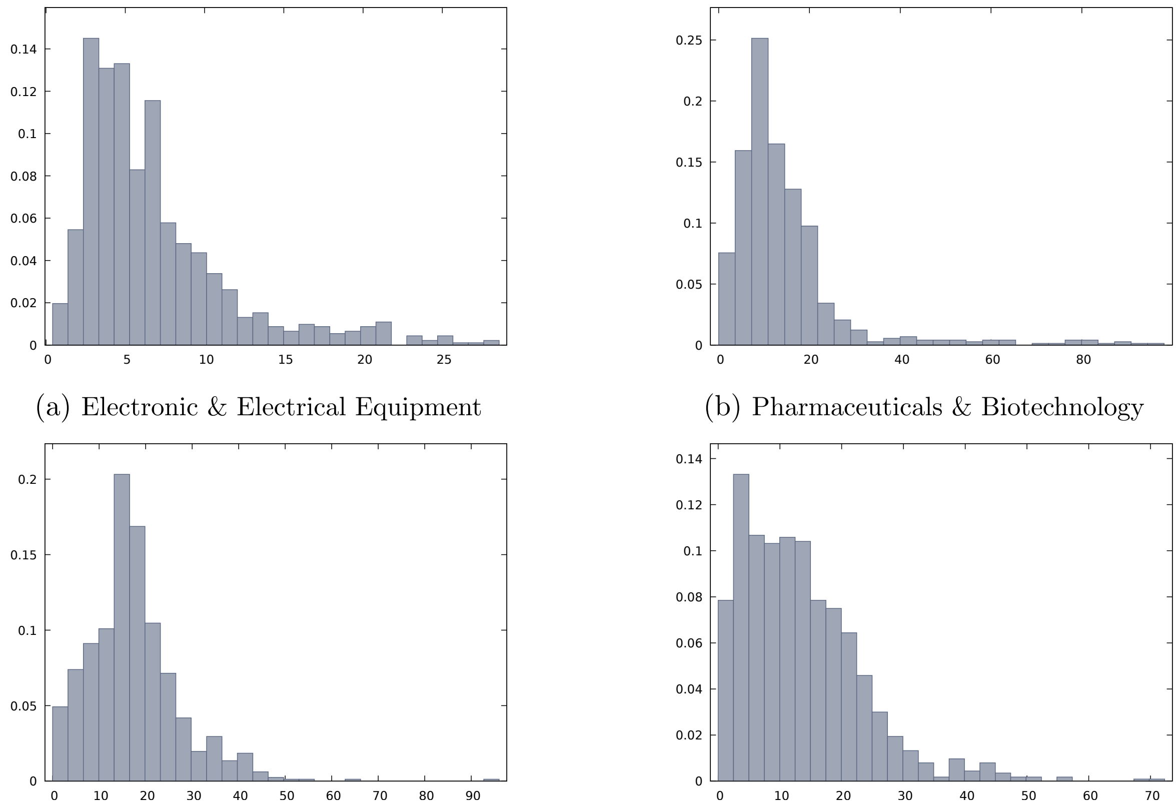
[9]The sectoral ranking is not consisted with that previously mentioned in the text because in Table 1

Table 1: Descriptive statistics by sector

| | EEE | PHB | SCS | THE |
|---|---|---|---|---|
| Number of companies | 131 | 104 | 116 | 162 |
| Number of observations | 917 | 728 | 812 | 1134 |
| | R&D intensity | | | |
| Mean | 6.811 | 14.49 | 17.24 | 13.34 |
| St. Dev. | 4.769 | 13.32 | 9.767 | 9.778 |
| Min | 0.900 | 0.200 | 0.200 | 0.200 |
| Max | 28.10 | 94.70 | 92.90 | 70.00 |
| | R&D expenditure (M €) | | | |
| Mean | 258.3 | 1003 | 501.4 | 528.7 |
| St. Dev. | 565.9 | 1942 | 1476 | 1351 |
| Min | 15.60 | 16.60 | 16.90 | 16.20 |
| Max | 6086 | 10753 | 17152 | 14436 |
| | Sales (B €) | | | |
| Mean | 6.133 | 6.872 | 4.495 | 6.040 |
| St. Dev. | 15.23 | 12.14 | 12.641 | 17.49 |
| Min | 0.081 | 0.043 | 0.058 | 0.047 |
| Max | 158.0 | 73.05 | 127.3 | 232.0 |
| | Capex intensity | | | |
| Mean | 5.135 | 6.106 | 4.609 | 6.630 |
| St. Dev. | 5.073 | 5.886 | 5.509 | 9.270 |
| Min | 0.000 | 0.000 | 0.000 | 0.000 |
| Max | 51.90 | 54.40 | 50.70 | 94.60 |
| | Profitability | | | |
| Mean | 9.765 | 12.72 | 10.41 | 8.513 |
| St. Dev. | 9.264 | 20.40 | 24.13 | 15.59 |
| Min | -74.50 | -162.0 | -111.6 | -197.3 |
| Max | 55.60 | 68.00 | 430.0 | 65.70 |
| | Geographical areas (frequency) | | | |
| US | 0.229 | 0.225 | 0.557 | 0.383 |
| East Asia | 0.382 | 0.262 | 0.052 | 0.368 |
| China | 0.154 | 0.139 | 0.106 | 0.076 |
| Europe | 0.230 | 0.361 | 0.224 | 0.142 |
| RoW | 0.005 | 0.013 | 0.061 | 0.031 |

**Notes:** EEE = Electronic & Electrical Equipment; PHB = Pharmaceuticals & Biotechnology; SCS = Software & Computer Services; THE = Technology Hardware & Equipment. R&D intensity, Capex intensity and Profitability are expressed in percentage points.

Figure 1: Distribution of R&D Intensity across sectors



(a) Electronic & Electrical Equipment

(b) Pharmaceuticals & Biotechnology

(c) Software & Computer Services

(d) Technology Hardware & Equipment

higher than that of (tangible) capital investment and very large variations in terms of profitability.

## 3.2 R&D intensity: a closer look

Figure 1 shows the distribution of the R&D intensity of the firms in our sample for the four sectors considered (for each firm all the values for the years 2013-2019 are included). Consistently with the regularity stressed by the relevant literature (Cohen and Klepper, 1992), all the distributions are markedly asymmetric and exhibit a sizable right-skeweness.

Since the distributions above give us a static representation of the nature of the R&D intensity of the firms, we expand our analysis as follows: we categorize the R&D intensity into sector-year specific quintiles and we compute the empirical transition probability matrix across the five classes, for each sector. The frequencies, standardized by row, are summarized in Table 2. For each sector, the panels report the probability of moving from

the sample averages rather than the weighted ones are considered.

Table 2: Sample transition probabilities

| EEE | | | | | | |
|---|---|---|---|---|---|---|
| | | $y_{it}$ | | | | |
| | | 1 | 2 | 3 | 4 | 5 |
| $y_{i,t-1}$ | 1 | 90.7 | 6.8 | 1.9 | 0.6 | |
| | 2 | 9.6 | 78.3 | 11.5 | | 0.6 |
| | 3 | 0.7 | 12.4 | 71.9 | 15.0 | |
| | 4 | | 1.9 | 14.4 | 71.9 | 11.9 |
| | 5 | | | 0.6 | 12.3 | 87.1 |

| PHB | | | | | | |
|---|---|---|---|---|---|---|
| | | $y_{it}$ | | | | |
| | | 1 | 2 | 3 | 4 | 5 |
| $y_{i,t-1}$ | 1 | 85.3 | 13.2 | 1.6 | | |
| | 2 | 14.4 | 68.8 | 14.4 | 2.4 | |
| | 3 | 0.8 | 15.4 | 64.2 | 19.5 | |
| | 4 | 0.8 | 2.4 | 15.9 | 63.5 | 17.5 |
| | 5 | | 0.8 | 1.7 | 16.5 | 81.0 |

| SCS | | | | | | |
|---|---|---|---|---|---|---|
| | | $y_{it}$ | | | | |
| | | 1 | 2 | 3 | 4 | 5 |
| $y_{i,t-1}$ | 1 | 90.8 | 8.5 | 0.7 | | |
| | 2 | 8.0 | 75.4 | 14.5 | 2.2 | |
| | 3 | 0.7 | 14.2 | 67.4 | 15.6 | 2.1 |
| | 4 | 0.7 | 2.2 | 19.0 | 66.4 | 11.7 |
| | 5 | | | | 14.4 | 85.6 |

| THE | | | | | | |
|---|---|---|---|---|---|---|
| | | $y_{it}$ | | | | |
| | | 1 | 2 | 3 | 4 | 5 |
| $y_{i,t-1}$ | 1 | 94.4 | 5.6 | | | |
| | 2 | 5.2 | 85.0 | 9.8 | | |
| | 3 | 0.5 | 8.8 | 76.3 | 13.4 | 1.0 |
| | 4 | | 0.5 | 14.9 | 70.3 | 14.4 |
| | 5 | | | | 14.9 | 85.1 |

a given state at time $t-1$ (rows) to a state at time $t$ (columns). For example, looking at sector THE, a firm in the first quintile at time $t-1$ has a 94.4% probability to remaining in the same quintile and a 5.6% to move in the second quintile. Overall, consistently with Klette and Johansen (2000), we observe a strong stickiness across quintiles, i.e., firms tend to persist in the same part of the distribution over years.

