LEM
WORKING PAPER SERIES

What do firms know? What do they produce? A new look at the relationship between patenting profiles and patterns of product diversification

Giovanni Dosi °
Marco Grazzi §
Daniele Moschella °*

° Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy
§ Department of Economics, University of Bologna, Italy
* IBIMET-CNR

What do firms know? What do they produce? A new look at the relationship between patenting profiles and patterns of product diversification[*]

Giovanni Dosi[†1], Marco Grazzi[‡2], and Daniele Moschella[§1,3]

[1]Institute of Economics, Scuola Superiore Sant'Anna
[2]Department of Economics, University of Bologna
[3]IBIMET-CNR

March 30, 2015

Abstract

In this work we analyze the relationship between the patterns of firm diversification, if any, across product lines and across bodies of innovative knowledge, proxied by the patent classes where the firm is present. Putting it more emphatically we investigate the relationship between "what a firm does" and "what a firm knows". Using a newly developed dataset matching information on patents and products at the firm level, we provide evidence concerning firms' technological and product scope, their relationships, the size-scaling and coherence properties of diversification itself. Our analysis shows that typically firms are much more diversified in terms of products than in terms of technologies, with their main products more related to the exploitation of their innovative knowledge. The scaling properties show that the number of products and technologies increase log-linearly with firm size. And the directions of diversification themselves display coherence between neighboring activities also at relatively high degrees of diversification. These findings are well in tune with a capability-based theory of the firm.

JEL codes: C81, D22, L20, L25, O31

Keywords: Diversification, Coherence, Patents, Products, Capabilities, Firm behaviour

1   Introduction

In this work we analyze the relationship between the patterns of firm diversification, if any, across product lines and across bodies of innovative knowledge, proxied by the patent classes where the firm is present. Putting it more emphatically we investigate the relationship between "what a firm does" and "what a firm knows".

Firms, and large companies in particular, tend to be active in more than one product. In this respect, on the one side the process of firm growth is related to the choice of the diversification pattern of the existing product portfolio and to the entry into new sub-markets (Bottazzi et al., 2001). On the other side, the successful accomplishment of the introduction of a new product crucially depends upon, and is constrained by, the existing stock of knowledge embedded in the organization (see among the others Nelson and Winter, 1982; Teece et al., 1994).

The technological and product breadth of a firm co-evolve over time, with technological diversification generally anticipating the introduction of new products (Pavitt, 1998). Knowledge accumulation and innovation, the evidence shows, are indeed positively related to productivity and contribute to boost sales, both on domestic and international markets (for empirical evidence see, among others, Hall and Mairesse, 1995; Wakelin, 1998; Dosi et al., 2014). Product innovation is one of the main forms of such innovative efforts: introducing a new good is likely to benefit the innovating firm in terms of market shares and penetration in new markets. Moreover, by producing new goods, the firm can diversify its product portfolio and exploit economies of scope of various kind (Hirsch and Lev, 1971; Montgomery, 1994). Also the process of continuous feedback between internal research activities and the actual production of *artefacts* contributes to the accumulation of capabilities and skills (see, among the others, Dosi, 1988; Pavitt, 1998).

In all that a crucial question concerns the precise relation between the expansion of the scope of technological knowledge on the one hand and of product portfolios on the other. Relatedly, a central issue regards the *pattern* and *direction* of diversification: how do firms choose to enter new technological fields or add products, conditional upon the existing stock of knowledge and product portfolio? Teece et al. (1994) investigates the diversification of US corporates larger than 20 employees into different industrial sectors and find that firms tend to diversify into *related* or neighboring industries. Patel and Pavitt (1997), who focus instead on the technological knowledge, report an analogous finding of path dependency in the accumulation of firm-specific technological competencies. Several studies have further contributed to support the relatedness hypothesis (Breschi et al., 2003; Valvano and Vannoni, 2003; Nesta and Saviotti, 2006; Bottazzi and Pirino, 2010). And more recent work of Piscitello (2000, 2004); Miller (2006) started to investigate the relation between the diversification pattern followed by firms and corporate performance, finding a positive relation between corporate coherence and firm performance.

In terms of *scope* of diversification Patel and Pavitt (1997), looking at large diversified firms, find that "firms know more than they do": the scope of knowledge diversification is wider than product diversification. Brusoni et al. (2001) confirm the finding in the case of the highly specialized industry of aircraft engine control systems and show that in a highly interdependent environment, in which a change in one firm within the network affects all the others, corporates *"need to have knowledge in excess of what they need for what they make, to cope with imbalances caused by uneven rates of development in the technologies on which they rely and with unpredictable product-level interdependencies."*

This work provides several contributions to this whole stream of literature. *First,*

we analyze the characteristics of the universe of Italian exporting firms and those which display innovative technological knowledge, as proxied by patents[1]. Patenting firms turn out to be bigger than non patenting firms, both in terms of total exports and in terms of product scope.

Second, we investigate the properties of the diversification breadth of both technological knowledge and product portfolios, using detailed product information from custom data. In this respect, employing the algorithm recently developed by (Lybbert and Zolas, 2014), we match the technological field in which a firm holds a patent to the corresponding 4-digit industry in which the firm is active with at least one product. This matching, although with some caveats, enables us to shed light on relevance of specific technological knowledge in backing the product in international markets and also on the Pavittian question of "knowing" *vs.* "making". In our Italian case we find an anti-Pavittian answer. The overwhelming majority of patenting firms appear to be more specialized in terms of innovative knowledge than products.

Third, we study the scaling relation between the size of the firm and diversification of both technological knowledge and product scope. We do find that both scaling relations are essentially log-linear: this supports the idea that firm diversification unfolds as a branching process characterized by incremental accumulation of capabilities.

Finally, we investigate to what extent firms are coherent in their technological and product diversification. Using a refined version of the Teece et al. (1994) measure of corporate coherence, developed in Bottazzi and Pirino (2010), we find that as firms introduce new products and develop new technologies, the coherence between neighboring activities is high for relatively low levels of diversification, but remain present also for sufficiently diversified firms.

2   Dataset description

In order to investigate the characteristics of technological and product diversification, three different microdata have to be linked together.

We take information on Italian firms' patent applications from AMADEUS, a commercial database provided by Bureau van Dijk. The last update at our access (March 2015) covers around one million of Italian companies, of which more than 20,000 turn out to own at least one patent. For those firms, AMADEUS also provides some basic information on patents, including the IPC classification code, the application date, and whether the patent has been granted or not. In order to have a better proxy of technological capabilities, here we focus only on patents that have been applied to USPTO or EPO offices, and that have been effectively granted; this choice restricts the original dataset to a smaller one of 49,803 patents owned by 7,311 firms.[2]

Data on the production activities of Italian firms come from two firm-level datasets collected by the Italian statistical office (ISTAT), namely the Business Register known as "Archivio Statistico Imprese Attive" (ASIA) and Statistiche del Commercio Estero (COE). ASIA is the register of all active Italian businesses. It covers the period 1998-2006 and contains information on firms' operations including the number of employees,

---

[1]On the limitations, but also the relevance of the use of patents as a proxy for innovation, see the critical review by Griliches (1990).

[2]Notice that a single patent may be owned by more than one firm. In case of co-patenting, our analysis credits the patent to each co-patentee, as it is usually done in the literature (see, for example, Breschi et al., 2003).

