INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

What a firm produces matters: diversification, coherence and performance of Indian manufacturing firms

Giovanni Dosi [a]
Nanditha Mathew [b,a,c]
Emanuele Pugliese [d,e]

[a] Institute of Economics & EmbeDS, Scuola Superiore Sant'Anna, Pisa, Italy.
[b] UNU-MERIT, Maastricht, Netherlands.
[c] IBIMET-CNR, Florence
[d] European Commission, Joint Research Centre (JRC), Seville, Spain.
[e] Institute of Complex Systems, CNR, Rome.

2019/10 April 2019
ISSN(ONLINE) 2284-0400

What a firm produces matters: diversification, coherence and performance of Indian manufacturing firms *†

Giovanni Dosi[1], Nanditha Mathew[2,1,3], and Emanuele Pugliese[4,5]

[1]Institute of Economics, Sant'Anna School of Advanced Studies, Pisa
[2]UNU-MERIT, Maastricht
[3]IBIMET-CNR, Florence
[4]European Commission, Joint Research Centre (JRC), Seville, Spain
[5]Institute of Complex Systems, CNR, Rome

Abstract

Economic growth and development of a country involves accumulation of knowledge and dynamic capabilities (Cimoli et al., 2009). Past research has begun to investigate the capability accumulation and macro-economic development of countries and sectors (Dosi et al., 1990), also by means of introduction of new products (Hausmann and Rodrik, 2003). In this work, recognizing that firms are the actual domain in which production takes place, we focus on the firm-level process of capability accumulation and diversification in a developing country. We investigate the relationship between diversification (and coherent diversification) and firm performance by employing an extensive database of Indian manufacturing firms with detailed information on product mix of firms. We claim that such an understanding of firms' incentives to diversify is relevant not only for the corporate management, but also for the diversification of countries and thereby its development.

First, we explore the reasons behind firms' strategy to diversify, i.e, which firms choose a broad product scope and whether the change in the scope of the firm results in improved performance in terms of firm profitability and sales growth. *Second*, we look at the idiosyncratic characteristics of different products, by emphasizing the synergies of a product line with respect to the overall product basket of the firm. In this line, we develop a measure that captures the synergies and economies of scope between different products, and observe that the firms' future performance crucially depend on the interactions between the products that comprise its basket. Overall, our results are consistent with an intangible-capabilities model of firm diversification: diversification results in improved firm performance if the firm has underused capabilities and the new production line is able to exploit them.

1 Introduction

In this work we study the firm level process of diversification and its effect on performances, focusing in particular the Indian manufacturing sector. The understanding of the motives behind firms' diversification

*Acknowledgements: We thank Antonio Andreoni, Masud Cader, Mercedes Campi, Alex Coad, Davide Consoli, Marco Duenas, Davide Fiaschi, Andrea Mina, Lorenzo Napolitano, Luciano Pietronero, Andrea Tacchella and Andrea Zaccaria, as well as participants at ISS 2016 (Montreal), EMAEE 2017 (Strasbourg) and INGENIO (Valencia) for their insightful comments. We gratefully acknowledge the support by the European Union Horizon 2020 Research and Innovation programme under grant agreement No. 822781 - GROWINPRO. Nanditha Mathew acknowledges the research support by the IBIMET-CNR (grant CRISISLAB-ProCoPe). Emanuele Pugliese acknowledges the financial contribution of project CRISISLAB (Progetto d'Interesse Nazionale, Italian Ministry of Research). The usual disclaimers apply.

†The content of this article does not reflect the official opinion of the institutions they belong to. Responsibility for the information and views expressed therein lies entirely with the authors.

process is not only a central validation of the capability approach to firm dynamics, but also a requirement to design industrial policies, and this is particularly true for developing countries.

The evolutionary and institutionalist literature, both empirical and theoretical, describe the development of a country as a great societal transformation involving the emergence of new institutions, structural change, and the accumulation of knowledge and dynamic capabilities (Prebisch, 1950; Hirschman, 1958; Nelson, 1994; Cimoli et al., 2009), in open contrast with simplistic mainstream theories predicting an "automatic" quantitative economic convergence without qualitative economic transformation (see also Dosi and Tranchero (2018) for a discussion). Differences in capabilities between countries are reflected both quantitatively, for instance through persistent differences in productivity (Dosi et al., 1990; Cimoli and Dosi, 1995), and on several qualitative aspects of the economy: industrial and market organization (Dosi, 1984), structural change and industrial composition (McMillan and Rodrik, 2011) etc. At the same time, the industrial composition and the choices of sectoral specialization are main drivers of long term growth (Cimoli et al., 2008; Castaldi et al., 2009), leading to overall different country trajectories. While some of these qualitative aspects are not easy to measure or to model, it is clear how the co-evolution between capabilities accumulation, production and export diversification, and economic growth can hardly be overvalued when looking at development. A recent attempt to exploit such link to quantitatively measure unobservable capabilities comes from the Economic Complexity literature (Hausmann et al., 2007; Hidalgo et al., 2007; Tacchella et al., 2012), looking at properties of the trade network and at the export basket of countries to extract information on their under-laying capabilities endowments.

The relationship between country-level capability accumulation and the evolution of industrial organizations in the country, i.e. the connection between macro-development and micro-adaptation of firms, is clearly central in this capability approach to development. It is therefore surprising that the micro-dimension is almost absent in the Economic Complexity literature. Although the possible barriers in terms of market failures at the micro level are acknowledged (Hausmann and Rodrik, 2003), researchers studying country diversification patterns often ignore the parallel process of capability accumulation at firm level and, consequently, the constraints faced by the firms to diversify their production in relation to their capabilities. However, such an understanding is crucial to properly address development through capability accumulation in relation to an expansion of country's export basket. Indeed, the introduction of a new product in a country does not happen in vacuum: firms are the domain in which economic development really takes place (Teece, 2000). To give an example, by looking at the 128 new products[1] that are introduced in India in the years from 1990 to 2012, only 9 are produced by new entrant firms. In the vast majority of cases, the country diversification is also an outcome of a diversification event of an existing firm.

To address the question of why and how firms diversify has therefore significance both from a macro and micro perspective. A firm diversify since it possess the "critical resources" (as Barney (1997) puts it) that are shared by similar business units and hence a firm can put to use the organizational capabilities accumulated in different activities as well. It is important to notice that, while such critical resources and capabilities can be generic management skills (book-keeping, human resource management,...), they can be as well technological and product specific capabilities. Both are equally important (Teece et al., 1997), in particular in developing countries where basic management capabilities can be harder to find (Cireira and Maloney, 2017). However, the answer to why related diversification is more successful than unrelated diversification lies in the the nature of product and sector specific organizational capabilities, which is essentially a product of organizational learning (Dosi et al., 2000). As pointed out by (Dosi, 1988, p. 1130), because technological change is cumulative in nature, "what a firm can hope to do in the future is narrowly constrained by what it has been capable of doing in the past". What the firm knows narrowly constraints what the firm can do (Brusoni et al., 2001; Dosi et al., 2017b), and a firm is therefore better in doing what the firm has already learned to produce. Hence, from this perspective, firms with related diversification strategies can outperform those with unrelated diversification strategies. Bottazzi and Secchi (2006) observes a firm diversification structure that can be described by a stochastic branching process, where the ability of a firm to enter in new sub-markets is related to the number of sub-markets it is already active in. This is again in

[1] see section 3 for the definition of products

line with the idea that activities of a firm is constrained by the competencies that it inherited from its past.

When a firm enters a related market it can therefore perform better by exploiting its existing knowledge. In this work, we analyze the relationship between diversification, coherence of the product basket and firm performance. Our findings are consistent with a capability-based theory of the firm (Teece, 1982)[2]. This gives the necessary theoretical and empirical microfoundation to models connecting country diversification and specialization with catching up and development.