# 4 Methodology

As stated in Section 2, the purpose of our analysis is to investigate whether there is a different degree of time persistence along the R&D intensity distribution. To this aim we categorize the intensity measure so as to specify a discrete-state dynamic model that allows us to evaluate the probability of transitioning in and out of different parts of the R&D intensity distribution. We take categories based on the sector-year-specific quintiles of R&D intensity distribution and characterize them as the states of a first-order Markov process, which leads to the specification of a dynamic ordered model.

Let $y^*_{it}$ denote the R&D intensity for firm $i = 1, \ldots, n$ at time $t = 1, \ldots, T$, in each of the sectors considered. We discretise $y^*_{it}$ in 5 categories as

$$y_{it} = j \quad \text{if} \quad c_j < y^*_{it} \leq c_{j+1},\ j = 0, \ldots, 4, \tag{1}$$

where $c_j$ is the $j$-th sector-year quintile of $y^*_{it}$, with $c_0 = -\infty$ and $c_5 = +\infty$, $\forall\ t$. The probability of $y_{it}$ being equal to $j$, for $j = 0, \ldots, 4$ is

$$\pi_{itj} = \Phi(c_{j+1} - \mu_{it}) - \Phi(c_j - \mu_{it}) \tag{2}$$

where $\Phi(\cdot)$ is the standard normal distribution function and

$$\mu_{it} = \sum_{j=0}^{J} \gamma_j \mathbf{I}(y_{i,t-1} = j) + \alpha_i + \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta}, \tag{3}$$

where $\mathbf{I}(\cdot)$ is an indicator function and the normalization $\gamma_0 = 0$ applies. In the above formulation, the parameters $\gamma_j$s capture the persistence in the R&D intensity, as they are related to the lags of the dependent variable, $\alpha_i$ is the company-specific unobserved heterogeneity, vector $\boldsymbol{x}_{i,t-1}$ collects the lagged values of the regressors described in Section 3, so as to avoid simultaneity issues, and $\boldsymbol{\beta}$ is the vector of related parameters. As for the dependent variables, along with dichotomisations of $y_{i,t-1}$, we include one lag of three firms' characteristics: the logarithm of sales, the profitability index and the Capex intensity, both the latter expressed in percentages. Further, we include four geographical area fixed-effects[10] (we leave US as the reference region) and year fixed-effects (reference is 2014) to control for potential heterogeneity.

Consistent estimation of the transition probabilities rests on properly identifying the so-called *true* state dependence, i.e., how having been in a certain state in $t-1$ affects the probability of being in that same state at time $t$. The sources of such time persistence,

[10]In some model specifications, the RoW and China dummy variables have been omitted due to quasi-collinearity.

which are broadly discussed in Section 2, have to be disentangled from the permanent unobserved heterogeneity $\alpha_i$, that affects the propensity of having a certain level of R&D intensity at all times (Heckman, 1981a). To this aim, we rely on a random-effects approach where we assume that $\alpha_i$ is normally distributed, with zero mean and variance $\sigma^2_\alpha$, and independent of $\boldsymbol{x}_{i,t-1}$. Because only a limited set of variables is available in the *Scoreboard*, $\alpha_i$ might embed information that is not properly captured by the model regressors, thereby poorly approximating the time-invariant component of $y_{it}$. We attempt to alleviate this issue by including the level of the R&D expenditure at time $t = 0$, i.e., 2013, denoted as $R\&D_{2013}$, in the set of regressors. Notice that this information is exogenous with respect to $y_{it}$ in (1), for which only the information in $t = 1, \ldots, T$ is used.[11]

The dynamic structure of the model poses the so-called "initial-conditions" problem: the correlation between $y_{i,t-1}$ and $\alpha_i$ requires that the process is initialized by specifying a conditional distribution for $y_{i0}$ given $\alpha_i$. We address this issue following Heckman (1981b)[12] and specify an additional set of probabilities for $y_{i0}$ as in Equation (2) where the linear index becomes

$$\mu_{i0} = \theta\alpha_i + \boldsymbol{x}'_{i0}\boldsymbol{\lambda}. \tag{4}$$

The parameters $\theta$ and $\boldsymbol{\lambda}$ are estimated along with $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_J)'$, $\boldsymbol{\beta}$, and $\sigma^2_\alpha$ by maximum likelihood relying on numerical integration methods.

The dynamic ordered probit model allows us to investigate the evolution of R&D intensity from an initial set of probabilities of belonging to each category to a set of steady state probabilities, representing the long-term behavior. Let us denote as $\boldsymbol{\pi}_{it} = (\pi_{it0}, \ldots, \pi_{itJ})'$ the vector collecting the probabilities $j$ defined in (2). Also define the transition matrix $\boldsymbol{P}_i$ with typical element $(s, j)$ the conditional probability

$$Pr(y_{it} = j | y_{i,t-1} = s) = \Phi\left(\frac{c_{j+1} - \gamma_s - \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta}}{\sqrt{1+\sigma^2_\alpha}}\right) - \Phi\left(\frac{c_j - \gamma_s - \boldsymbol{x}'_{i,t-1}\boldsymbol{\beta}}{\sqrt{1+\sigma^2_\alpha}}\right).$$

Then, under standard regularity conditions on the Markov model,[13] the set of probabilities

[11]We also considered and discarded the fixed-effects approach as the remarkably high degree of persistence in the dependent variable gave rise to quasi-complete separation problems and quasi-collinearity issues when company-specific dummies were included. Any alternative strategy based on model transformations to eliminate $\alpha_i$ would not allow us to compute predicted probabilities.

[12]The alternative solution provided by Wooldridge (2005), which consists of including the dichotomisations of $y_{i0}$ in (3) is again unfeasible in our case because of the high persistence in the R&D intensity generating quasi-collinearity between $y_{i,t-1}$ and $y_{i0}$. In a way, the inclusion of the initial value of the R&D expenditure acts as a linear approximation in the spirit of Wooldridge (2005).

[13]The Markov model is time-homogeneous, i.e., the transition probabilities from $t-1$ to $t$ are the same for all $t$, and the Markov chain must be irreducible and aperiodic.

after the $h$-th transition between the states can be obtained as

$$\boldsymbol{\pi}_{ih} = \boldsymbol{P}_i \boldsymbol{\pi}_{i,h-1} = \boldsymbol{P}_i^h \boldsymbol{\pi}_{i0}. \tag{5}$$

The steady–state probabilities, denoted as $\boldsymbol{\pi}_i$, are such that $\boldsymbol{P}_i \boldsymbol{\pi}_i = \boldsymbol{\pi}_i$, meaning that transitions no longer affect the probability of R&D intensity being in each of the categories in the long term. Operatively, we compute the set of estimated steady-state probabilities using the last term of expression (5) as $\hat{\boldsymbol{\pi}}_i = \hat{\boldsymbol{P}}_i^h \hat{\boldsymbol{\pi}}_{i0}$, with $h = 20$ chosen so that $\hat{\boldsymbol{\pi}}_i$ is stabilized across steps and where $\hat{\boldsymbol{P}}_i$ and $\hat{\boldsymbol{\pi}}_{i0}$ are evaluated at the maximum likelihood estimates of the parameters.