Table 1: Patents and firms, by period of application and patent office

Period	Total		USPTO		EPO	
	PATENTS	FIRMS	PATENTS	FIRMS	PATENTS	FIRMS
1949-1978	1,086	187	1,086	187		
1979-1995	8,055	1,426	3,929	863	4,126	1,168
1996-2006	**21,305**	**2,946**	**9,817**	**1,647**	**11,488**	**2,499**
2007-2014	9,340	1,948	4,871	1,006	4,469	1,550
1949-2014	39,786	4,411	19,703	2,586	20,083	3,709

Note. Number of USPTO and EPO granted patents owned by Italian firms. The period refers to the application date. Data from AMADEUS, ASIA, and COE.

total turnover,[3] and firm's age, defined as the year of incorporation. We are able to link around 85% of firms in AMADEUS and 90% of their patent applications granted in USPTO and EPO to 2006 ASIA archive. The higher rate of matching for patent applications signals that the firms left out of the combined database are mostly small patentees (the median value of their patent stock is 1). This is partly due to the fact that ASIA archive refers to 2006, whereas AMADEUS also contains information on firms born later on. These young firms are most likely small patentees (see below).

The COE dataset consists of all cross-border transactions performed by Italian firms and it covers the period 2000-2007. COE includes the annual value and quantity of export transactions by the firm at the product level. A product is defined as a six digit category in the Harmonized System (HS6). [4] Using the unique identification code of the firm, we link the firm-level export data from COE 2006 to the ASIA-AMADEUS dataset. After this final link, we are able to maintain information for around 90% of patent applications (with respect to the ASIA-AMADEUS dataset) and 70% of firms. Such a great proportion of patenting firms that are present both in ASIA and in COE dataset is explained by the general high export propensity of patenting firms (see Dosi et al., 2014). Finally, notice that as COE reports information on exported goods, patenting firms that are active in non manufacturing sectors are less likely to be present in COE: in our case, around 70% of non matched firms are active in non manufacturing firms. Once again, the very high rate of matching for patent applications says that patenting firms that exports tend to be, on average, bigger (in terms of patents) of patenting firms that do not export.

Table 1 reports the number of patents and firms present in our final dataset, broken down by date of application and patent office. The first thing to note is that the number of patent applications during the period 1996-2006 overtake all the patent applications that took place before (and after[5]). These figures are consistent with the patent explosion that have been observed both in US and in Europe starting from the mid of 90's (see the discussion in Dosi et al., 2010). Secondly, we observe that the number of patents applied to USPTO by Italian firms is similar to patents applied to EPO, even if there are much more firms applying to EPO. This implies that small patentees are more likely to apply for a patent only in Europe. The joint analysis of USPTO and EPO patents allows us not to loose information about such small patentees while at the same time considering

---

[3]Information on total turnover is available only in 2000 and 2003.

[4]More on the dataset in Grazzi et al. (2013).

[5]Notice that figures relative to more recent years are less reliable due to the time lag between a patent application and its grant.

Table 2: Distribution of firms and patents by number of patents

	No. of patents										Obs.
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10+	
SHARE OF FIRMS	41.68	19.14	9.37	6.45	4.41	3.29	1.66	1.60	1.15	11.24	2,946
SHARE OF PATENTS	5.73	5.26	3.86	3.54	3.03	2.71	1.60	1.75	1.43	71.08	21,441

Source. Amadeus, ASIA, and COE, 1996-2006.

only patents of higher average quality (with respect to the universe of patents).

In the following analysis, we shall proxy firms' technological capabilities in 2006 by considering only patents applied during the period 1996-2006 (in boldface in the table). This ten-year counting period seems a reasonable choice as it discards too "old" patents as a meaningful proxy of firms' capabilities (for similar time periods, see Patel and Pavitt, 1997; Breschi et al., 2003).

Table 2 reports the distribution of firms by number of patents (first row) and also the share of patents accounted for by those firms (second row). The distribution is, as expected, extremely skewed: the typical firm owns just one patent (41.68%), and more than 60% of firms do not have more than two patents. On the right tail of the distribution, there is about only a 10% of firms which have ten or more patents: these firms, however, account disproportionately for the whole stock of patents (71.08%).

In the next sections, we shall analyze the patterns of technological diversification of patenting firms and their relationship with product diversification. In order to do so, we need a classification of technological fields which is consistent with standard industrial and product classification. In this respect, existing IPC-industry concordances are far from being satisfactory as they do not allow to go further than using a 2-digits (in some cases 3-digits) classification of industrial activities[6]. To overcome this limitation, we resort to the probabilistic algorithm recently developed by Lybbert and Zolas (2014) which allows to build a correspondence between technological and production activities of the firm at a more disaggregated level. In particular, we link IPC codes to 125 4-digits ISIC codes (Rev. 3), and SH6 codes to 145 4-digits ISIC codes (Rev. 3). Firms in our sample turn out to patent in 118 different technological fields and produce 138 different products.[7]

Before moving to the main analysis, let us flag some of the characteristics that distinguish patenting from non patenting firms.

3 Characteristics of patenting firms

In this section we exploit our integrated dataset to compare patenting firms with the universe of exporters in 2006. Table 3 reports the differences between patenting and non-patenting firms in terms of size, age, and number of products. Columns (1), (2), and (3) report the β's estimated from the following regression:

$$X_i = \alpha + \beta D_{PAT_i} + \epsilon_i$$

where D_{PAT_i} is a binary variable taking value one if the firm has a granted patent in

---

[6]See in particular the Schmoch et al. (2003) concordance table.

[7]In what follows, we will assume that "products" and "technological fields" map one-to-one into each other (a 4-digit ISIC Rev.3 sector).

Table 3: Size, age, and #products: patenting *vs.* non-patenting firms

Dep. var.	(1) $size_i$	(2) age_i	(3) age_i	(4) $\#products_i$	(5) $\#products_i$
D_{PAT_i}	3.092***	0.320***	0.189***	1.003***	0.371***
	(0.047)	(0.016)	(0.017)	(0.015)	(0.012)
$size_i$			0.042***		0.205***
			(0.001)		(0.001)
N	139,360	139,360	139,360	139,360	139,360
adj. R^2	0.195	0.060	0.074	0.122	0.484
Sector dummies	Yes	Yes	Yes	Yes	Yes

Note. Standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the period 1996-2006, and zero otherwise, and X is either size, measured as the (log) of trade volumes, or (log) age or (log) number of products. Columns (2) and (4) estimate the same equation also controlling for size. Size, age and number of products are measured in 2006.

Patenting firms are both bigger (column 1) and older (columns 2 and 3) than non patenting firms, and they are also more diversified in terms of products, even when controlling for size (columns 4 and 5).

We also investigate within each product-country pair the price (unit value) of paten-tees and non-patentees. For each product-country pair we get two price distributions, one for patentees and one for non-patentees. By employing non-parametric Fligner-Policello (FP) test we study to what extent one distribution dominate stochastically the other. Notice that to compute the statistics we need at least 20 observations both for paten-tees and non-patentees. We find that, when the two distributions statistically differ, in (around) 80% of the product-country pair, patentees have higher price.

4 Technological and product diversification: some stylized facts

There is a remarkable degree of firm heterogeneity both in terms of technological and product diversification. A summary account of the diversification structure of firms in our database is provided in Figures 1,2,3, and 4.