The paper is structured as follows. Section 2 provides a brief discussion on previous works on diversification, coherence and performance, while section 3 describes the dataset employed. Following this, empirical questions will be divided into two sections. Section 4 frames the econometric methodology and presents results concerning diversification and firm performance. Section 5 presents the measure of relatedness and details on the relationship between the coherence of the product basket of the firm and its performance. Finally, section 6 concludes.

2 Theoretical background and review of literature

The empirical discussion about how firms diversify has been neglected not only by the Economic Complexity literature, as we pointed out before. As argued in McKelvie and Wiklund (2010), much of the micro-econometric research on firm growth has been concentrated on "how much" firms grow, instead of seeking answers to "how" firms grow. This is particularly true in the case of understanding diversification patterns of growth and how firms diversify to achieve improved performance. As Penrose (1959/1995) pointed out, growth is a process in which "increases in size [are] accompanied by changes in the characteristics of the growing object" (Penrose 1995, p. 1). To bridge this gap will be the main purpose of this work: to investigate "how firms grow" by assessing their diversification patterns and how those relates to their performance.

In this section, let us look carefully to the incumbent literature on the relationship between products and performances of the firm. This section is divided into two parts. The first part concentrates on diversification and firm performance, while the second details on coherent diversification and firm performance, also reviewing the existing measures of relatedness.

2.1 The scope of firms and their performance

Developing on the resource-based theory of the firm Penrose (1959/1995), the capability-based theory suggests that diversification of the firm is significantly dependent on the intangible capabilities it possesses. If the firm possesses underused "excess" capabilities, it could benefit by putting those excess capabilities to use by diversifying its production lines. Differently from Penrose (1959/1995), however, the capability approach do not focus on physical resources, since they are highly mobile. If there are no transaction costs involved, the "excess" resources could just be traded. Therefore, as Teece et al. (1994) points out, the focus is on intangible capabilities – i.e. the skills of the firm that are mostly embedded in its routines (Nelson and Winter, 1982). Indeed, given that routines involve a strong tacit dimension, they can not be imitated or transferred. Even though such intangible resources are not transferable across firms, the transfer of such excess capabilities inside the firm to related and complementary businesses (for instance, another related product) could be an optimal strategy.

There is indeed no clear agreement in the empirical literature on which firms diversify, or whether diversification improves firm performance. The results from the existing work on diversification and firm performance are still inconclusive: Studies have found a positive linear relationship (Michel and Shaked, 1984; Rumelt, 1982; Delios et al., 2008; Jose et al., 1986), a negative relationship (Lang and Stulz, 1994; Chen and Ho, 2000; Berger and Ofek, 1995; Markides, 1995), a U-shaped curvilinear relationship (Khanna and Palepu, 2000; Palich et al., 2000), and no relationship (Christensen and Montgomery, 1981; Delios and Beamish, 1999). Most of these studies focused on the single, causal effect of diversification on firms' subsequent performance, neglecting the question of which firms diversify. The mixed empirical results from studies

[2]For a critical survey see Dosi et al. (2008); a detailed discussion follows however in section 2.

mentioned before leave us with an important question: Why not all firms engage in diversification activities or benefit equally from their diversification strategies? To answer this question, one should investigate the antecedent factors determining a firm's likelihood to diversify.

Some previous works attempt to identify the differences between diversifying firms and those specializing in a single-industry. Along these lines, among others, Maksimovic and Phillips (2002) and Gomes and Livdan (2004) argue that it is when the firms become relatively unproductive in their current businesses that they diversify. Matsusaka (2001) and Bernardo and Chowdhry (2002) state that firms choose to engage in diversification as a way to search for new opportunities to leverage their capabilities and resources. Similarly, Lang and Stulz (1994) also proposed that poor performers diversify in search of growth opportunities because they have exhausted such opportunities in their existing activities. Goldberg et al. (2010), with the same data on Indian firms used in this work, provide evidence that multi-product firms are significantly different from single-product firms - they are larger, more productive, and more likely to export than single-product firms. In short, these empirical studies show that the firms that decide to diversify exhibit significant inherent differences from those that do not diversify.

Identifying this selection bias, recent empirical contributions have explicitly incorporated in their works the endogeneity of the diversification decision. Recent works (Campa and Kedia, 2002; Villalonga, 2004a,b; Graham et al., 2002; Santarelli and Tran, 2013) permit the possibility that underlying differences between diversifiers and non-diversifiers might influence their performance, and employ empirical methods (Heckman, 1979; Dehejia and Wahba, 2002) that control for such selection bias. Some studies, such as, Campa and Kedia (2002) and Villalonga (2004b), indicate that diversification does not reduce firm value. In fact, when corrected for selection bias, the diversification discount disappears or even becomes a premium. In contrast, few other studies (Lamont and Polk, 2001; Ammann et al., 2012) provide evidence of a significant diversification discount even after controlling for the endogeneity of firms' decision to diversify. Recent empirical studies (Coad and Guenther, 2013), argue, in line with the present paper, that a firm has incentive to diversify when it has excess capabilities.

To tackle the reasons behind the disagreement of previous studies investigating the complex relationship between diversification and firm performance, will be the first objective of this paper. Our explanation will be based on the heterogeneity between diversifying firms. Indeed, both finance and strategy scholars have noted that prior firm performance (or, prior firm characteristics) influence the diversification decision of firms (Campa and Kedia, 2002; Graham et al., 2002; Park, 2003; Villalonga, 2004b; Miller, 2004; Santarelli and Tran, 2013). Therefore, a key ingredient of the exercise we perform in this work is the non-random firm's decision on their products' basket. While one might find that multi-product firms grow less than single-product firms, this could be due to an inverse casual relationship, i.e, firms decide to diversify their portfolio when their growth potential is reduced. In other words, there is pre-selection among the firms that decide to diversify. In order to correct for this non-random selection bias while investigating the relationship between diversification and performance, we use the method developed by Maddala (1986), the Endogenous Switching Model, which helps in estimating the effect with respect to their counter-factual stories, and separately assessing the effects of scope on firm performance.

Another prediction of the capability based approach is that firms diversify in fields related to their expertise, thereby exploiting the complementarities and synergies between production lines, due to the overlap between capabilities required for their production. Hence, while investigating the relationship between diversification and firm performance, it is important to analyze whether a firm diversifies in related or unrelated fields, relative to its core capabilities. Subsequently, the question of measurement arises in assessing what is a "related" field for a firm, given its existing field of expertise.

2.2 Relatedness between products, firms' coherence and their performance

The idea of coherence was introduced by Dosi et al. (1992). They pointed out that, in order for a firm to secure continuous accumulation of its capabilities, the main aim of the firm should be to achieve coherence between its existing and new activities. The first systematic empirical treatment of coherence was provided by the work of Teece et al. (1994). They measured coherence of a firm based on the inter-business relatedness of its outputs. They state that firms diversifying coherently enjoy economies of scope, as they share similar and complementary technical competencies and assets.[3] The presence of economies of scope is therefore a clear and falsifiable prediction of the capability theory of the firm.

The second objective of our paper is therefore to investigate whether firms benefit from entering into related business activities or not. For this, first and foremost, we need a measure of relatedness between products. In the following, we will discuss on the different ways to measure relatedness between products. We mainly consider two major methods: a categorical approach (Rumelt, 1974) and an evolutionary approach (Teece et al., 1994).

The measure developed by Rumelt is a categorical approach to determine relatedness. The procedure involves subjective classification of businesses into related and unrelated, using similarities in inputs, production technology, distribution channels etc, which is cited as the main drawback of this method (Markides and Williamson, 1996; Robins and Wiersema, 1995). Studies using Rumelt's method (Geringer et al., 1989; Dubofsky and Vandarajan, 1987; Christensen and Montgomery, 1981; Bettis, 1981) have broadly concluded that related diversifiers performed better than unrelated diversifiers.