Looking for convergence towards the middle of the R&D intensity distribution is therefore rather straightforward if we compare the initial and steady state estimated probabilities. On the other hand, a dynamic ordered model does not easily lend itself to the identification of within sector heterogeneity in the transitions in and out of the categories at time $t$, especially if we also want to investigate how such heterogeneity differs for firms who belonged to specific categories in $t-1$.[14] In order to study the effects of covariates on the transition probabilities, we rely on simpler models where only one dichotomisation is considered at the time. Specifically, we study the probability of belonging to the bottom and top 20% of the R&D intensity distribution. Considering the former, for $t = 0, \ldots, T$, let

$$d_{it} = \mathbf{I}(y_{it} = 0), \quad \text{with} \quad \pi_{it} = \Phi(\nu_{it})^{d_{it}} \left[1 - \Phi(\nu_{it})\right]^{d_{it}}, \tag{6}$$

with $\nu_{it}$ specified as

$$\begin{aligned} \nu_{it} &= d_{i,t-1}\boldsymbol{x}'_{i,t-1}\boldsymbol{\psi}_1 + (1 - d_{i,t-1})\boldsymbol{x}'_{i,t-1}\boldsymbol{\psi}_0 + \eta_i \text{ for } t > 0 \\ \nu_{i0} &= \vartheta\eta_i + \boldsymbol{x}'_{i0}\boldsymbol{\delta}, \end{aligned} \tag{7}$$

where $\eta_i$ denotes the firm-specific unobserved heterogeneity, again assumed to be normally distributed with zero mean and variance $\sigma^2_\eta$. The set of exogenous explanatory variables is left unchanged.

The formulation in (7) is that of a transition probability model, typical of the empirical literature on poverty dynamics (Jenkins, 2000; Cappellari and Jenkins, 2004). The specification is designed to study heterogeneous effects of subject-specific characteristics on the probability of entering/remaining in the state identified by the dependent variable at time $t$. In practice, it amounts to including interaction terms between the regressors and the lagged dependent variable so that, in our case, we can estimate different coeffi-

[14] Partial effects for an ordered model should be computed for the probability of each category. If interactions between the lags and covariates were also included, the effects of interest become $J^2$ for each sector.

cients for those firms who were and were not in the bottom 20% of the R&D intensity distribution in $t-1$. A standard dynamic binary choice model can readily be obtained by imposing $\boldsymbol{\psi}_0 = \boldsymbol{\psi}_1 = \boldsymbol{\psi}$.[15]

Two different Average Partial Effects (APE) for the $p$-th regressor can therefore be obtained as

$$\text{APE}_{pz} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Delta\Phi\left(\frac{x'_{i,t-1}\boldsymbol{\psi}_z}{\sqrt{1+\sigma_\eta^2}}\right),\ z = 0, 1, \tag{8}$$

where $\Delta(\cdot)$ here denotes partial continuous differentiation or discrete changes with respect to $x_p$ according to whether it is continuous or discrete. The above expression denotes the APEs of the $p$-th regressor on the entry ($z = 0$) and persistence ($z = 1$) probability in the bottom 20%, and their estimated counterparts are obtained by evaluating (8) at the parameters maximum likelihood estimates. Finally, we also specify a model for the top 20%, i.e. $d_{it} = \mathbf{I}(y_{it} = 4)$, for which the expressions (7) and (8) can readily by obtained by the appropriate change of notation.

# 5 Estimation results

## 5.1 Dynamic ordered probit model

The first set of results we consider are those from the dynamic ordered probit model. It is worth recalling that in this case the dependent variable consists in five categories corresponding to sector-specific quintiles of R&D intensity. The related estimated coefficients are reported in Table 3. Although coefficients cannot be interpreted in a straightforward manner, we observe some interesting features. With only one exception, state dependence parameters are always positive and statistically significant, denoting a strong persistence across states. Further, for all sectors, sales have statistically significant negative (positive) effect on the probability of being in the fifth (first) quintile, so that it is possible to conjecture that larger firms would tend to be placed in the "left-tail" of the R&D intensity distribution. Only in PHB and SCS there is evidence of a similar effect exerted by profitability: i.e. more profitable companies are likely to shift towards lower levels of R&D intensity. Instead, only in THE the intensity of physical capital has an opposite effect.

The dynamic ordered probit model gives us the possibility of computing predicted membership probabilities for each state, the related transition matrix and then the shape

[15] As the ordered probit, we rely on a random-effects transition probability model. Fixed-effects specifications cannot be considered in this case as companies for which the response variable is always equal to 0 cannot be used in the estimation sample. The large number of such cases leaves us with an estimation sample that is too small (15 to 30 firms) to yield reliable inference.

Table 3: Dynamic Ordered Probit model: estimated coefficients

| | EEE | PHB | SCS | THE |
|---|---|---|---|---|
| $\mathbf{I}(y_{i,t-1} = 2)$ | 0.848*** | 0.561** | 0.630 | 1.211*** |
| | (0.293) | (0.262) | (0.395) | (0.278) |
| $\mathbf{I}(y_{i,t-1} = 3)$ | 1.633*** | 1.265*** | 1.540*** | 2.333*** |
| | (0.374) | (0.338) | (0.459) | (0.541) |
| $\mathbf{I}(y_{i,t-1} = 4)$ | 2.403*** | 1.640*** | 1.991*** | 3.374*** |
| | (0.487) | (0.456) | (0.526) | (0.603) |
| $\mathbf{I}(y_{i,t-1} = 5)$ | 3.400*** | 2.472*** | 3.203*** | 4.530*** |
| | (0.512) | (0.588) | (0.558) | (0.660) |
| $\log(\text{Sales})_{t-1}$ | -0.967*** | -0.445*** | -0.531*** | -0.428*** |
| | (0.095) | (0.102) | (0.109) | (0.116) |
| $\text{Profitability}_{t-1}$ | 0.013 | -0.011*** | -0.003 | -0.016*** |
| | (0.009) | (0.004) | (0.003) | (0.006) |
| $\text{Capex\_int}_{t-1}$ | -0.005 | 0.009 | -0.016 | 0.035*** |
| | (0.013) | (0.017) | (0.014) | (0.009) |
| East Asia | 0.631** | -2.229*** | -0.230 | -1.606*** |
| | (0.322) | (0.323) | (0.148) | (0.248) |
| China | 0.158 | -3.472*** | -1.374*** | -1.056*** |
| | (0.227) | (0.753) | (0.387) | (0.297) |
| Europe | 1.527*** | -1.120*** | -0.611*** | 0.638 |
| | (0.362) | (0.273) | (0.215) | (0.575) |
| RoW | - | -4.227*** | -0.610*** | -0.727 |
| | - | (0.561) | (0.201) | (0.582) |
| $\sigma^2_\alpha$ | 2.053*** | 1.979*** | 2.387*** | 2.112*** |
| | (0.217) | (0.250) | (0.253) | (0.271) |
| $c_1$ | -7.873*** | -5.686*** | -5.047*** | -4.681*** |
| | (0.793) | (1.100) | (0.795) | (1.125) |
| $c_2$ | -5.500*** | -3.578*** | -2.674*** | -1.761 |
| | (0.728) | (1.093) | (0.781) | (1.101) |
| $c_3$ | -3.276*** | -1.785* | -0.711 | 0.746 |
| | (0.636) | (1.061) | (0.795) | (1.112) |
| $c_4$ | -0.793 | 0.222 | 1.508* | 3.165*** |
| | (0.636) | (1.065) | (0.783) | (1.106) |

**Notes:** $p$-values: $^* < 0.1$, $^{**} < 0.05$ and $^{***} < 0.01$. Model specifications also include an intercept term, time dummies, and $R\&D_{2013}$.

of their steady-state distribution. Figures 2 and 3 illustrate the estimated initial- and steady-state probability distributions among quintiles of R&D intensity. For each sector, the probability distributions are concerned with five groups of companies: from those starting with the bottom quintile (Group 1) to those starting with the top one (Group 5). In case of convergence of the level of R&D intensity we would expect to observe a shift of the estimated state probability masses from extreme to central quintiles.