Figure 1 reports the distribution of firms according to the number of different tech-nological fields in which they applied (and got granted) patents in the period 1996-2006. More than 65% of all firms are present in just one technological field, and around 20% in two. Thus, the distribution appears even more skewed than the one observed in Ta-ble 2: many firms do own more than one patent in the same technological field. Another implication of Figure 1 is that around 55% of diversified firms, i.e. firms which are ac-tive in more than one field, are active in just two fields. This figure is similar the one summarized in Breschi et al. (2003) who, however, consider a narrower classification of 30 technological fields. Very few firms are present in several technological fields: the most diversified firm is active in 35 fields (out of 125), with the second most diversified patenting in 23 fields and three firms in 20.

These (few) very diversified firms are large patentees, as it is apparent from Figure 2, which reports the distribution of patents held by firms according to the number of tech-
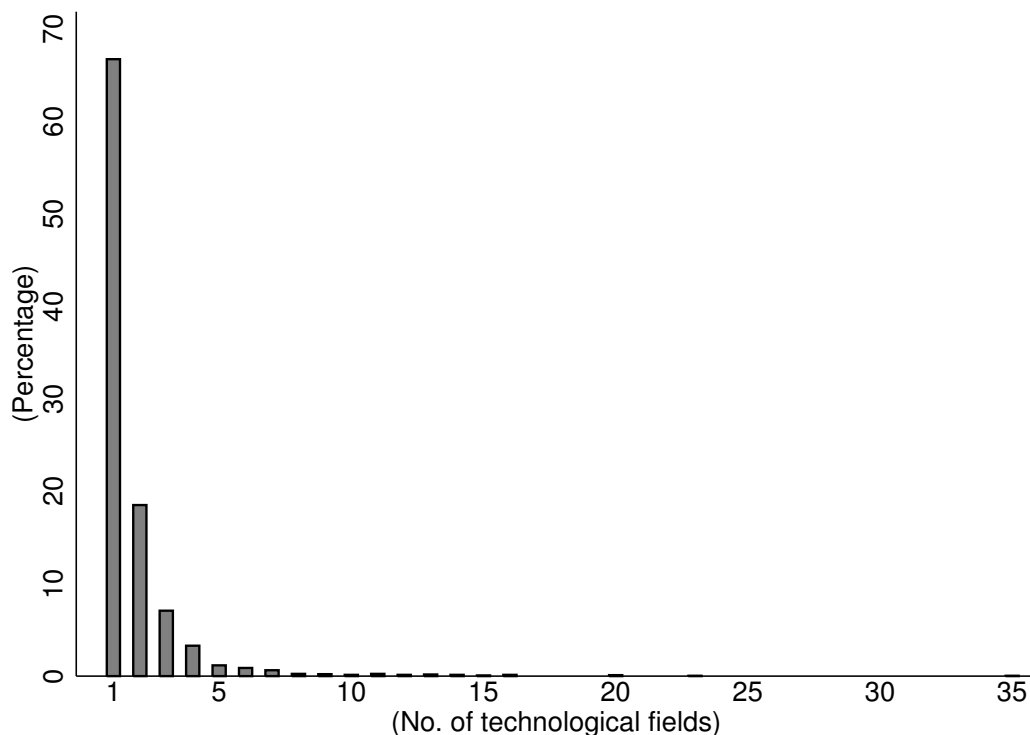
Figure 1: Distribution of firms according to the number of technological fields in which they have a granted patent

nological fields in which those firms applied for a patent in the period 1996-2006. The single firm patenting in 35 different fields owns around 15% of all patents; more in general, firms diversified in at least 10 technological fields account for more than 35% of all patent. On the other hand, firms which patent in less than 3 fields, and that represent around 85% of all firms, account for 30% of all patents.

Figures 3 and 4 report similar distributions for products, with weights given by the export value of each category (Fig. 4). Figure 3 reveals that patenting firms are, on average, quite diversified in terms of products: around 40% of firms produce 10 or more products, and there is also a firm which is active in more than half of the total available products (77 over 145). As it is apparent from Figure 4, firms who diversify a lot across products account disproportionately for total export value, even if there is no single category which produces more than 6% of total export value.

The relative higher frequency of multiproduct firms with respect to multitechnology firms suggests that firms, on average, diversify more across products than across technologies. This result stands in contrast to the evidence put forward by Patel and Pavitt (1997) and Brusoni et al. (2001), according to which firms' technological reach is greater than their product reach. How do we interpret such different views? We try to shed some light on this issue by analyzing the joint distribution of firms across number of products and technological fields, reported in Table 4.

Each row distributes the total number of firms active in a given number of technological fields according to the number of products they produce. Considering the first row, which contains all the firms active in just one technological field, we see that around 90% of them produce more than one product. In general, firms on the left of the main
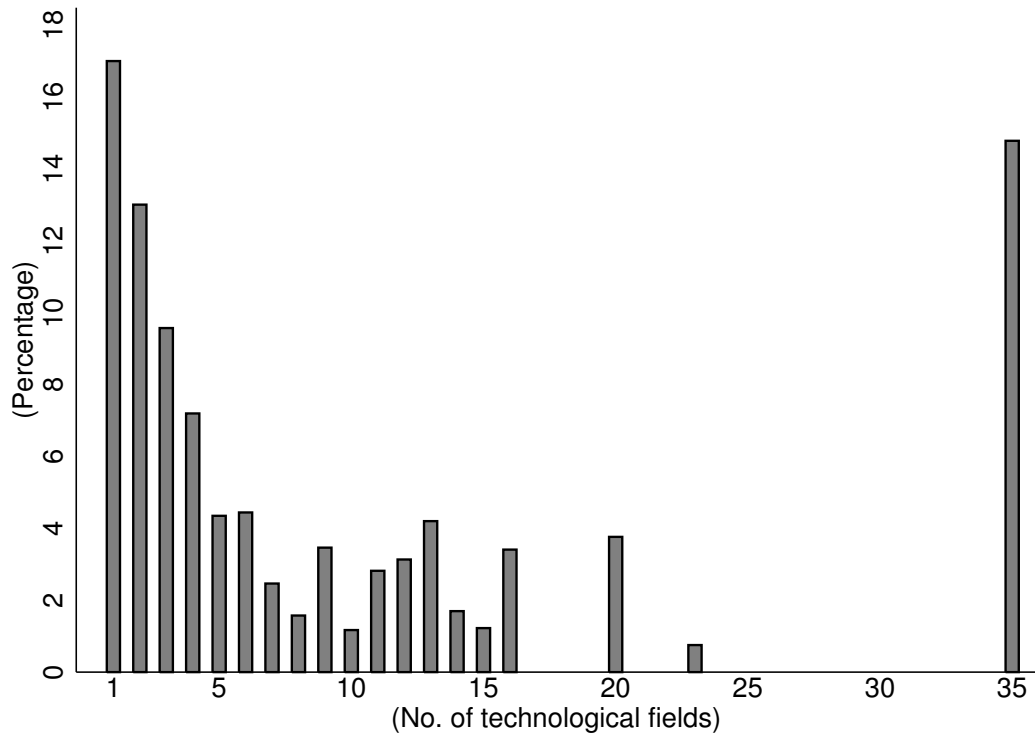
7