The first evolutionary measure of coherence was introduced in Teece et al. (1994), which looked at inter-business relatedness to measure coherence. The authors define two industries as closely related if the observed firms are frequently active in those industries and coherent firms are defined as those which are active (or in other words, survivors) in a set of closely related sectors. Indeed, contrary to the measure by Rumelt (1974), such a method allows the data to speak, without necessarily assuming ex-ante which activities are to be considered as related. Teece et al. (1994) illustrate the diversification patterns of U.S. firms, and show that, as opposed to the idea that firms diversify at random, there is some coherence in the way firms diversify, i.e coherence (non-randomness) is a salient feature of firms' diversification patterns. They show that non-coherent diversifiers gradually disappear (or exit from the market) and surviving firms exhibit common patterns of diversification across industries.

Studies that used the measure proposed by Teece et al. (1994) include, among others, Piscitello (2000), Vonortas (1999) and Zuckerman (2000). They primarily focus on providing descriptive statistics of relatedness (Vonortas, 1999) and coherence (Piscitello, 2000), and the few cases that presented econometric exercises, make little reference to the theory of corporate coherence. Results of Piscitello (2000) show, consistently with the predictions, that relatedness of products was positively related to the firm's exploitation of complementary assets (i.e., R&D and advertising resources), and to its presence at different stages of the vertical chain.

Few other studies have applied the idea of Teece et al. (1994) to patents and industrial sectors. Breschi et al. (2003) use a similar perspective to explore the knowledge diversification of firms, by looking at the coherence of patent profiles. They investigate the determinants of corporate coherence and propose that relatedness in knowledge is a key factor affecting firms' technological diversification. Similar to Breschi et al. (2003), Nesta and Saviotti (2006) measure technological relatedness as co-occurrences of technology classes within patents and show that the coherence of the knowledge base within firms is a significant explanatory variable of the firms' stock market values. Another branch of studies have assessed industrial sectors in which the firm is active. Valvano and Vannoni (2003), using data on the top 5 firms in 95 3-digit Italian manufacturing industries, show that coherent-diversifiers are more probable to survive as top leaders, in contrast with the less coherent firms that have a higher probability to exit from the top 5 position. Piscitello (2004) adopts the Teece-approach and measures coherence using technological fields and output markets, where the market is defined as industrial sectors. He shows that only the firms that are able to exploit the synergies between their competencies and downstream activities adopt a coherent corporate strategy and are more likely to perform better. Bryce and Winter (2009) provide a similar metric to measure the relatedness between industrial sectors, aiming to inform managers and other stakeholders. Interestingly, in the paper their measure is tested for forecasting accuracy, and the good performance in this respect is considered a confirmation of the capability theory of the firm in general and of this empirical strategy in particular.

Bottazzi and Pirino (2010) have contributed to this literature with a technical improvement to the

[3] Some works emphasize managerial competencies, like 'core capabilities' (Leonard-Barton, 1992), 'dynamic capabilities' (Teece et al., 1997) and 'architectural competencies' (Henderson and Cockburn, 1994).

measure by Teece et al. (1994), by computing the statistical significance of the coherence between products with a stricter null model. Dosi et al. (2017b) uses this measure to analyze the relationship between different patterns of firm diversification, in particular, firm's knowledge base and production base. They find that, generally, firms are much more diversified in terms of products than in terms of technologies. However, their main products tend to be more related to their innovative knowledge.

To summarize, when considered as a whole, the empirical literature is in broad agreement with a capability-based view of diversification. The present work will try to give a unified empirical picture, tackling in particular two weak links of this literature. *First*, there is no explicit empirical consideration in integrating the firm's situation preceding the diversification event and the time of diversification with the future performance of the firm along all its trajectory. *Second*, as we discuss in detail in section 5.1, there are issues related to the proper definition of relatedness: often the measure used by previous studies has been created for a different purpose (for example, for the purpose of product classification).

3 Data: Firm and product level

This study employs firm-level data from the Prowess database, provided by the CMIE (Centre For Monitoring Indian Economy Pvt. Ltd.). The CMIE collects information from the annual financial reports of companies, which includes balance sheets and income statements. It covers both publicly listed and unlisted firms (one-third of the firms in the database are publicly listed firms). The companies covered account for around 70 percent of industrial output, 75 percent of corporate taxes, and more than 95 percent of excise taxes collected by the Government of India. In this work, we use the data on manufacturing firms over the period 1995 to 2015.

This database is a firm-level panel and the only Indian database, to our knowledge, that records detailed annual information on firms' product mix. Indian firms are required by the 1956 Companies Act to disclose product-level information on capacities, production and sales in their annual reports. The CMIE compiles these detailed data in time, which allows us to track changes in the product basket of firms. The product classification used by CMIE is similar to the Harmonised System classification. In this study, we use the products that are classified at 6 digit level, which in total are 383 unique products. For more details on the nested industry and product structure of Prowess, please refer to Dosi et al. (2017a).

3.1 Definition of the variables used

We will focus on two dimensions of performance, namely firm growth (in terms of sales), and relative profitability (the share of profits in sales of the firm with respect to the other firms in the sector). In particular,

$$SalesGrowth_{i,t} = \log Sales_{i,t} - \log Sales_{i,t-1}, \quad (1)$$

$$RelativeProfitability_{i,t} = \left(\frac{Profits_{i,t}}{Sales_{i,t}}\right) / \left\langle \frac{Profits_{j,t}}{Sales_{j,t}} \right\rangle_{j \in S_i} \quad (2)$$

where $\langle \cdot \rangle_{j \in S}$ is the average with respect to all the firms in the sector S, and S_i is the sector of main activity of firm i.

A key narrative element of our analysis is that diversifying firms are not a random sample of firms; they are different from non-diversifying firms. We will see this difference in multiple dimensions, defined by different firm-level variables, and we will use the same vector of variables X_{it} for all the analysis. Notice that, since in the econometric exercises we will exploit this heterogeneity to identify the effect of diversification, we want to use several dimensions to have a well defined counterfactual. We also want to use the same set of controls in all the analysis: some of these dimensions could be not significant for some relationships, but we want to avoid forcing our assumptions on the data. The variables that will be excluded to identify the causal direction (exclusion restriction) will be introduced separetely.

In particular, in all the analysis we will employ the following firm-level variables as controls in all the analysis: i) Firm size, defined in terms of total sales; ii) Age of the firm, defined in terms of number of years after incorporation; v) Export dummy to identify exporters, that takes value 1 if the firm exported and zero otherwise; vi) R&D dummy, to identify innovative firms, that takes value 1 if the firm spend on R&D and 0 otherwise; iii) Process complexity of the firm, proxied by the number of inputs used by the firm; iv) Distribution complexity of the firm, proxied by share of selling expenses to total sales; vii) Leverage and cash balance, to control for financial status; viii) Process innovation proxied by investment intensity, defined as additions to gross fixed assets over sales; and ix) Profitability growth (the shrinking or growth of the firm's margin of profit) to control for the growth momentum of the firm.

In our econometric analysis, we also control for the squared log of sales to control for non-linearities. Sector and time specific effects are captured through dummies, where sector dummies considered relate to the four Pavitt's taxa as defined in the next sub-section.

Finally, to define coherence of the product basket of the firm (detailed in section 5.1), we use the detailed product-level sales for each firm.

3.2 Definition of sectors

In this study, exercises are frequently performed at a disaggregated level, i.e, clustering firms separately for 2-digit industrial sectors. However, at times, we require larger sectoral aggregations for higher sample statistics. In these cases, we aggregate sectors according to the Pavitt Taxonomy (Pavitt, 1984):

- SECTOR 1 - Supplier dominated: Sectors include textiles, wearing apparel, leather, wood, paper, printing and basic metals;
- SECTOR 2 - Scale intensive: Sectors include food, beverages and tobacco, coke and petroleum, chemicals, rubber and plastics, other non-metallic minerals and fabricates metals;
- SECTOR 3 - Specialized suppliers: Sectors include electrical equipment, machinery and equipment, motor vehicles, transport except air spacecraft and military;
- SECTOR 4 - Science based: Sectors include computer and electronics, electrical equipment, machinery and equipment, motor vehicles and other transport.