For all sectors, the above figures show that this has not happened. Indeed, the companies show similar patterns of high persistence as witnessed by the small differences between the initial and steady state probability distributions in all the groups considered, with the partial exceptions of Group 5 in EEE and THE. In both sectors, in fact, companies in the highest quintiles (Group 5) record a steady state probability of being in the same quintile higher than that of the initial state, suggesting a process of entry (cf. Figures 2i and 3j). Although to a lower extent, the same occurs to Group 4 of THE (see Figure 3h).

To summarise, the orderd probit estimates indicate the presence of a strong persistence. There is no convergence to central quintiles: on the contrary, for two sectors, there is some evidence of a transition process towards "the right tail of the distribution" for firms characterised by an initial high level of R&D intensity.

## 5.2 Transition probability model

We now focus on the transition probability model outlined in Equations (6)-(8). In this case the dependent variables are binary and take value one when the R&D intensity level of a firm lies in the first (bottom 20%) or the fifth (top 20%) quintile, respectively.

Table 4 reports the results for the analysis concerning the probability of being in the first quintile. The first part of the table reports the APE for $y_{i,t-1}$ in a standard random-effects dynamic probit model.[16] This parameter is higher in PHB: on average and *ceteris paribus*, firms that were in the bottom 20% at time $t-1$ have a probability of being in the same quintile at time $t$ 9.6 percentage points higher than firms that were in the top 80%. Relatively lower values emerge for EEE, SCS and THE, with 8.6, 5.8 and 3.4 percentage points (pp henceforth), respectively.

The rest of Table 4 reports the estimated APEs for the transition probability models, related to the probabilities of remaining (top panel) and entering (bottom panel) in the lowest quintile according to firm characteristics (sales, profitability and capital intensity), and regional dummies. With the exception of PHB, both the probabilities of remaining and entering in the bottom 20% of the R&D intensity distribution increase with sales. Considering, for example, EEE, a 1% increase of sales raises the probability of persistence in the first quintile by 8.2 pp while raising that of entry by 7.5 pp.

Regarding the impacts of other company characteristics, additional findings emerge for the THE sector only: the probability of persisting in the first quintile decreases with the intensity of capital expenditures and the location in China (with a smaller effect in the first

[16] According to the notation of Section 4, this model is defined by Equations (6) and (7) under the restriction $\boldsymbol{\psi}_0 = \boldsymbol{\psi}_1 = \boldsymbol{\psi}$.

case). In the same sector and in SCS the probability of entry is also affected, with different impacts, by country/regional dummies. For instance, companies with headquarters in Europe, compared to their US counterparts, have a probability of entry in the bottom 20% 31.4 pp lower in THE while 4.3 pp higher is SCS.

The set of results for the top 20% of the R&D intensity distribution is reported in Table 5. As we can see, with the partial exception of SCS, the state dependence is strong in all sectors and, most importantly, stronger than that found for the lowest quintile. With the above exception, firms that were in top 20% at time $t-1$ have a probability of being in the highest quintile at time $t$ about 18-23 pp larger than those that were in the bottom 80%.

With respect to the transition probability model, in EEE and THE, sales exert a statistically significant negative effect on the probability of persisting in the fifth quintile of the R&D intensity distribution (see the top panel of Table 5). When sales rise by 1% the probability of remaining in the top 20% decreases, on average, by 8.3 pp in EEE and by 5.3 pp in THE. In the latter sector a negative impact also emerges for companies with headquarters in Europe (minus 11 pp) and for those more profitable (though, in this case, the magnitude of the effect is very small).

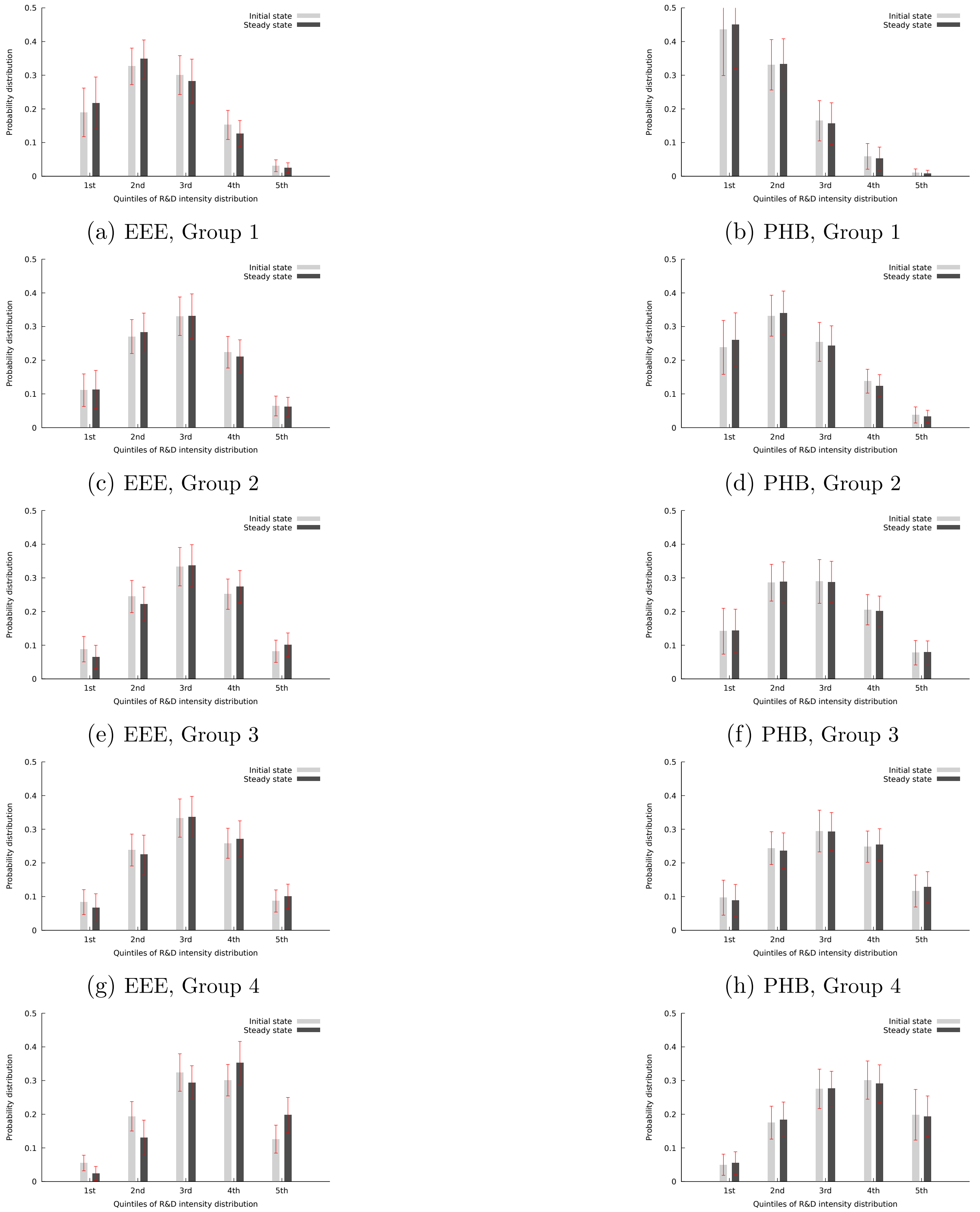
The bottom panel of Table 5 reports instead the estimated effects on the probability of entering in the top 20%. In this case, with the exception of PHB, it emerges that sales have a statistically significant negative effect on the probability of entry: in particular, a 1% increase of sales reduces such a probability by 12 pp in EEE and by 9 pp in THE. Only in the latter sector there is evidence that more profitable companies have a lower probability of entry while only in SCS the intensity of capital expenditures exerts an opposite effect (in both cases, however, the magnitude of the impact is relatively small). Finally, in SCS the probability of entry increases by about 4 pp for the companies with headquarters in China.