Figure 2: Distribution of total patents owned by firms according to the number of technological fields in which they have a granted patent
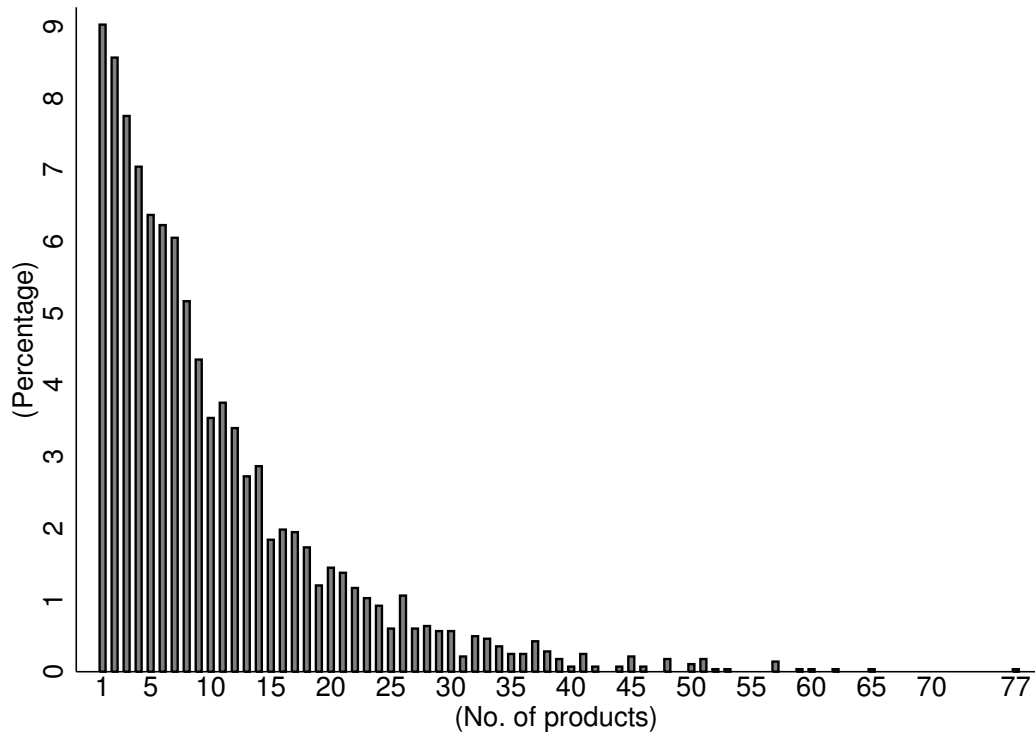

Figure 3: Distribution of firms according to the number of products they produce
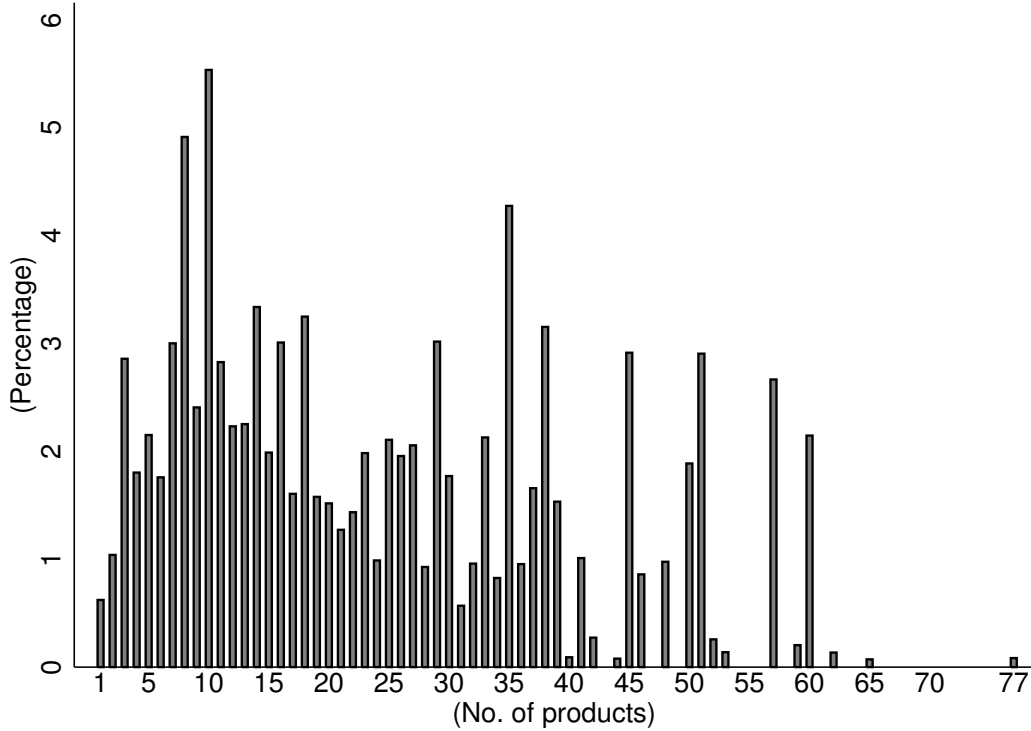
Figure 4: Distribution of total export value produced by firms according to the number of products they produce (export)

diagonal (i.e. firms diversified more in terms of technologies than in terms of products) are quite rare. They are relatively more present among firms active in many technological fields: "Pavitt" firms, who know more than they make, are mostly limited to the few large patentees.

Finally, we analyze to what extent firms knowledge is related to their product portfolios. In Table 5 we report the percentage matching between the product classification and the technological classification for firms producing different number of products. In general, the matching is not very high: for example, it is about 18% for firms producing only one product. This of course may reflect some noise in the concordance provided by the Lybbert and Zolas (2014) algorithm. However, interestingly, as Table 5 illustrates, the matching is much higher for the products that, within each firm, account for most of its export. This evidence seems to suggest that firms tend to exploit innovative knowledge more related to their main products.

Table 4: Distribution of firms by number of technological fields and products

#Technological fields	1	2	3	4	5	6	7	8	9	10+	Total
1	208	194	168	156	136	130	115	95	81	603	1,886
	(11.03)	(10.29)	(8.91)	(8.27)	(7.21)	(6.89)	(6.10)	(5.04)	(4.29)	(31.97)	(100.00)
2	35	38	38	31	32	27	40	31	23	228	523
	(6.69)	(7.27)	(7.27)	(5.93)	(6.12)	(5.16)	(7.65)	(5.93)	(4.40)	(43.59)	(100.00)
3	5	6	7	9	6	15	10	13	10	119	200
	(2.50)	(3.00)	(3.50)	(4.50)	(3.00)	(7.50)	(5.00)	(6.50)	(5.00)	(59.50)	(100.00)
4	4	4	3	1	3	1	4	3	6	64	93
	(4.30)	(4.30)	(3.23)	(1.08)	(3.23)	(1.08)	(4.30)	(3.23)	(6.45)	(68.82)	(100.00)
5	2	0	0	1	0	2	0	1	3	24	33
	(6.06)	(0.00)	(0.00)	(3.03)	(0.00)	(6.06)	(0.00)	(3.03)	(9.09)	(72.73)	(100.00)
6	0	0	0	0	0	0	1	1	0	23	25
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(4.00)	(4.00)	(0.00)	(92.00)	(100.00)
7	0	0	0	1	0	0	0	1	0	16	18
	(0.00)	(0.00)	(0.00)	(5.56)	(0.00)	(0.00)	(0.00)	(5.56)	(0.00)	(88.89)	(100.00)
8	0	0	2	0	1	0	0	0	0	4	7
	(0.00)	(0.00)	(28.57)	(0.00)	(14.29)	(0.00)	(0.00)	(0.00)	(0.00)	(57.14)	(100.00)
9	0	0	0	0	0	0	1	0	0	5	6
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(16.67)	(0.00)	(0.00)	(83.33)	(100.00)
10+	1	0	1	0	2	1	0	1	0	29	35
	(2.86)	(0.00)	(2.86)	(0.00)	(5.71)	(2.86)	(0.00)	(2.86)	(0.00)	(82.86)	(100.00)
Total	255	242	219	199	180	176	171	146	123	1,115	2,826
	(9.02)	(8.56)	(7.75)	(7.04)	(6.37)	(6.23)	(6.05)	(5.17)	(4.35)	(39.46)	(100.00)

Note. Absolute and percentage (in brackets) frequencies.