4 Diversification and firm performance

In this section, we examine the relationship between scope of the firm and its performance. In line with the framework of under-utilized capabilities of firms leading to diversification (Teece, 1982), while there are clear expectations relating an increase in scope and the growth in firm sales, the relation between increase in scope and relative profitability is more ambiguous. Indeed, while one could argue that diversifying firms are moving away from their "core capabilities", it could also be argued that such multi-product firms enjoy economies of scope helping them to attain higher profitability, in contrast with the case in which the two activities are performed by two single product firms.

Since the effects of the scope of the firm are expected to be non linear, we will limit our analysis to the difference in performance between single-product firms and multi-product firms.

In line with the argument presented in the previous sections that diversifying firms are different from their counterparts, here we demonstrate that, in order to answer the question of whether firm scope affects its performance, it is necessary to examine the differences in firm characteristics between single and multi-product firms.

4.1 Growth rate and profitability of firms that diversify

A cursory look at descriptive statistics on some firm characteristics shows that single and multi-product firms are inherently different. Tables 1 and 2 present the main characteristics of the two classes of firms in different manufacturing sectors: the differences between single-product and multi-product firms are evident.

It is notable that, while some differences are consistent across all the sectors, for example the average size of multi-product firms is higher than the size of single-product firms in all sectors, some other differences are more nuanced and sector-specific. Shares of single and multi-product firms also vary across sectors, which is consistent with the observed differences in performance among sectors if one assumes that firm diversification is an endogenous process at least partly related to the firm's expectation of growth and profits. Indeed, this is consistent with the capability-based, technologically driven narrative: if the trade offs influencing the firm's decision to diversify their production are (also) due to the firm moving further away from their core capabilities, such trade offs can behave differently in different contexts and sectors.

Table 1: **Characteristics of single and multi product firms by sector of economic activity for the year 2000**

	No. of firms		Sales		Firm Growth		Profitability	
Sectors	Single	Multi	Single	Multi	Single	Multi	Single	Multi
Textiles	293	87	553.19	1192.57	0.017	0.031	0.037	0.030
Wearing Apparel	20	8	213.26	803.70	0.139	0.142	0.038	0.065
Leather and leather products	28	11	361.71	208.35	0.056	-0.200	0.025	0.019
Wood	12	5	213.09	644.87	0.146	0.007	0.029	0.049
Paper	122	25	796.12	811.51	0.140	0.101	0.063	0.057
Printing	6	1	228.23	31.10	0.241	.	0.089	0.000
Basic metals	101	107	284.87	14168.62	-0.034	0.035	0.035	0.023
Food Products	98	92	842.28	1031.60	0.056	0.068	0.029	0.066
Beverages	60	40	443.39	1850.21	0.022	0.090	0.038	0.055
Tobacco	6	1	3863.18	.	0.061	.	0.023	.
Chemicals	266	221	713.01	5865.85	0.043	0.060	0.047	0.036
Rubber and plastics	125	103	379.96	885.92	0.044	0.031	0.043	0.042
Other non-metallic minerals	128	54	1181.23	2221.26	0.068	0.027	0.078	0.059
Fabricated metals	188	175	779.28	1737.06	0.015	0.047	0.029	0.028
Electrical Equipment	94	126	968.01	1817.21	0.052	0.084	0.055	0.039
Machinery and equipment	118	196	418.11	1700.40	0.015	0.068	0.057	0.042
Motor vehicles	6	2	9114.65	7702.70	0.177	-0.136	0.006	0.023
Transport except Air spacecraft and military	15	12	450.23	434.03	0.176	0.047	0.090	0.047
Pharmaceuticals	203	171	504.45	2841.20	0.028	0.049	0.060	0.052
Computer and electronics	341	73	1075.57	939.04	0.106	0.034	0.076	0.046
Transport-Air spacecraft and military	7	3	8986.60	656.30	0.000	0.200	0.000	0.001

As single and multi-product firms differ in most dimensions, simply looking at descriptive statistics can be misleading. Therefore, in this section, we perform an econometric analysis to understand the differential performance of single and multi-product firms. We begin with a baseline OLS model:

$$Y_{it} = \beta X_{it} + \delta D_{it} + \eta_{it}, \tag{3}$$

where Y_{it} represents two dimensions of performance of firm i at time t as defined in the previous section, namely, sales growth and relative profitability; X_{it} is the vector of firm-level variables as defined in section 3.1; D_{it} is a dummy variable which takes the value 0 for single-product and 1 for multi-product firms; η_{it} are exogenous shocks. The independent variables are lagged by one year.

Estimating the model through OLS gives an estimate of δ, the difference in performance between single and multi-product firms. We also exploit the panel structure of our data to control for unobservable heterogeneity of firms and perform a fixed effects estimation of equation 3. Results are presented in table 3. Columns 1 and 2 report results from an OLS and fixed effects estimation of variables of interest on firm growth, while columns 3 and 4 report results on firm profitability.

First, we focus on the determinants of *sales growth* of the firm. Quite evidently, we observe that, our variable of interest, i.e. whether the firm is single or multi-product, does not have a clear and significant effect on firm growth.

Instead, we find some clear indication with respect to other variables in determining firm growth. Firm momentum, in terms of past profitability growth, has a significant and positive relation to firm growth. The

Table 2: **Characteristics of single and multi product firms by sector of economic activity for the year 2000 *(continued)***

	R&D		Export		Selling exp.		No. of inputs		Invest. Int.		Log leverage	
Sectors	Single	Multi	Single	Multi	Single	Multi	Single	Multi	Single	Multi	Single	Multi
Textiles	0.071	0.155	0.669	0.824	0.036	0.044	2.11	1.79	0.044	0.083	1.235	4.134
Wearing Apparel	0.000	0.000	0.800	1.000	0.062	0.080	1.38	2.64	0.097	0.000	0.618	0.781
Leather and leather products	0.000	0.000	0.762	0.800	0.063	0.045	1.00	1.04	0.071	0.001	0.692	0.801
Wood	0.000	0.333	0.500	1.000	0.100	0.094	1.62	3.90	0.361	0.000	0.269	0.337
Paper	0.085	0.176	0.265	0.625	0.028	0.037	1.27	1.15	0.126	0.135	0.545	0.455
Printing	0.000	0.000	1.000	0.000	0.044	0.035	3.17	1.20	0.010	0.006	0.332	0.248
Basic metals	0.033	0.117	0.467	0.552	0.032	0.029	1.39	1.67	0.049	0.059	0.504	0.588
Food Products	0.105	0.134	0.542	0.341	0.063	0.051	1.76	1.87	0.032	0.051	0.714	0.620
Beverages	0.047	0.407	0.258	0.312	0.133	0.063	1.01	2.03	0.040	0.061	0.445	0.375
Tobacco	0.500	0.000	0.667	0.000	0.061	0.000	1.14	6.34	0.013	0.000	0.156	0.000
Chemicals	0.201	0.301	0.545	0.650	0.057	0.051	1.97	1.81	0.066	0.242	0.573	0.573
Rubber and plastics	0.117	0.153	0.691	0.734	0.047	0.042	1.91	1.94	0.066	0.167	0.556	0.587
Other non-metallic minerals	0.198	0.190	0.600	0.579	0.087	0.108	2.09	1.69	0.060	0.043	0.736	0.725
Fabricated metals	0.098	0.107	0.545	0.620	0.025	0.026	1.53	1.60	0.090	0.047	0.703	0.536
Electrical Equipment	0.172	0.346	0.562	0.761	0.049	0.063	2.03	2.40	0.207	0.037	0.426	0.499
Machinery and equipment	0.176	0.432	0.677	0.824	0.049	0.053	2.12	2.20	0.057	0.031	0.777	0.487
Motor vehicles	0.667	0.000	1.000	0.500	0.054	0.043	1.20	1.57	0.042	0.031	0.308	0.402
Transport[4]	0.167	0.000	0.667	0.400	0.043	0.037	4.07	3.23	0.352	0.000	1.066	0.549
Pharmaceuticals	0.206	0.310	0.639	0.812	0.052	0.047	1.55	1.67	0.127	0.126	0.447	0.474
Computer and electronics	0.193	0.351	0.761	0.657	0.049	0.038	5.51	3.12	1.274	0.101	0.552	0.545
Transport-Air spacecraft and military	0.000	0.500	1.000	0.500	0.003	0.049	5.40	3.55	0.095	0.000	0.411	0.305