Table 4: Bottom 20%. Transition probability model: Average partial effects

| | EEE | PHB | SCS | THE |
|---|---|---|---|---|
| | Dynamic probit model | | | |
| $y_{i,t-1}$ | 0.086** | 0.096*** | 0.058** | 0.034* |
| | (0.036) | (0.035) | (0.028) | (0.018) |
| | Transition probability model: $y_{i,t-1} = 1$ | | | |
| log(Sales)$_{t-1}$ | 0.082** | 0.013 | 0.062*** | 0.035*** |
| | (0.034) | (0.050) | (0.023) | (0.005) |
| Profitability$_{t-1}$ | 0.005 | -0.009 | -0.001 | -0.000 |
| | (0.003) | (0.008) | (0.001) | (0.001) |
| Capex_int$_{t-1}$ | 0.002 | -0.008* | 0.016 | -0.001** |
| | (0.005) | (0.004) | (0.010) | (0.000) |
| East Asia | -0.011 | -0.038 | 0.044 | -0.005 |
| | (0.089) | (0.095) | (0.034) | (0.007) |
| China | -0.078 | -0.064 | - | -0.047*** |
| | (0.068) | (0.095) | - | (0.018) |
| Europe | 0.043 | - | 0.036 | 0.022 |
| | (0.037) | - | (0.021) | (0.044) |
| | Transition probability model: $y_{i,t-1} = 0$ | | | |
| log(Sales)$_{t-1}$ | 0.075*** | 0.069* | 0.107*** | 0.104*** |
| | (0.018) | (0.040) | (0.040) | (0.015) |
| Profitability$_{t-1}$ | 0.001 | -0.001 | 0.000 | 0.000 |
| | (0.002) | (0.001) | (0.000) | (0.001) |
| Capex_int$_{t-1}$ | -0.004 | 0.001 | 0.002 | 0.000 |
| | (0.004) | (0.004) | (0.004) | (0.001) |
| East Asia | 0.005 | 0.048 | 0.083*** | 0.035*** |
| | (0.047) | (0.039) | (0.030) | (0.012) |
| China | -0.033 | 0.072 | - | -0.128*** |
| | (0.115) | (0.053) | - | (0.043) |
| Europe | -0.045 | - | 0.043** | -0.314*** |
| | (0.093) | - | (0.022) | (0.048) |

**Notes:** $p$-values: $^{*} < 0.1$, $^{**} < 0.05$ and $^{***} < 0.01$. Model specifications also include an intercept term, time dummies (not interacted with $y_{i,t-1}$) and $R\&D_{2013}$.

Figure 2: Initial and steady state probability distributions by firm group: EEE and PHB




(a) EEE, Group 1

(b) PHB, Group 1

(c) EEE, Group 2

(d) PHB, Group 2

(e) EEE, Group 3

(f) PHB, Group 3

(g) EEE, Group 4

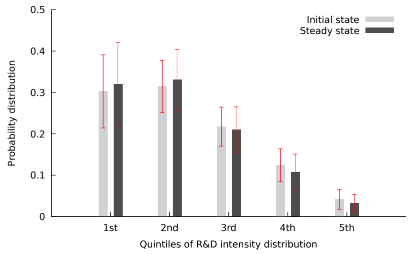
(h) PHB, Group 4

(i) EEE, Group 5

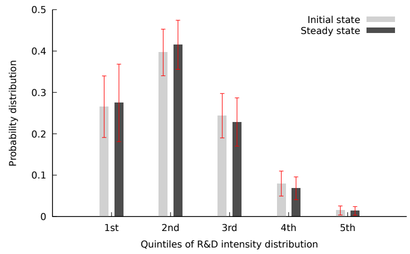
(j) PHB, Group 5

**Notes:** The red vertical bar represents the 95% confidence interval.

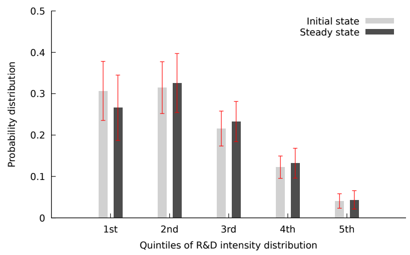
Figure 3: Initial and steady state probability distributions by firm group: SCS and THE



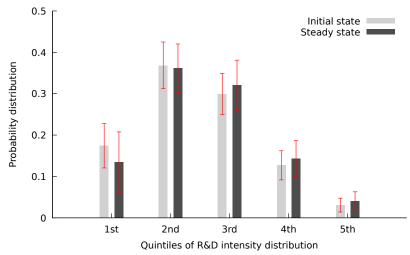

(a) SCS, Group 1



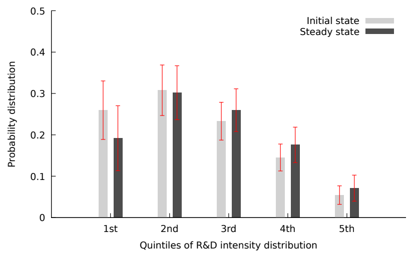

(b) THE, Group 1



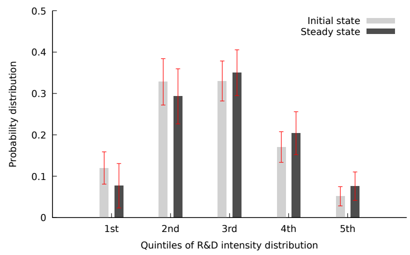

(c) SCS, Group 2



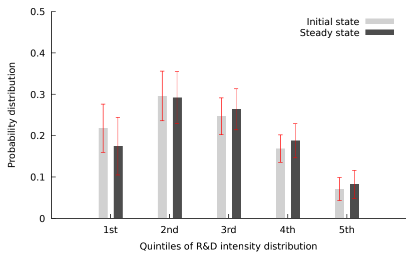

(d) THE, Group 2



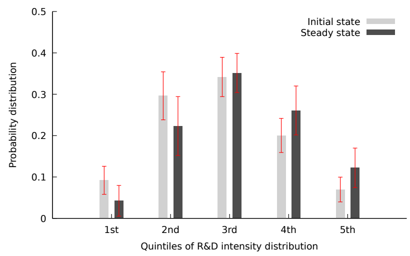

(e) SCS, Group 3



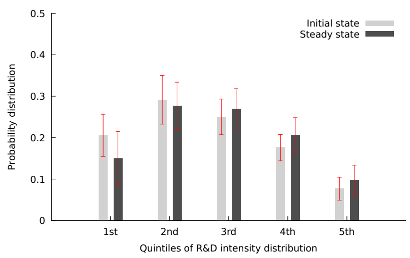

(f) THE, Group 3



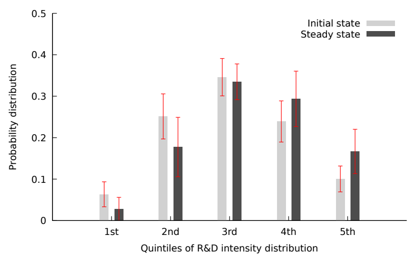

(g) SCS, Group 4




(h) THE, Group 4




(i) SCS, Group 5




(j) THE, Group 5

**Notes:** The red vertical bar represents the 95% confidence interval.

Table 5: Top 20%. Transition probability model: Average partial effects

| | EEE | PHB | SCS | THE |
|---|---|---|---|---|
| | Dynamic probit model | | | |
| $y_{i,t-1}$ | 0.233** | 0.213** | 0.072** | 0.176*** |
| | (0.112) | (0.090) | (0.029) | (0.064) |
| | Transition probability model: $y_{i,t-1} = 1$ | | | |
| log(Sales)$_{t-1}$ | -0.083*** | -0.026 | -0.024 | -0.053*** |
| | (0.028) | (0.022) | (0.036) | (0.014) |
| Profitability$_{t-1}$ | -0.001 | -0.001 | -0.003** | -0.002*** |
| | (0.002) | (0.001) | (0.001) | (0.001) |
| Capex_int$_{t-1}$ | -0.003 | -0.001 | 0.000 | 0.007 |
| | (0.002) | (0.002) | (0.003) | (0.005) |
| East Asia | 0.005 | -0.049 | -0.129 | -0.083* |
| | (0.037) | (0.069) | (0.106) | (0.048) |
| China | -0.136* | - | -0.197 | -0.096 |
| | (0.076) | - | (0.122) | (0.071) |
| Europe | -0.034 | -0.039 | -0.058 | -0.109** |
| | (0.045) | (0.060) | (0.084) | (0.051) |
| | Transition probability model: $y_{i,t-1} = 0$ | | | |
| log(Sales)$_{t-1}$ | -0.118*** | -0.017 | -0.038** | -0.089*** |
| | (0.032) | (0.018) | (0.016) | (0.019) |
| Profitability$_{t-1}$ | -0.000 | -0.001 | -0.000 | -0.001** |
| | (0.002) | (0.001) | (0.001) | (0.001) |
| Capex_int$_{t-1}$ | -0.001 | -0.002 | 0.004** | -0.001 |
| | (0.003) | (0.005) | (0.002) | (0.002) |
| East Asia | -0.038 | -0.017 | 0.022 | -0.083 |
| | (0.050) | (0.068) | (0.024) | (0.058) |
| China | -0.063 | - | 0.038** | -0.097 |
| | (0.061) | - | (0.017) | (0.074) |
| Europe | -0.051 | -0.016 | 0.027 | -0.100 |
| | (0.060) | (0.058) | (0.019) | (0.064) |

**Notes:** $p$-values: $^{*} < 0.1$, $^{**} < 0.05$ and $^{***} < 0.01$. Model specifications also include an intercept term, time dummies (not interacted with $y_{i,t-1}$) and $R\&D_{2013}$.