Table 5: Matching between technological fields and products

Product rank	# Products									
	1 (745)	2 (566)	3 (880)	4 (564)	5 (816)	6 (501)	7 (652)	8 (1,107)	9 (561)	10+ (13,941)
1	17.32	21.38	17.27	24.82	19.98	19.56	13.50	36.04	29.23	23.61
2		4.95	4.89	10.99	10.17	6.79	4.60	2.17	3.74	6.17
3			6.14	1.06	1.96	2.79	6.60	2.44	1.25	3.92
4				2.49	2.21	4.59	1.23	3.52	2.14	1.74
5					1.03	2.40	2.45	3.99	0.18	1.46
6						0.60	0.31	0.45	0.71	1.75
7							2.15	0.90	3.92	0.91
8								0.45	1.25	3.72
9									1.07	1.66
10+										0.81

Note. Each cell reports the percentage matching between technological fields and product categories across the relevant set of firm-products. In parentheses, the number of patents for the relevant set of firm-products.

5 Scaling and diversification

The previous section has highlighted two main phenomena. First, multitechnology firms are quite rare, much more than multiproduct firms. Second, firms that diversify their activities across different products or different technologies account disproportionately for the total exports and the total patents. This evidence hints at some underlying relation between size and firm diversification: bigger firms in terms of patents or exports are also those which explore more opportunities in the product and technological space. Here, let us analyze in greater detail the scaling properties that relate products and technological fields to exports and patents.

In Figure 5(a), we plot the average number of products of the firms belonging to different export bins against the bin average log export. A simple linear fit does not seem to capture very well the observed correlation, whereas a clear positive exponential relation between the two variables emerges. As a consequence, we plot on a log-log scale the (binned) number of products against (binned) total exports (Figure 5(b)). In this case, fitting a linear relation of the form:

$$log(\#products_i) = \alpha + \beta log(exports_i) + \epsilon_i \tag{1}$$

delivers a strongly significant slope $\beta = 0.26$ with a standard error of 0.01.

In Figure 6(a), we repeat the same exercise as before by plotting the binned relationship between technological fields and log patents. It is worth stressing that in this case the scantiness of observations related to large patentees make it more difficult to appreciate the exponential fit in the linear-log space. However, a log-log specification like (1) seems to capture adequately the relationship in Figure 6(b), with a slope $\beta = 0.43$ and a standard error of 0.01.

The evidence about the scaling properties of products and technological fields has two main implications, already pointed out in Bottazzi et al. (2001). First, the increase in products and technologies we observe as firms get bigger is proportional to their growth rate. To put it differently, firms are able to enter in more and more activities (products and technologies) as they become bigger. The scaling properties themselves are quite robust: bigger firms diversify more in terms of both technological fields and products (with the causal mechanism obviously running both ways: firms might well grow via diversification!).

Second, the scaling factor < 1 implies that a large firm is less diversified than a collection of small (single product) firms which add up to the same size of the large one. In turn that implies that a large firm is more risky than a collection of smaller firms: a notional investor would face a lower risk by diversifying their portfolio in the N (independent) firms of size S each rather than betting on one single firm of size NS.

From an interpretative perspective all this militates against the hypothesis that diversification is driven by risk-minimizing considerations. Rather, the evidence may be plausibly interpreted in terms of competence-driven diversification processes, in the presence of knowledge spillovers across products and lines of search. In fact, as formally discussed in Bottazzi and Secchi (2006), the observed diversification patterns can essentially be described by a stochastic branching process: its economic interpretation may be plausibly grounded in the incremental development of knowledge basis, driving the exploration of an expanding range of products/technologies.
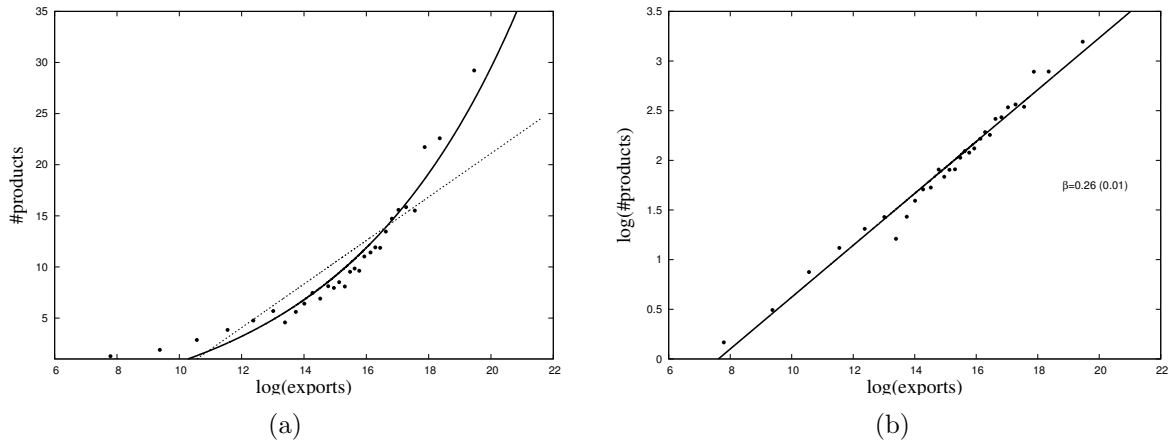
Figure 5: (a) Binned relationship between log(exports) and #products together with a linear (dotted line) and an exponential (solid line) fit
(b) Binned relationship between log(exports) and log(#products) together with a linear fit
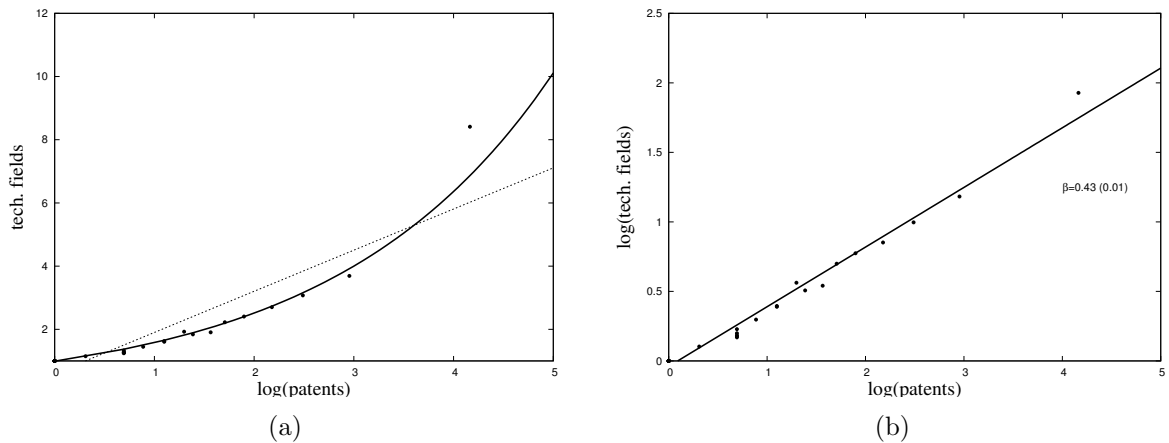