results concerning the size and age of the firm are in line with other studies, i.e, the smaller and younger firms exhibit high growth (even if in our sample age is not significant once we control for fixed effect). Further, we observe that fast growing firms have simple processes and high selling and distribution expenses. Innovative firms, here proxied by firms investing in R&D, and firms competing in the international market do not grow faster in India – as also reported in other studies on Indian manufacturing sectors (Mathew, 2017). The firm's financial status does not seem to matter significantly in explaining the sales growth of the firm.

Now, we move to the analysis of performance in terms of the *relative profitability* of the firm, which is its profitability with respect to other firms in the same sector. Columns 3 and 4 in table 3 report results of an OLS and fixed effects estimation. Most importantly, we observe that multi-product firms are significantly less profitable than their single-product peers when an OLS analysis is performed, while the results from employing a fixed effect estimation are not significant. Concerning other variables, while the momentum of the firm is still crucial, firms with higher relative profitability (with respect to the average profitability in the sector to which they belong) tend to be big and young, while having simple manufacturing processes and no in-house R&D. Financial leverage and lack of cash, though not very significant, seem to reduce the firm's profitability relatively to its peers.

Overall, the estimation is consistent with the simple narrative that firm performance is hindered, or at least not enhanced, when firms move away from their core capabilities. But what does the coefficient δ – i.e, the coefficient of the single-multi-product dummy – estimate? If the single or multi-product status of the firm were random and exogenous to the performance variables, this would be an unbiased estimate of the conditional estimate $E[Y_{1it}|X_{it}] - E[Y_{0it}|X_{it}]$, that is, the difference in expected performance for a random firm if the *same* firm were single or multi-product. However, as previously mentioned, this is not usually the case. While earlier studies on diversification interpreted the negative performance of firms after diversification ($\delta < 0$) as a loss of competitiveness due to moving afar from the core production of the firm (Markides, 1995; Lang and Stulz, 1994), recent studies have questioned this conclusion. A different, more recent, explanation is that struggling firms, or firms that have hit a barrier in further growth due to unobservable variables like competition, demand or technical issues, or simply satisficing aspirations of the firm led by its idiosyncratic routines, are more likely to diversify (Santarelli and Tran, 2013). The choice of

Table 3: **Firm scope and performance: OLS and Fixed Effects Regression**

Log Sales	-0.0614***	-0.3955***	0.0114***	0.0069*
	(0.0069)	(0.0152)	(0.0022)	(0.0040)
Squared Log of Sales	0.0014***	0.0075***	-0.0005***	-0.0004
	(0.0005)	(0.0010)	(0.0001)	(0.0003)
Log Age	-0.0399***	-0.0013	-0.0087***	-0.0312***
	(0.0031)	(0.0137)	(0.0009)	(0.0035)
Export Dummy	0.0028	-0.0079	0.0078***	-0.0009
	(0.0050)	(0.0069)	(0.0014)	(0.0016)
R&D Dummy	-0.0077	0.0022	-0.0051***	-0.0045**
	(0.0048)	(0.0071)	(0.0013)	(0.0018)
Log no. of inputs	-0.0149***	-0.0057	-0.0061***	-0.0040***
	(0.0028)	(0.0052)	(0.0008)	(0.0015)
Log of Selling Expenses	0.0406***	0.1138***	0.0009*	0.0003
	(0.0018)	(0.0032)	(0.0005)	(0.0008)
Log Leverage	0.0018	0.0061**	-0.0077***	-0.0009
	(0.0020)	(0.0030)	(0.0006)	(0.0007)
Log of Cash Balance	0.0004	0.0035**	-0.0012***	-0.0011**
	(0.0013)	(0.0017)	(0.0004)	(0.0004)
Investment Intensity	-0.0213***	-0.0479***	0.0079***	-0.0084***
	(0.0020)	(0.0031)	(0.0007)	(0.0010)
Profitability Growth	0.0544***	0.0538***	0.2653***	0.2194***
	(0.0099)	(0.0096)	(0.0117)	(0.0086)
Time & Sector Dummies	Yes	Yes	Yes	Yes
Single-multi dummy	0.0079*	0.0056	-0.0065***	-0.0034
	(0.0043)	(0.0080)	(0.0012)	(0.0022)
Observations	16346	16346	12868	12868
R^2	0.094	0.238	0.099	0.136
Number of firms	3205	3205	2430	2430

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Column I - OLS with firm growth,
Column II - Fixed effects with firm growth,
Column III - OLS with profitability,
Column IV - Fixed effects with profitability

the firm scope, D_{it}, therefore cannot be considered exogenous and random. This study will, therefore, more deeply investigate the characteristics of single and multi-product firms.

4.2 Which firms diversify?

In this section, we investigate the different characteristics of single and multi-product firms, starting with a Probit estimation:

$$D_{it} = \begin{cases} 1 & \text{if} \quad \theta X_{it} + \epsilon_{it} \geq 0, \\ 0 & \text{if} \quad \theta X_{it} + \epsilon_{it} < 0. \end{cases} \tag{4}$$

where we use D_{it}, the single or multi-product status of the firm, as dependent variable and the same vector of controls we employed before, X_{it}, as explanatory variables. The results are reported in table 4.

As is immediately clear, the choice of the firm's scope is neither random nor exogenous to the performance of the firm. Firms that decide to diversify are substantially different: they are bigger, older, have complex manufacturing processes and simple distribution processes. The presence of such big firms in the country – i.e. firms with the required management capabilities to expand their production – could be a key variable for development through the expansion of the export basket of the country. From an econometric point of view, the results in the previous section, on the determinants of firm performance, and here, on the characteristics of single and multi-product status of the firm, suggest that there is a substantial overlap between the characteristics that determine which firms diversify and the characteristics of low performing firms. This suggests that firm diversification cannot be treated as an exogenous variable.

In the next sub-section, we will re-investigate the relationship between diversification and firm performance. However, this time we treat the scope of the firm as an endogenous variable.

4.3 Non-random selection

A commonly used econometric technique to tackle non-randomness in the selection mechanism and to determine the unbiased effect of the non-exogenous choice is the Heckman's treatment effect. Below, we couple equations 3 and 4:

$$Y_{it} = \beta X_{it} + \delta^H D_{it} + \eta_{it}^H, \tag{5}$$

$$D_{it} = \begin{cases} 1 & \text{if} \quad \theta Z_{it} + \epsilon_{it} \geq 0, \\ 0 & \text{if} \quad \theta Z_{it} + \epsilon_{it} < 0. \end{cases} \tag{6}$$

where Z_{it} includes the same independent variables used in the main equations, and X_{it} includes (other than the set of variables used in the main equations) two further variables which satisfies the exclusion restriction. Following Campa and Kedia (2002), the exclusion restrictions we use here are i) the share of diversified firms (PNDIV) and ii) the share of their sales with respect to overall sales (PSDIV) in the main sector of economic activity of the firm. The right hand side variables are lagged by one year.