## 5.3 State dependence in the right tail of the R&D intensity distribution

We now extend our analysis by considering alternative dichotomisations of the R&D intensity distribution. Specifically, we are interested in investigating whether the state dependence we found for the observations in the last quintile (see Table 5) is homogeneous for all the Top 20% of firms in terms of R&D intensity. In order to shed light on this aspect we perform the estimation of the dynamic probit model for two different dependent variables obtained by dichotomisation at specific sector-year quantiles of the R&D intensity distributions, namely the 90th (Top 10%) and the 85th (Top 15%). Average partial effects for the state dependence parameter is reported in Table 6.

Results highlight a clear pattern. The state dependence is not significant or even becomes negative (EEE) in the rightmost part of the R&D intensity distribution, i.e., Top 10%, while for SCS the maximum likelihood estimate does not converge. These findings are not surprising given the small number of firms that are placed in the top decile of the distribution each year (from 10 in PHB to 16 in THE; see Table 1). Indeed, state dependence turns out to be positive and significant as we consider a larger portion of firms, namely those characterized by "smaller", even though large, R&D intensities, i.e. Top 15%. The most significant results are obtained by selecting, as done in our previous estimates, the highest quintile of the R&D intensity distribution.

## 5.4 Extending the analysis up to 2020

In this further extension of our analysis we add the year 2020 to our dataset, using the 2021 wave of the *Scoreboard*. In the previous analysis we avoided doing so because of the negative economic shock caused by the Covid-19 pandemic at the global level. As stressed in the *Scoreboard* report (Grassano et al., 2021), in spite of the pandemic, the total R&D expenses of the world's top R&D investors belonging to ICT producers, ICT services and Health industries (i.e. to the sectors considered in our analysis) continued to increase in 2020. Instead other economic indicators, such as sales, profits and capital expenditures, exhibited a decline. Obviously, these are global sectoral changes which may hide remarkable differences between companies. Since our main variable of interest is the ratio of R&D expenditures on company sales we have no clue as to what changes may have occurred in the pandemic year compared to the previous tranquil periods. Therefore, in this section, we examine whether these changes have affected the persistence in low or high levels of R&D intensity.

In principle, extending the period of analysis should improve estimation accuracy. However, for the reasons stressed in Section 3, we have chosen to work with a balanced

Table 6: State dependence for different quantiles

| | EEE | PHB | SCS | THE |
|---|---|---|---|---|
| | Top 10% | | | |
| $y_{i,t-1}$ | -0.028** | 0.023 | - | 0.065 |
| | (0.012) | (0.023) | - | (0.103) |
| | Top 15% | | | |
| $y_{i,t-1}$ | 0.394* | 0.038* | 0.071* | 0.205** |
| | (0.230) | (0.020) | (0.041) | (0.084) |
| | Top 20% | | | |
| $y_{i,t-1}$ | 0.233** | 0.213** | 0.072** | 0.176*** |
| | (0.112) | (0.090) | (0.029) | (0.064) |

**Notes:** $p$-values: $^{*} < 0.1$, $^{**} < 0.05$ and $^{***} < 0.01$. For SCS the APE for the dependent variable limited at the top decile is not reported due to the lack of convergence.

panel of companies so that adding one more year entails losing those who exited from the ranking of the 2500 world's top R&D investors in 2020.[17] In total, these companies are 43 and Table 7 shows their distribution between the four sectors considered in our study.

Table 8 shows the transition probability model estimate, based on the extended panel, for the bottom part of the R&D intensity distribution. Almost all the results are consistent with those obtained with the 2013-2019 panel (see Table 4). The differences to highlight are: in the dynamic probit model for the PHB, companies show a much higher probability of persisting in the first quintile (plus 24.2 versus 9.6 pp); in EEE both profitability and the European location increase the same probability, while in SCS the same effects are exerted by the capital intensity and the location in East Asia (the former variable also enhances the likelihood of entering in the bottom 20%).

Consistent results also emerge for the top 20% of the R&D intensity distribution (see Table 9 and, for comparison, Table 5). In this case, SCS companies record a much higher probability of persistence in the top quintile (37.8 versus 7.2 pp). Both in SCS and EEE Chinese companies are characterised by a lower likelihood to remain in the same quintile, while in EEE negative effects are also due to the capital intensity and the European

[17] As already mentioned in Section 3, most of the exited companies were placed in the fringe of the ranking (i.e. performing low level of R&D expenses) while those in the top part were acquired by of merged with other companies (often included in the ranking).

Table 7: Number of companies by balanced panels

| | EEE | PHB | SCS | THE |
|---|---|---|---|---|
| 2013-2020 | 119 | 92 | 108 | 151 |
| 2013-2019 | 131 | 104 | 116 | 162 |
| Difference | -12 | -12 | -8 | -11 |

location.

It should be stressed that the stronger state dependence parameters recorded by PHB firms in the bottom and by SCS firms in the top 20% could be ascribed to a weak identification problem (as witnessed by an almost threefold increase in the estimated standard errors of the same parameters), probably due to a lack of variation in the binary dependent variable. Indeed, dropping from the sample even a few firms that exhibit transitions in and out from the bottom and top quintiles may result into a non-negligible loss of identifying information.

Table 8: 2013-2020 Bottom 20%. Transition probability model: Average partial effects

| | EEE | PHB | SCS | THE |
|---|---|---|---|---|
| | Dynamic probit model | | | |
| $y_{i,t-1}$ | 0.051** | 0.242** | 0.066* | 0.029 |
| | (0.023) | (0.109) | (0.039) | (0.018) |
| | Transition probability model: $y_{i,t-1} = 1$ | | | |
| log(Sales)$_{t-1}$ | 0.057*** | 0.030 | 0.053*** | 0.041*** |
| | (0.011) | (0.033) | (0.012) | (0.003) |
| Profitability$_{t-1}$ | 0.002** | -0.001 | 0.000 | 0.000 |
| | (0.001) | (0.001) | (0.000) | (0.001) |
| Capex_int$_{t-1}$ | 0.004** | -0.001 | 0.009** | -0.001 |
| | (0.002) | (0.004) | (0.005) | (0.001) |
| East Asia | -0.017 | -0.128 | 0.044*** | 0.008 |
| | (0.028) | (0.099) | (0.011) | (0.006) |
| China | -0.038 | -0.075 | - | -0.051*** |
| | (0.027) | (0.087) | - | (0.018) |
| Europe | 0.050*** | - | -0.008 | 0.052*** |
| | (0.016) | - | (0.014) | (0.014) |
| | Transition probability model: $y_{i,t-1} = 0$ | | | |
| log(Sales)$_{t-1}$ | 0.092*** | 0.047* | 0.084*** | 0.108*** |
| | (0.023) | (0.027) | (0.013) | (0.009) |
| Profitability$_{t-1}$ | 0.002* | -0.000 | 0.000 | -0.000 |
| | (0.001) | (0.001) | (0.000) | (0.000) |
| Capex_int$_{t-1}$ | -0.000 | -0.001 | 0.004** | -0.000 |
| | (0.003) | (0.003) | (0.002) | (0.000) |
| East Asia | 0.023 | 0.049 | 0.123*** | 0.009 |
| | (0.036) | (0.035) | (0.016) | (0.014) |
| China | -0.029 | 0.034 | - | -0.200*** |
| | (0.062) | (0.063) | - | (0.030) |
| Europe | 0.063 | - | 0.059*** | -0.102*** |
| | (0.048) | - | (0.013) | (0.013) |

**Notes:** $p$-values: * $< 0.1$, ** $< 0.05$ and *** $< 0.01$. Model specifications also include an intercept term, time dummies (not interacted with $y_{i,t-1}$) and $R\&D_{2013}$.