Figure 6: (a) Binned relationship between log(patents) and #technological fields together with a linear (dotted line) and an exponential (solid line) fit
(b) Binned relationship between log(patents) and log(#technological fields) together with a linear fit

6 Diversification and firm coherence

If diversification is a competence-driven process, one should expect it to be associated with non-decreasing levels of firm coherence. Indeed, coherent patterns of firm diversification are predicted by capabilities-based and evolutionary views of the firm: these theories show how opportunities of diversification are shaped by technological imperatives and path-dependent learning dynamics within the firm (Teece et al., 1994; Dosi et al., 2000; Winter, 2003). Notice that in these perspectives the technological and the product side are strictly connected in defining corporate coherence, following the definition proposed in Teece et al. (1994):

> A firm exhibits coherence when its lines of business are related, in the sense that there are certain technological and market characteristics common to each.

Hence one should expect firms to exhibit a coherent pattern of diversification both with respect to products and with respect to technologies. In the following, let us test this prediction.

6.1 Measuring firm coherence

In order to define a measure of corporate coherence, one first needs a notion of *relatedness*. We adopt a "survivor" measure of relatedness which has been proposed by Teece et al. (1994) and then widely used in the literature, with reference both to industrial sectors and products (Piscitello, 2000, 2004; Valvano and Vannoni, 2003) and to patent classes (Breschi et al., 2003; Nesta and Saviotti, 2006; Bottazzi and Pirino, 2010).

The basic notion underlying such a survivor measure is that if some selection mechanism, however rough, is at work "selecting out" relatively inefficient techniques and products, and if relatedness confers some advantage, then in Teece et al. (1994) one captures that related activities will appear with higher frequency within the same firm. To operationalize the concept, let $C_{ik} = 1$ if firm k is active in product (or technological field) i, and 0 otherwise. The total number of firms active in both products (or technological fields) i and j is therefore given by:

$$J_{ij} = \sum_k C_{ik} C_{jk} \tag{2}$$

By applying (2) to all possible pairs of products or technological fields, we calculate a square symmetrical matrix C whose generic cell J_{ij} reports the number of firms active in both products i and j (or in both technologies i and j).[8]

Finally, we compare the observed value of each J_{ij} in the matrix with the value that would be expected under the null hypothesis that diversification is random. A natural measure of relatedness, suggested by Bottazzi and Pirino (2010) (see also Breschi et al., 2003), is then given by the P-value:

$$p_{ij}(J, H) = Prob[\tilde{J}_{ij} \leq J_{ij}|H] \tag{3}$$

where $\tilde{J}_{ij}$ is the value of the relative cell under the null hypothesis H.[9] Notice that p_{ij} can be greater than, equal, or less than 0.5, implying that activities i and j are positively related, not related, or negatively related.

It is clear that p_{ij} as from Equation (3) depends, in general, on the choice of the null hypothesis H. Starting from Teece et al. (1994), the relatedness literature has usually assumed as a null hypothesis a random distribution of firms across activities, under the only constraint that the total number of firms active in a given sector (or product or patent class) is fixed and equal to the one observed in the actual data. Under this hypothesis, which we will refer to as the standard null hypothesis, the probability to obtain a given value of $\tilde{J}_{ij}$ is distributed according to a hypergeometric random variable.

---

[8]The matrix is calculated taking into account only firms which are active in at least two products (technological fields), and only products (technological fields) in which at least two firms are active. The resulting matrix is $136x136$ in the case of products (taking into account 2,571 firms) and $102x102$ in the case of technological fields (taking into account 940 firms).

[9]When making inference about skewed distributions, as for example distributions of firms across products or technological fields, P-value is to be preferred to statistics-based inference, like the t-statistics used in Teece et al. (1994). The point is discussed at greater length in Bottazzi and Pirino (2010).

The standard null hypothesis does not impose any constraint on the number of activities in which each firm is present: in principle, a firm could be present in all products or technological classes. As noted in Bottazzi and Pirino (2010), one implication of this assumption is that the implied distribution of firm scope converges to a binomial, in contrast with the Paretian shape that is usually observed in real data (and also in our dataset: see section 4, Figures 1 and 3). To overcome this drawback, Bottazzi and Pirino (2010) propose a different null hypothesis, according to which both firms scope and the number of firms per activities are fixed and correspond to the observed ones. We will refer to this hypothesis as the alternative null hypothesis.

In the following analysis, we shall use both null hypotheses. The standard null hypothesis allows to compare our results with the ones usually found in the literature. The alternative null hypothesis tests the robustness of the results with respect to more restrictive assumptions. Notice that deriving the P-value under the alternative null hypothesis is, in general, not trivial. Here, we adopt the Monte Carlo algorithm developed and explained at greater length in Bottazzi and Pirino (2010).

Equipped with our P-values, we can now calculate a modified version of two largely diffused measures of firm coherence. The first one is based on the weighted-average relatedness of activity i to all other activities within the firm and is defined as:

$$WAR_k(H) = \frac{1}{n} \sum_i C_{ik} \left(\frac{\sum_{j \neq i} p_{ij}(H) w_{jk}}{\sum_{j \neq i} w_{jk}} \right) \tag{4}$$

where n is total number of products (technological fields) in which a firms is active and w_{jk} the weight of product (technological field) j with respect to firm k. We weight products with export share and technological fields with patent count.

The second one is based on the weighted average relatedness of neighbors and is defined as:

$$WARN_k(H) = \frac{1}{n} \sum_i C_{ik} \left(\frac{\sum_{j \neq i} p_{ij}(H) m_{ij}^k w_{jk}}{\sum_{j \neq i} m_{ij}^k w_{jk}} \right) \tag{5}$$

where $m_{ij}^k = 1$ if the pair ij is in the maximum spanning tree of firm k, defined as the graph with $n-1$ links such that the sum of the relatedness measures on each link is largest.

Notice that WAR and $WARN$ capture two different aspects of firm diversification structure. WAR measures the inverse of the average distance from a firm activity to all other activities, whereas $WARN$ measures the inverse of the average distance from a firm activity to its neighbor activity. As firms diversify in more and more products and technological fields, one would expect the WAR measure, on average, to increase. However, if the competence-driven branching process is a reasonable account of firm diversification, the $WARN$ measure should not be affected by the degree of diversification, as firm should add products or technologies that are near to some of its previous products and technologies.