If the two error terms η_{it} and ϵ_{it} were independent, i.e. $E[\eta_{it}\epsilon_{it}] = 0$, the two equations could be estimated separately without any selection issue. Instead, if the two error terms are dependent, the simple estimation of the two equations does not give an unbiased estimation of δ^H, which is the expected effect of diversification decision by a random firm. An alternative way of looking at it, is as a problem of censored data: we are assuming $D_{it} = 1$ both for i) firms that could have opted for $D_{it} = 0$ with equal probability (i.e., if $\theta Z_{it} \approx 0$) and ii) firms that could not have acted in a different way (i.e. $\theta Z_{it} \gg 0$).

The Heckman treatment effect estimation tackles this issue, and can be executed in several ways. In the following analysis we will use a Full Information Maximum Likelihood estimation, estimating both equations at the same time with one likelihood function (Amemiya, 1985). The technique has a clear counterfactual interpretation. We can interpret equation 5 as unbiased for firms that are at the threshold, i.e. firms with $\theta Z_{it} = 0$ are assumed to randomly select their behavior between D_{it} equal 0 or 1. Instead, for firms that are far from the threshold, the estimated value of Y_{it} is corrected proportionally to i) their distance from

Table 4: **Determinants of firm scope: Which firms divesify?**

	(1) I	(2) II	(3) III
Log Sales	0.3610***	0.2821***	0.1587***
	(0.0313)	(0.1031)	(0.0087)
Squared log of sales	-0.0030	0.0315***	-0.0048***
	(0.0022)	(0.0073)	(0.0005)
Log Age	0.2080***	0.6906***	0.0617***
	(0.0171)	(0.0760)	(0.0052)
Export Dummy	0.0305	-0.1190	0.0192**
	(0.0265)	(0.0957)	(0.0092)
R&D Dummy	0.0273	-0.3156***	0.0336***
	(0.0267)	(0.0955)	(0.0086)
Log no. of inputs	0.2229***	0.8832***	0.0608***
	(0.0156)	(0.0606)	(0.0045)
Log of Selling Expenses	-0.0851***	-0.2076***	-0.0247***
	(0.0099)	(0.0352)	(0.0029)
Log Leverage	0.0608***	0.1242***	0.0217***
	(0.0110)	(0.0328)	(0.0037)
Log of Cash Balance	0.0085	0.1210***	-0.0005
	(0.0072)	(0.0214)	(0.0022)
Investment Intensity	-0.0037	0.0012	-0.0080***
	(0.0115)	(0.0394)	(0.0031)
Profitability Growth	0.0838	-0.0832	0.0133
	(0.0743)	(0.1668)	(0.0099)
Time & Sector Dummies	Yes	Yes	Yes
Observations	18320	18320	18320
R^2			0.204
Pseudo R^2	0.189		
Number of firms	3505	3505	3505

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column I - Probit, Column II - Probit random effects, Column III - Linear Probability the random threshold – measured by the inverse Mill's ratio – and ii) the correlation between ϵ_{it} and η_{it}. The technique gives rise to a value of δ^H as an unbiased estimate of $E[Y_{1it}|X_{it}] - E[Y_{0it}|X_{it}]$, the estimated counterfactual value of choosing $D_{it} = 1$ over $D_{it} = 0$ for a random firm in the sample. Several studies have used this technique to estimate gains from diversification, as we reviewed in section 2, usually redeeming the value of diversification for firms (e.g., Brendel et al. 2015). The results from a Heckman treatment effect estimation are presented in table 5.

The estimate indicates that, when controlling for non-random selection, results are very different. Columns 1 and 2 in table 5 report results from the performance equation (eq. 5) and selection equation (eq. 6) with firm growth as the dependent variable, and columns 3 and 4 report results from performance and selection equations for profitability as the dependent variable, respectively. Quite strikingly, we observe that the effect of expanding the firm's scope on firm growth is now positive and significant: firms deciding to expand their scope grow 6.4% more than firms that, in similar conditions, decide to stay focused. The effect on the relative profitability is still negative.

However, the target of our research question is not to understand the effect of an increase in scope of a *random firm*. If different firms opt for different strategies, we should expect that their returns for such strategies would be different. The average effect of diversification of a *random firm* could well be not significant, whereas the average effect of a firm strategy (here diversification) could be positive for the firms selecting that strategy, and negative for others. A difference in the expected returns of the possible strategic choices among firms choosing those strategies is not only possible, but theoretically expected. Once we accept that the firm's choice to diversify is not exogenous, it is not reasonable to assume that the expected return from diversification on firm performance would be the same for any random firm. Instead, we would expect that the effect of diversification on firm performance is different for firms that actually diversified and for those that did not.

We will therefore apply Endogenous Switching Regression (Maddala, 1986), which features a selection equation to determine which firms diversify, and 2 equations that separately estimate the relationship between diversification and performance for single and multi-product firms. In practice, all three equations are estimated at the same time:

$$Y_{0it} = \beta X_{0it} + \eta^E_{0it}, \tag{7}$$

$$Y_{1it} = \beta X_{1it} + \eta^E_{1it}, \tag{8}$$

$$D_{it} = \begin{cases} 1 & \text{if} \quad \theta Z_{it} + \epsilon_{it} \geq 0, \\ 0 & \text{if} \quad \theta Z_{it} + \epsilon_{it} < 0. \end{cases} \tag{9}$$

here Y_{0it} (Y_{1it}) and X_{0it} (X_{1it}) are the performance variables and independent variables for the single (multi-product) firms respectively. The equations are estimated simultaneously with a Full Information Maximum Likelihood method. In a similar fashion to the Heckman's treatment effect, the correction for censoring is both proportional to i) the distance of a firm from the threshold $\theta Z_{it} = 0$ (Inverse Mill's ratio), and ii) the correlation between the error terms.

The effect of the multi-product status of the firm in this case is estimated separately after the regression, by looking at the counterfactual case in which a multi-product firm is estimated as a single-product firm, and vice versa. This allows separate estimates of two effects, which are: i) the average change in performance of a multi-product firm with respect to the hypothetical case in which it had been a single-product firm (Average Treatment Effect on the Treated, ATT); ii) the performance of a single-product firm with respect to the hypothetical case in which it had expanded its scope to multi-product (Average Treatment Effect on the Untreated, ATU). The results of the estimation are reported in table 6.

The first three columns report results with firm growth as the performance variable, while the last three columns present results with relative profitability. In each case, the first column shows the estimation results for single product firms, while the second shows that of multi-product firms, and the third shows the results from a selection equation. The results show that the effects of increasing the scope of the firm on its performance are very different according to the context in which the choice is taken: the effects on the

Table 5: **Firm scope and performance: Heckman Treatment**

	Firms' Growth		Relative Profitability	
	Main	Selection	Main	Selection
Log Sales	-0.072***	0.310***	0.043***	0.034
	(9.01)	(5.95)	(27.23)	(0.98)
Squared Log of Sales	0.002***	-0.003	-0.002***	0.017***
	(3.77)	(0.79)	(19.58)	(6.57)
Log Age	-0.044***	0.211***	-0.004***	0.215***
	(12.93)	(11.60)	(4.25)	(11.27)
Export Dummy	0.002	0.019	0.007***	0.024
	(0.36)	(0.69)	(5.03)	(0.83)
R&D Dummy	-0.009*	0.030	-0.005***	0.028
	(1.91)	(1.06)	(3.68)	(0.98)
Log no. of inputs	-0.019***	0.248***	-0.002*	0.239***
	(5.99)	(14.66)	(1.88)	(12.95)
Log of Selling Expenses	0.042***	-0.050***	-0.003***	-0.041***
	(22.79)	(4.72)	(5.87)	(3.54)
Log Leverage	0.001	0.057***	-0.008***	0.059***
	(0.35)	(4.85)	(14.28)	(4.82)
Log of Cash Balance	0.000	0.011	-0.001**	0.009
	(0.35)	(1.39)	(2.39)	(1.15)
Investment Intensity	-0.021***	0.027**	0.011***	0.024*
	(10.38)	(2.07)	(16.78)	(1.65)
Profitability Growth	0.054***	0.143	0.238***	0.295
	(5.41)	(1.64)	(19.96)	(1.15)
Single-multi dummy	0.064***		-0.047***	
	(3.04)		(6.82)	
PNDIV		0.869***		0.386*
		(4.09)		(1.71)
PSDIV		0.964***		1.106***
		(6.51)		(7.32)
Time & Sector Dummies	Yes	Yes	Yes	Yes
Observations	16346	16346	12868	12868
Number of firms	3205	3205	2430	2430

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$
Column I - Performance equation with firm growth,
Column II - Selection equation (firm growth)
Column III - Performance equation with firm profitability,
Column IV - Selection equation (firm profitability).