Table 9: 2013-2020 Top 20%. Transition probability model: Average partial effects

| | EEE | PHB | SCS | THE |
|---|---|---|---|---|
| | Dynamic probit model | | | |
| $y_{i,t-1}$ | 0.218*** | 0.250** | 0.378*** | 0.169*** |
| | (0.075) | (0.110) | (0.109) | (0.054) |
| | Transition probability model: $y_{i,t-1} = 1$ | | | |
| log(Sales)$_{t-1}$ | -0.084*** | -0.008 | -0.025 | -0.048*** |
| | (0.011) | (0.018) | (0.027) | (0.011) |
| Profitability$_{t-1}$ | -0.001 | -0.001 | -0.004*** | -0.002*** |
| | (0.002) | (0.001) | (0.001) | (0.001) |
| Capex_int$_{t-1}$ | -0.002** | -0.001 | 0.002 | -0.001 |
| | (0.001) | (0.002) | (0.002) | (0.001) |
| East Asia | -0.020 | 0.015 | -0.032 | -0.046 |
| | (0.038) | (0.036) | (0.063) | (0.036) |
| China | -0.111** | - | -0.256*** | - |
| | (0.051) | - | (0.092) | - |
| Europe | -0.103** | -0.012 | 0.039 | -0.066 |
| | (0.045) | (0.049) | (0.040) | (0.043) |
| | Transition probability model: $y_{i,t-1} = 0$ | | | |
| log(Sales)$_{t-1}$ | -0.083*** | -0.018 | -0.024** | -0.079*** |
| | (0.030) | (0.020) | (0.010) | (0.017) |
| Profitability$_{t-1}$ | 0.001 | 0.001 | -0.000 | -0.002** |
| | (0.002) | (0.001) | (0.001) | (0.001) |
| Capex_int$_{t-1}$ | 0.001 | 0.002 | 0.003** | -0.001 |
| | (0.002) | (0.003) | (0.001) | (0.002) |
| East Asia | -0.033 | -0.071 | 0.021 | -0.072 |
| | (0.067) | (0.075) | (0.019) | (0.046) |
| China | -0.108 | - | 0.037*** | - |
| | (0.089) | - | (0.009) | - |
| Europe | -0.034 | -0.041 | 0.021 | -0.077 |
| | (0.067) | (0.060) | (0.014) | (0.062) |

**Notes:** $p$-values: $^{*} < 0.1$, $^{**} < 0.05$ and $^{***} < 0.01$. Model specifications also include an intercept term, time dummies (not interacted with $y_{i,t-1}$) and $R\&D_{2013}$.

# 6 Discussion and concluding remarks

The first and most important finding of our analysis is that time persistence (state dependence) appears to be stronger in the highest as opposed to the lowest quintile of the R&D intensity distribution. In other words, companies with high R&D intensities are less likely to move towards medium levels than companies recording low R&D intensities. Hence, the lack of convergence towards a sectoral mean is mainly due to a set of companies that persistently invest in R&D more resources with respect to their sales. This is consistent with the results obtained, through different methods, by Coad (2019) (see the last part of Section 2). It should be stressed that our finding emerges from a balanced panel of selected corporations (top R&D investors in the world) competing in the global market for which, at least in principle, it should be more likely to observe a process of convergence. Why, instead, several of these companies belonging to high-tech sectors keep investing heavily in R&D with respect to their sales?

Different reasons can be invoked in this respect (see Section 2). We can refer to the hypothesis of "success breed success" coupled with that of increasing returns in R&D: if past R&D efforts makes current efforts more productive companies will tend to maintain high R&D/sales ratios (Klette and Johansen, 2000). This can also be due to the presence of R&D sunk-costs, which could be substantial as far as companies undertake ambitious research projects with expected returns only in the medium- or long-term. The latter explanation appears to be particularly suitable for the Pharmaceutical & Biotechnology sector in which incremental innovations are not appreciated by both private and public payers so that R&D investment tend to focus on new though more uncertain and difficult outcomes (Pammolli et al., 2011); moreover, due to the need of extensive clinical trials before being approved by regulators, to translate new discoveries into marketable products is a quite risky and long process (OECD, 2018; Hernández et al., 2019).[18]

The firms' characteristics associated with the above explanatory hypotheses are many and most of them are not easily observable (e.g. managerial practices, risk attitude, product diversification, breadth of the knowledge bases). However, although with a limited set of variables, our analysis provides interesting findings which are also useful to highlight some differences among sectors.

As expected, also in the light of how the intensity of R&D is measured, the company size, proxies by sales, plays a relevant role. Larger companies are less likely to carry on higher R&D intensities but this effect is statistically significant only for those belonging to Electronic & Electrical Equipment and Technology Hardware & Equipment. Instead, higher sales increase the probability of keeping a low intensity of R&D: however, this

[18] According to OECD (2018) the average probability of getting marketing approval for a new drug is estimated at 14% while its clinal development takes around eight years

occurs in the same sectors and in Software & Computer Services but not in Pharmaceutical & Biotechnology. The fact that only in the latter sector companies' sales, with a one-year lag, do not influence the persistence in both low and high levels of R&D intensity could be due to the already mentioned characteristic: in the biopharmaceutical sector the current revenues depend on the R&D carried out in the distant past while current R&D investment will generate revenues in the distant future (OECD, 2018). Among other possible reasons, we can refer to the processes of mergers and acquisitions (M&A) which have been particularly intense in this sector.[19] By acquiring smaller inventive firms characterised by high R&D efforts, larger pharma companies can keep a stable intensity of R&D expenditures even in presence of sales' increases.

Regarding other explanatory variables, there are few significant effects. For instance, only in Technology Hardware & Equipment companies with higher shares of capital investments are less likely to carry on low R&D intensities (suggesting complementarity between the two types of investment). In Software & Computer Services only, the share of profits decreases the likelihood of keeping high intensities of R&D (suggesting a negative relationship).

With respect to the transition processes, in all sectors but Biopharmaceuticals the probability of moving towards lower (higher) R&D intensities is positively (negatively) affected by the lagged level of sales. Again, aside from the size variable, results across sectors are not homogeneous. For example, Software & Computer Services companies with a greater intensity of capital expenditures are more likely to move towards high intensities of R&D (pointing, again, to complementarity) while in Technology Hardware & Equipment the level of profitability reduces the likelihood of entering in the group of high R&D intensive companies. Hence, also in this case, there seems to be a negative relationship between the share of profits and that of R&D expenses.

Finally, with regard to the role of geographical location, companies with headquarters in East Asia, China and Europe, as opposed to their US counterparts, tend to move away from low R&D intensities in Technology Hardware & Equipment, while in Software & Computer Services only Chinese companies tend to shift towards high intensities of R&D expenditures. The latter result somewhat reflects the growing challenge of Chinese companies to the US leadership in ICTs.

Moving to the empirical limitations and drawbacks of our study, a seven-year period (2013-2019) could be viewed as too short for analysing R&D intensity persistence. We

[19] The reason for this is twofold (Hernández et al., 2019). On the one hand, the early stages of research giving rise to patented inventions are increasingly carried out by smaller biopharma firms (IQVIA (2019, 2020)). However, on the other hand, to develop biologic drugs up to the commercial stage more R&D in clinical trials is needed and more regulatory requirements must be fulfilled. Larger pharma companies are better equipped to perform these demanding and expensive tasks, obviously after acquiring or merging with smaller inventive firms.

show that by adding the year 2020 (albeit affected by the Covid-19 pandemic) the results do not significantly change. If, as we argue, the choice of using a balanced panel is needed, extending the time dimension too far is not advisable because it would remarkably reduce the number of companies considered. Instead, the set of company-level variables used to control for observable characteristics should be extended. For instance, as in Cincera and Veugelers (2013), the companies' age could be included for inspecting whether the (relatively) younger ones are characterised by persistently higher R&D intensities. Having also the advantage of being time-variant, another variable worth to be considered is the size and type of acquisitions made by the companies included in the balanced panel. The employment of further variables connected to corporate strategies and product diversification processes would be desirable but not at the price of reducing the number of companies analysed.