6.2 Results

In Figure 7 (a) and (c), we plot the average WAR_k of the firms belonging to different #products bins against the bin average #products under, respectively, the standard and the alternative null hypothesis. As expected, as firms increase their product scope, the coherence across all its activities decrease. The observed relationship seems to be linear
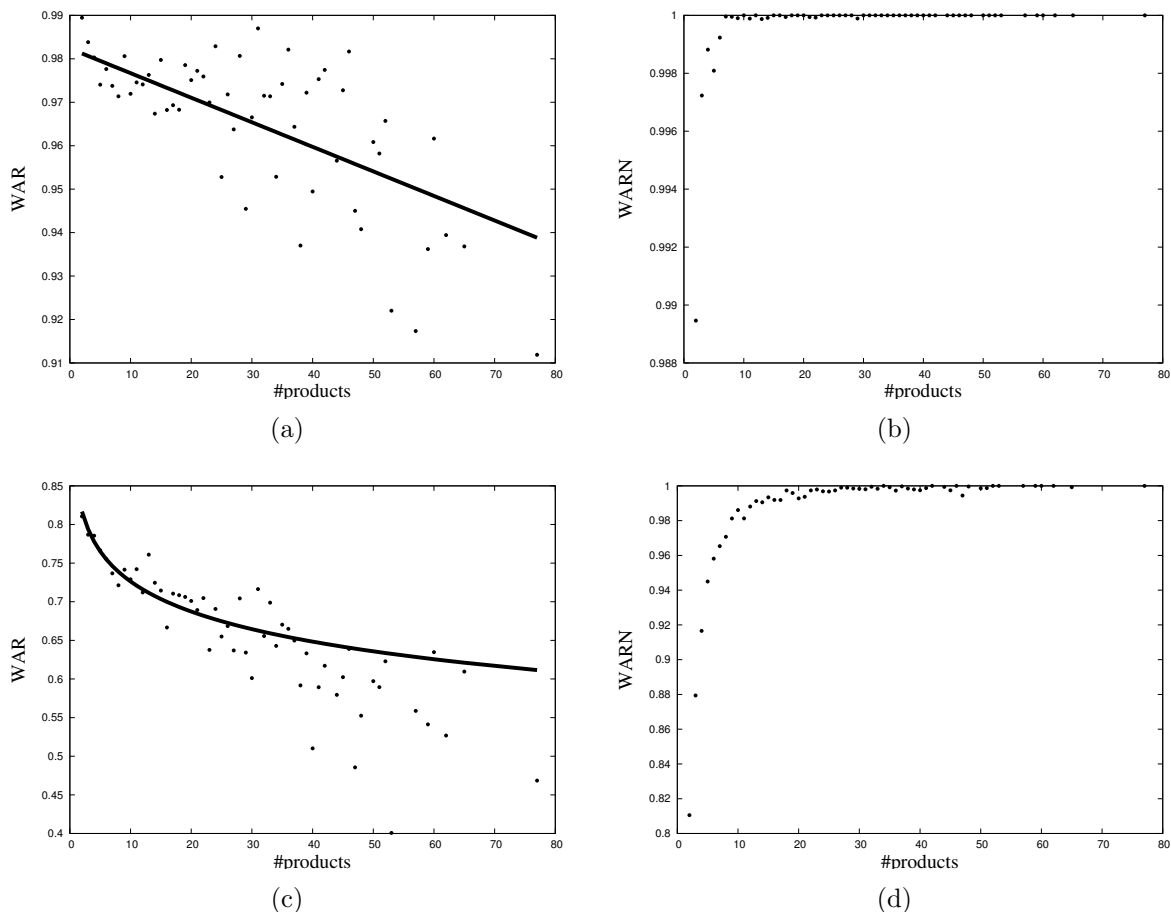
Figure 7: Binned relationship between product WAR (left) and WARN (right) as a function of #products. Standard null hypothesis (top) and alternative null hypothesis (bottom). Linear fit in (a) and log-linear fit in (c).

under the standard null hypothesis, but log-linear under the alternative hypothesis. The interpretation is that if heterogeneity in firm scope is properly accounted for, then the addition of a new product to small firms reduce coherence much more than in large firm (see Bottazzi and Pirino, 2010).

In Figure 7 (b) and (d), we plot the same binned relationship for the $WARN_k$ measure. It is apparent that as firms introduce new products the coherence between neighboring activities slightly increase for relatively low levels of diversification, and stay constant for sufficiently diversified firms.

In Figure 8 (a)-(d), we repeat the same exercise as before by plotting the binned relationship between #technological fields and coherence measures. The scantiness of observations related to diversified firms make it difficult to appreciate any real difference between (a) and (c): in general, no clear relationship seems to emerge between levels of diversification and average coherence across all technological fields. It is clear, however, from (b) and (d) that also in the case of technological fields, diversification is associated to constant levels of coherence, in agreement with our hypothesis of competence-driven diversification and with the original intuition of Teece et al. (1994).
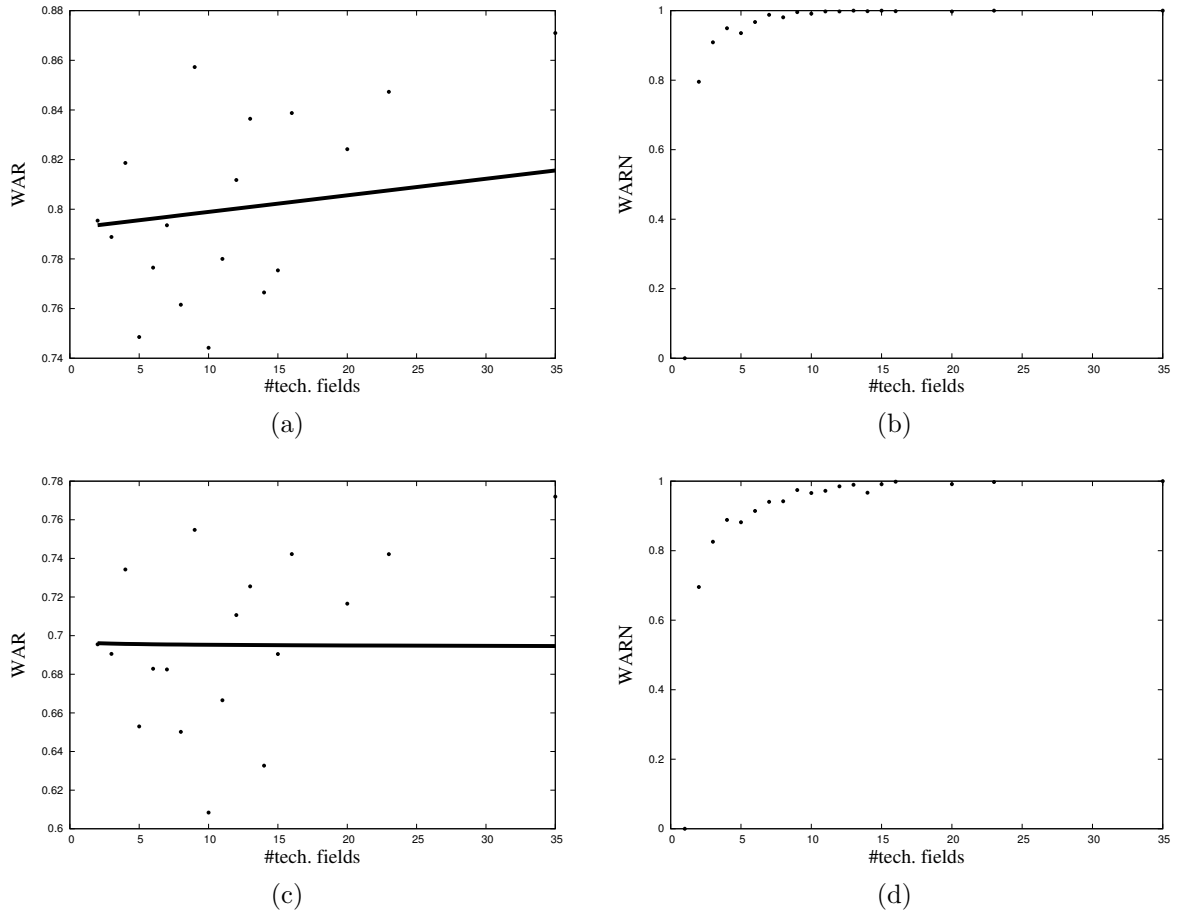
Figure 8: Binned relationship between technological WAR (left) and WARN (right) as a function of #technological fields. Standard null hypothesis (top) and alternative null hypothesis (bottom). Linear fit in (a) and log-linear fit in (c).