Table 6: **Firm's performances and firm's scope: Endogeneous Switching Regression**

	Firm Growth			Relative Profitability		
	Single	Multi	Selection	Single	Multi	Selection
Log Sales	-0.116***	-0.050***	0.352***	-0.002	0.027***	0.164***
	(7.21)	(5.20)	(6.81)	(0.79)	(13.14)	(6.14)
Squared Log of Sales	0.002**	0.001*	-0.005	-0.001***	-0.001***	0.001
	(2.03)	(1.79)	(1.47)	(4.20)	(4.43)	(0.69)
Log Age	-0.072***	-0.035***	0.217***	-0.017***	0.007***	0.141***
	(12.27)	(8.79)	(11.99)	(11.13)	(5.95)	(9.43)
Export Dummy	0.010	-0.003	0.029	0.004*	0.003*	-0.005
	(1.19)	(0.42)	(1.03)	(1.86)	(1.83)	(0.21)
R&D Dummy	0.012	-0.017***	0.026	-0.003	-0.001	0.008
	(1.28)	(2.87)	(0.93)	(1.38)	(0.41)	(0.34)
Log no. of inputs	-0.043***	-0.012***	0.250***	-0.012***	0.007***	0.139***
	(7.75)	(3.43)	(14.82)	(8.26)	(6.15)	(9.61)
Log of Selling Expenses	0.054***	0.038***	-0.053***	0.002**	-0.005***	-0.036***
	(16.29)	(17.32)	(4.93)	(2.18)	(7.56)	(4.04)
Log Leverage	0.007*	-0.004	0.054***	-0.001	-0.001*	-0.001
	(1.73)	(1.49)	(4.60)	(0.80)	(1.70)	(0.12)
Log of Cash Balance	0.004*	-0.002	0.010	-0.001	-0.000	0.002
	(1.71)	(1.39)	(1.39)	(1.31)	(0.55)	(0.39)
Investment Intensity	-0.046***	-0.011***	0.027**	-0.004***	0.002*	0.030**
	(11.02)	(4.93)	(2.07)	(3.11)	(1.91)	(2.57)
Profitability Growth	0.251***	0.033***	0.084	-0.046**	0.129***	0.684***
	(6.72)	(3.31)	(0.97)	(2.20)	(7.96)	(3.32)
PNDIV			0.692***			-0.128**
			(3.37)			(1.99)
PSDIV			0.997***			0.221***
			(7.03)			(4.98)
Time & Sector Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Total Observations	16346	16346	16346	12868	12868	12868
Number of firms	3205	3205	3205	2430	2430	2430
ATT		0.267***			0.166***	
		(2.75)			(6.94)	
ATU	0.020			-0.142***		
	(0.38)			(8.56)		

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Column I - Firm growth equation for single product firms,
Column II - Firm growth equation for multi product firms,
Column III - Selection equation (firm growth),
Column IV - Firm profitability equation for single product firms,
Column V - Firm profitability equation for multi product firms,
Column VI - Selection equation (firm profitability).

performance of a random firm are not informative. Instead, separately estimating the effects of diversification on firms that actually diversified and those that did not provides some interesting observations. In the case in which a firm actually did not diversify, if it had increased its scope, it would not have gained much in terms of firm growth and would have greatly decreased its profitability. However, firms that decided to increase their scope achieved both higher firm growth (in average around 27% more), and higher profit margin (around 17% more).

The results show a coherent narrative, but not a trivial one. Firms' dynamics is a complex process and the firm's choices have non-linear, context-dependent effects. Firms in good shape (i.e, doing well in terms of growth and profit margins) would reduce their profit margin by moving far from their core capabilities, without gaining much on the growth of their sales. If there are untapped sources of growth in existing markets (in other words, existing product basket), spreading their efforts in diversifying in more (and different) directions gives less additional growth. However, once those sources are exhausted and the firm is left with excess resources, not only does the firm grow less but its profit margins also reduce. At that stage, expanding the scope of the firm allows the firm to grow further. Profit margins may also increase through a more efficient use of the firm's resources and capabilities.

In the next sub-section, we investigate at a sectoral level the effect of diversification on firm performance.

4.4 Diversification and firm performance: Sector-wise analysis

As was shown in section 2, staying in line with the capability-based approach to firms dynamics, one would expect that sector specificities play a major role.

In table 7 we report the results of the endogenous switching regression – the average treatment effect on the untreated and treated, *ATU* and *ATT* – for firm growth and profitability, for each of the four Pavitt's taxa.

Table 7: **Firm's performances and scope: Sectorwise analysis**

	I		II		III		IV	
	ATT	ATU	ATT	ATU	ATT	ATU	ATT	ATU
Firm Growth	0.158**	0.085	0.188***	0.000	0.275***	-0.070	0.029	0.012
	(2.289)	(1.240)	(4.857)	(0.000)	(2.775)	(0.778)	(0.586)	(0.249)
Rel. Profitability	0.037	0.013	0.162***	-0.133***	0.146***	-0.127***	0.230***	-0.153***
	(1.890)	(0.689)	(7.429)	(5.845)	(4.294)	(3.735)	(4.035)	(4.811)

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
I - Supplier dominated, II - Scale intensive, III - Specialized suppliers, IV - Science based

First of all, the results we observe are consistent with minimal expectations of rational behavior. For all the Pavitt sectors and in both performance measures, *ATT* is higher than *ATU*, i.e. firms increasing their scope are the ones gaining more out of it, both in terms of firm growth and relative profitability. In addition, in all the sectors, the firms that decided to increase their scope would not have been better off in terms of firm growth or profitability if they decided to stay as single-product. Indeed, in all cases the value of *ATT* was found to be either positive or not significant. Similarly, in all sectors the value of *ATU* was found to be either negative or not significant. This means that the typical routines and decision making of firms in all sectors are not, in average, self-harming.

However, overall the results are extremely heterogeneous among sectors and measures of performance. Firm growth is not fostered by an increase in scope for firms in science based sectors, while these firms are more rewarded in terms of relative profitability. Conversely, firms in more traditional Supplier-dominated sectors do not have any advantage of diversification on profitability, but they are rewarded in terms of firm growth. This sectoral difference could be better understood by looking at the interaction between the diversification process of the firm with respect to production and with respect to its knowledge base (as in Dosi et al. (2017b)), and the heterogeneous importance of the firm knowledge base in different sectors.

Finally, we point out again that, even in the sectors were profit margins are not directly affected by diversification, profit margins may increase through a more efficient use of resources and capabilities of firms. As pointed out by Dosi (1988), since organizational capabilities are cumulative in nature and firms are better able to carry out tasks with which they already have some skills, an efficient use of a firm's existing resources and capabilities would imply developing products that are related to their existing product basket. In the next section, we detail how we capture this idea.

5 Coherence of the product basket of the firm

Until now, the study focused on investigating the relationship between diversification of firms into new markets and their performance. As previously stated, when a firm diversifies, expanding to potential new business areas that are related to its existing business can largely build on the firm's current experience, knowledge and its transferability. When most of the capabilities of a firm are not transferable, and given the specificity of such intangible assets, these capabilities become most beneficial when being used to diversify into related activities. In this section, we proceed to study the effects of related diversification on firm performance, by emphasizing the synergies of a product line with respect to the overall product basket of the firm.