# References

Antonelli C. 1989. A failure-inducement model of research and development expenditure: Italian evidence from the early 1980s. *Journal of Economic Behavior & Organization* **12**: 159–180.

Antonelli C, Crespi F, Scellato G. 2012. Inside innovation persistence: New evidence from Italian micro-data. *Structural Change and Economic Dynamics* **23**: 341–353.

Arqué-Castells P. 2013. Persistence in r&d performance and its implications for the granting of subsidies. *Review of Industrial Organization* **43**: 193–220.

Arroyabe MF, Schumann M. 2022. On the estimation of true state dependence in the persistence of innovation. *Oxford Bulletin of Economics and Statistics* .

Barge-Gil A, López A. 2014. R&D determinants: Accounting for the differences between research and development. *Research Policy* **43**: 1634–1648.

Baysinger B, Hoskisson RE. 1989. Diversification strategy and R&D intensity in multi-product firms. *Academy of Management journal* **32**: 310–332.

Cappellari L, Jenkins SP. 2004. Modelling low income transitions. *Journal of Applied Econometrics* **19**: 593–610.

Cefis E. 2003. Is there persistence in innovative activities? *International Journal of Industrial Organization* **21**: 489–515.

Christen P. 2006. A comparison of personal name matching: Techniques and practical issues. In *Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06)*. IEEE, 290–294.

Cincera M, Ravet J. 2014. Globalisation, industrial diversification and productivity growth in large european R&D companies. *Journal of Productivity Analysis* **41**: 227–246.

Cincera M, Veugelers R. 2013. Young leading innovators and the eu's r&d intensity gap. *Economics of Innovation and New Technology* **22**: 177–198.

Clausen T, Pohjola M, Sapprasert K, Verspagen B. 2012. Innovation strategies as a source of persistent innovation. *Industrial and Corporate Change* **21**: 553–585.

Coad A. 2019. Persistent heterogeneity of R&D intensities within sectors: Evidence and policy implications. *Research Policy* **48**: 37–50.

Cohen WM, Klepper S. 1992. The anatomy of industry R&D intensity distributions. *The American Economic Review* : 773–799.

Cohen WM, Klepper S. 1996. A reprise of size and r & d. *The Economic Journal* **106**: 925–951.

Cohen WM, Levin RC, Mowery DC. 1987. Firm size and R&D intensity: A re-examination.

Galaasen SM, Irarrazabal A. 2021. R&D heterogeneity and the impact of R&D subsidies. *The Economic Journal* **131**: 3338–3364.

Geroski PA, Van Reenen J, Walters CF. 1997. How persistently do firms innovate? *Research policy* **26**: 33–48.

Grassano N, Hernandez Guevara H, Fako P, Tuebke A, Amoroso S, Georgakaki A, Napolitano L, Pasimeni F, Rentocchini F, Compaño R, Fatica S, Panzica R. 2021. The 2021 EU Industrial R&D Investment Scoreboard. EUR 30902 EN, Publications Office of the European Union.

Grassano N, Hernandez Guevara H, Tuebke A, Amoroso S, Dosso M, Georgakaki A, Pasimeni F. 2020. The 2020 EU Industrial R&D Investment Scoreboard. EUR 30519 EN, Publications Office of the European Union.

Heckman JJ. 1981a. Heterogeneity and state dependence. In Rosen S (ed.) *Studies in Labor Markets*. Chicago IL: Chicago University Press, 91–139.

Heckman JJ. 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In Manski CF, McFadden D (eds.) *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge MA: MIT Press, 179–195.

Hernández H, Grassano N, Tübke A, Amoroso S, Csefalvay Z, Gkotsis P. 2019. The 2019 EU Industrial R&D Investment Scoreboard. EUR 30002 EN, Publications Office of the European Union.

Honoré F, Munari F, de La Potterie BvP. 2015. Corporate governance practices and companies' r&d intensity: Evidence from european countries. *Research policy* **44**: 533–543.

IQVIA. 2019, 2020. Emerging biopharma's contribution to innovation. IQVIA, Institute for Human Data Science.

Jenkins SP. 2000. Modelling household income dynamics. *Journal of Population Economics* **13**: 529–567.

Klette TJ, Johansen F. 2000. Accumulation of R&D capital and dynamic firm performance: a not-so-fixed effect model. In *The economics and econometrics of innovation*. Springer, 367–397.

Le Bas C, Scellato G. 2014. Firm innovation persistence: a fresh look at the frameworks of analysis. *Economics of Innovation and New Technology* **23**: 423–446.

Levin RC, Cohen WM, Mowery DC. 1985. R & d appropriability, opportunity, and market structure: new evidence on some schumpeterian hypotheses. *The American economic review* **75**: 20–24.

Manez JA, Rochina-Barrachina ME, Sanchis A, Sanchis JA. 2009. The role of sunk costs in the decision to invest in R&D. *The Journal of Industrial Economics* **57**: 712–735.

Moncada-Paternò-Castello P. 2022. Top R&D investors, structural change and the R&D growth performance of young and old firms. *Eurasian Business Review* **12**: 1–33.

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**: 381–393.

Nelson RR. 1988. Modelling the connections in the cross section between technical progress and R&D intensity. *The Rand Journal of Economics* : 478–485.

OECD. 2018. Pharmaceutical innovation and access to medicines. OECD Health Policy Studies, OECD Publishing.

Ortega-Argilés R, Piva M, Vivarelli M. 2015. The productivity impact of R&D investment: are high-tech sectors still ahead? *Economics of Innovation and New Technology* **24**: 204–222.

Pammolli F, Magazzini L, Riccaboni M. 2011. The productivity crisis in pharmaceutical r&d. *Nature reviews Drug discovery* **10**: 428–438.

Peters B. 2009. Persistence of innovation: stylised facts and panel data evidence. *The Journal of Technology Transfer* **34**: 226–243.

Raffo J. 2015. MATCHIT: Stata module to match two datasets based on similar text patterns. Statistical Software Components, Boston College Department of Economics. URL https://ideas.repec.org/c/boc/bocode/s457992.html

Raymond W, Mohnen P, Palm F, Van Der Loeff SS. 2010. Persistence of innovation in dutch manufacturing: Is it spurious? *The Review of Economics and Statistics* **92**: 495–504.

Sterlacchini A. 1994. Technological opportunities, intraindustry spillovers and firm R&D intensity: Some evidence for italian manufacturing industries. *Economics of Innovation and New Technology* **3**: 123–138.

Tavassoli S, Karlsson C. 2015. Persistence of various types of innovation analyzed and explained. *Research Policy* **44**: 1887–1901.

Triguero A, Córcoles D. 2013. Understanding innovation: An analysis of persistence for spanish manufacturing firms. *Research Policy* **42**: 340–352.

Woerter M. 2014. Competition and persistence of R&D. *Economics of Innovation and New Technology* **23**: 469–489.

Wooldridge JM. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* **20**: 39–54.

Zobel J, Dart P. 1995. Finding approximate matches in large lexicons. *Software: Practice and Experience* **25**: 331–345.

# A List of countries per Geographical Area

Table 10: Countries per Geographical Area

| East Asia | US | China | Europe | RoW |
|---|---|---|---|---|
| Hong Kong | USA | China | Austria | Argentina |
| Japan | | | Belgium | Australia |
| Singapore | | | Denmark | Bermuda |
| South Korea | | | Finland | Brazil |
| Taiwan | | | France | Canada |
| | | | Germany | Cayman Islands |
| | | | Hungary | India |
| | | | Ireland | |
| | | | Israel | |
| | | | Italy | |
| | | | Netherlands | |
| | | | Norway | |
| | | | Portugal | |
| | | | Slovenia | |
| | | | Spain | |
| | | | Sweden | |
| | | | Switzerland | |
| | | | Turkey | |
| | | | UK | |