7 Conclusions

The paper contributes to the analysis of the main patterns of firm diversification both in the technological and in the product space. Using a newly developed dataset matching information on patents and products at the firm level, we have provided evidence concerning firms' technological and product scope, their relationships, and the size-scaling properties of diversification itself.

Our analysis shows that typically firms are much more diversified in terms of products than in terms of technologies, even if their main products are related to the exploitation of their (seemingly narrow) innovative knowledge. Interestingly, "Pavitt" firms, who know more than they make, are quite rare, mostly limited to the few large patentees. At the risk of overtheorizing, the picture of the Italian national system of innovation and production appears to be one of complementarity between few relatively large members of international oligopolies - especially in relatively new technological paradigms - which "know more than they do", and a large cloud of relatively narrowly specialized firms.

At the same time, the analysis has also shown some similarities between the technological and the product diversification processes. In particular, the scaling properties of both show that the number of products and technologies increase log-linearly as firms grow. And the directions of diversification themselves display coherence between neighboring activities also at relatively high degrees of diversification (consistently with the conjecture put forward in Teece et al. (1994)). Both findings are well in tune with a capability-based theory of the firm (for a critical survey within a vast literature, see Dosi et al. (2008)). The patterns of diversification - both in technologies knowledge and in products - are consistent with a branching process whereby knowledge on production and innovation, so to speak, "spurs out" from what the firm already does and knows. And of course, the coherence in the directions of diversification reinforced the point: where a firm stands in term of pre-existing capabilities shapes to a good extent where it will go.

References

Bottazzi, G., G. Dosi, M. Lippi, F. Pammolli, and M. Riccaboni (2001): "Innovation and corporate growth in the evolution of the drug industry," *International Journal of Industrial Organization*, 19, 1161–1187.

Bottazzi, G. and D. Pirino (2010): "Measuring Industry Relatedness and Corporate Coherence," LEM Papers Series 2010/10, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy.

Bottazzi, G. and A. Secchi (2006): "Gibrat's Law and diversification," *Industrial and Corporate Change*, 15, 847–875.

Breschi, S., F. Lissoni, and F. Malerba (2003): "Knowledge-relatedness in firm technological diversification," *Research Policy*, 32, 69 – 87.

Brusoni, S., A. Prencipe, and K. Pavitt (2001): "Knowledge Specialisation, Organizational Coupling and the Boundaries of the Firm: Why Firms Know More Than They Make?" *Administrative Science Quarterly*, 46, 597–621.

Dosi, G. (1988): "Sources, Procedures, and Microeconomic Effects of Innovation," *Journal of Economic Literature*, 26, 1120–71.

DOSI, G., M. FAILLO, AND L. MARENGO (2008): "Organizational Capabilities, Patterns of Knowledge Accumulation and Governance Structures in Business Firms: An Introduction," *Organization Studies*, 29, 1165–1185.

DOSI, G., M. GRAZZI, AND D. MOSCHELLA (2014): "Technology and costs in international competitiveness: from countries and sectors to firms," LEM Papers Series 2014/10, Laboratory of Economics and Management (LEM), Sant'Anna School of Advanced Studies, Pisa, Italy.

DOSI, G., L. MARENGO, AND C. PASQUALI (2010): "How much should society fuel the greed of innovators? On the relations between appropriability, opportunities and rates of innovation," in *The Capitalization of Knowledge*, ed. by R. Viale and H. Etzkowitz, Edward Elgar.

DOSI, G., R. R. NELSON, AND S. WINTER (2000): "Introduction," in *The nature and dynamics of organizational capabilities*, ed. by G. Dosi, R. R. Nelson, and S. Winter, Oxford University Press: Oxford.

GRAZZI, M., R. SANZO, A. SECCHI, AND A. ZELI (2013): "The building process of a new integrated system of business micro-data 1989-2004," *Journal of Economic and Social Measurement*, 38, 291–324.

GRILICHES, Z. (1990): "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28, 1661–1707.

HALL, B. H. AND J. MAIRESSE (1995): "Exploring the relationship between R&D and productivity in French manufacturing firms," *Journal of Econometrics*, 65, 263–293.

HIRSCH, S. AND B. LEV (1971): "Sales Stabilization Through Export Diversification," *The Review of Economics and Statistics*, 53, 270–77.

LYBBERT, T. J. AND N. J. ZOLAS (2014): "Getting patents and economic data to speak to each other: An Algorithmic Links with Probabilities approach for joint analyses of patenting and economic activity," *Research Policy*, 43, 530–542.

MILLER, D. J. (2006): "Technological diversity, related diversification, and firm performance," *Strategic Management Journal*, 27, 601–619.

MONTGOMERY, C. A. (1994): "Corporate diversification," *The Journal of Economic Perspectives*, 8, 163–178.

NELSON, R. R. AND S. G. WINTER (1982): *An Evolutionary Theory of Economic Change*, The Belknap Press of Harvard University Press: Cambridge, MA.

NESTA, L. AND P.-P. SAVIOTTI (2006): "Firm knowledge and market value in biotechnology," *Industrial and Corporate Change*, 15, 625–652.

PATEL, P. AND K. PAVITT (1997): "The technological competencies of the world's largest firms: Complex and path-dependent, but not much variety," *Research Policy*, 26, 141–156.

Pavitt, K. (1998): "Technologies, Products and Organization in the Innovating Firm: What Adam Smith Tells Us and Joseph Schumpeter Doesn't," *Industrial and Corporate Change*, 7, 433–452.

Piscitello, L. (2000): "Relatedness and coherence in technological and product diversification of the world's largest firms," *Structural Change and Economic Dynamics*, 11, 295–315.

——— (2004): "Corporate diversification, coherence and economic performance," *Industrial and Corporate Change*, 13, 757–787.

Schmoch, U., F. Laville, P. Patel, and R. Frietsch (2003): "Linking technology areas to industrial sectors," *Final Report to the European Commission, DG Research.*

Teece, D. J., R. Rumelt, G. Dosi, and S. Winter (1994): "Understanding corporate coherence: Theory and evidence," *Journal of Economic Behavior & Organization*, 23, 1–30.

Valvano, S. and D. Vannoni (2003): "Diversification Strategies and Corporate Coherence Evidence from Italian Leading Firms," *Review of Industrial Organization*, 23, 25–41.

Wakelin, K. (1998): "Innovation and export behaviour at the firm level," *Research policy*, 26, 829–841.

Winter, S. G. (2003): "Understanding dynamic capabilities," *Strategic management journal*, 24, 991–995.