We start by defining a measure of relatedness between products, which we later use to define a measure of coherence of the product basket of the firm. Further, we will investigate the relationship between coherence of the product basket of the firm and its performance, which complements the narrative from previous sections of this paper.

5.1 A simple measure of relatedness between products

Before moving to a *quantitative* definition, we begin by clarifying the kind of coherence we are measuring. First of all, we are not interested in standard classifications of products. Hence, measures like Rumelt (1974) or classifications like Harmonised System would not be useful. Other than the issues already raised in the introduction about the subjective value of Rumelt (1974) measure, that family of measures also heavily relies on an external classification of products. However, any already existing classification has as a main objective to position the products in a hierarchical structure, and the distance between products is inferred only as a by-product. However, the kind of spillovers and common capabilities that defines two products as related could go deeper than skin deep, and therefore ignored in the effort of finding the proper position for each product. Attempts to classify the products in multiple dimensions show that several dimensions are needed (for instance, Pehrsson (2006) proposes at least five dimensions) and that a hierarchical representation is not a good candidate to represent relatedness and coherence.

In line with Teece et al. (1994), we define "relatedness" as the emergent property of an evolutionary process. We define two products as related if the probability of having them produced by the same firm is high. However, a further qualitative clarification is that we do not want to measure the statistical significance of the relatedness between two products, but the relatedness itself. Therefore we will not use, as a measure of relatedness, the statistical significance of the observed coherence (*i.e, the probability of observing that number of concurrences in the same firm between the two products assuming a random coupling between firms and products*), as in Teece et al. (1994) or Bottazzi and Pirino (2010), but the observed relatedness itself (*how often the two products concur in the same firm*). While the former is the proper measure if one wants to determine which concurrences among products are statistically significant, and therefore likely not random, it is not the proper measure of the strength of the coherence.

As an example, let us imagine that we have two coins and we want to see which one of them is more likely to show heads. Let us imagine that one coin has been flipped twenty times obtaining twenty heads, while the second has been flipped one thousand times and has obtained six hundred heads. If we imagine that both coins are fair, the first sequence had one in one million chances to happen by chance. The second sequence however had less than one chance in ten billion to happen by chance with a fair coin: the second experiment is ten thousands times more significant with respect to the falsification of the null hypothesis! If we were interested in the question "which coin is less likely to be fair?", the second coin is by far the better candidate. This is the kind of question that Teece et al. (1994) and Bottazzi and Pirino (2010) were interested in answering, but it is not the research question of this paper. What we want to know is which of the two coins you would flip if you have to bet on heads. While the second is less likely to be biased, the first is more likely to be *more* biased than the second. The null hypothesis that they are "as biased", against the alternative that the first coin is more biased, can be rejected with a significance of almost one in ten thousands.

In our case, it is not a trivial argument, because the size of the sample – the number of flips – is not without consequence: it is the number of firms producing those products. The products that are significantly more related to others are also those with more competition in their sectors. If we are looking at the effect of relatedness on performance by looking at the *significance* of their relatedness, we mix the effect of relatedness and the effect of producing in a market with a higher number of firms.

To measure the relatedness between products i and j we will therefore use the simple conditional probability of product i appearing in the product basket of the firm given the fact that the firm produces j. Given that the goal is to find a symmetric measure and the conditional probability of the firm producing i given j or j given i are not equal, we take the smaller of these. Therefore,

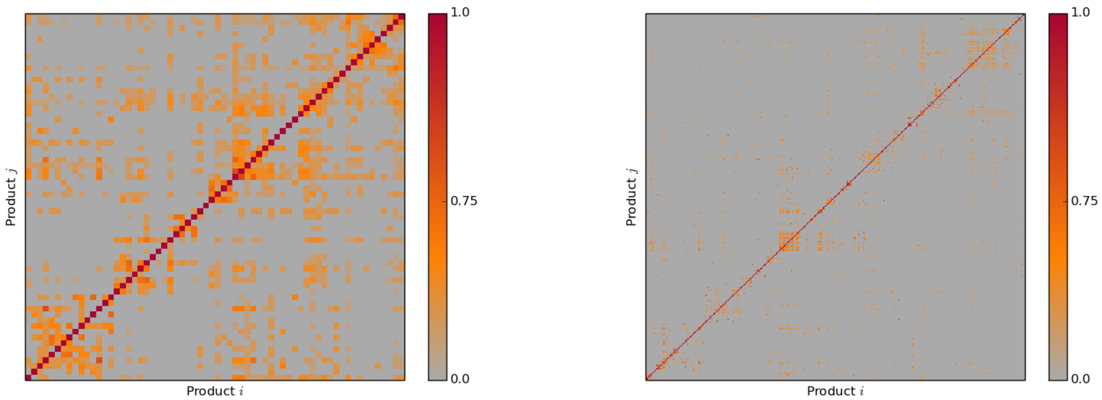
$$B_{i,j}(t) = \frac{1}{\max(u_i(t), u_j(t))} \sum_f M_{fi}(t) M_{fj}(t); \tag{10}$$

where $M_{fi}(t)$ is equal to 1 if the firm f produced the product i in the year t, and 0 otherwise; $u_i(t)$ is the ubiquity of product i in the year t, that is the total number of firms producing it: $u_i(t) = \sum_f M_{fi}(t)$. This probability defines a measure analogous to the one defined in Hidalgo et al. (2007) (and similar to Zaccaria et al. 2014) to connect countries and products. However, since the network exploited to build the matrix B is the network between firms, we are highlighting here Firm-level Capabilities instead of Country-level Capabilities. The relationship between the two is by itself an interesting research question; a priori it is possible to imagine two products requiring similar country-level capabilities but completely different firm-level ones, i.e, products often produced in the same countries but different firms.

Now, we look at how the measure introduced behaves qualitatively. The 6-digit product classification that will be used in the regressions consists of 383 different products, and therefore, $383^2 \approx 150,000$. For graphical purposes however, in figures 1a and 2 we use products defined with the 4-digit product classification. In this way, we have only 64 products and the figures are clearer and more informative.

Let's start by looking at figure 1a. Both rows and columns represent different products, ordered according to the hierarchical classification. In each cell, the color represents the relatedness between the product in the row and the product in the column. A darker color indicate stronger relatedness between products. If as a measure of relatedness we were to consider a measure based on the hierarchical categorization, i.e. two products are related if their first digits are equal, we would have found a block diagonal matrix: only nearby products would be related. As is visible, the hierarchical categorization is informative, as diagonal blocks are somehow visible, but there is far more information in the whole matrix: some products in the same category are not related, while some products in different categories are related. The equivalent matrix for products at 6-digit level is presented in figure 1b.

The same information can be recovered from figure 2. Each node is a 4-digit product, colored according to its 2-digit classification. Each product is connected with an edge to the product that is maximally related to it. It is evident that, while products in the same aggregated hierarchical category tend to be in the same cluster, we also observe many logical but non-trivial exceptions. A good example is product 0604, "Textiles based on vegetable fibers other than cotton", which is classified under textiles (according to the hierarchical classification), but we see that it is more related to the paper and pulp cluster than to the other textiles.



(a) 4 digits

(b) 6 digits

Figure 1: Relatedness $B_{i,j}$ between product i and j, defined at 4-digit product classification, with 64 different products (figure 1a) and 6-digit product classification, with 383 different products (figure 1b) for year 2010. The products are ordered on the axis accordingly to their Prowess hierarchical code. The diagonal $B_{i,i}$ has relatedness 1.

5.2 Derived measure: coherence of firm product basket

In this paper we are not interested directly in the relatedness between products. Rather, we are interested in assessing the average coherence of the whole portfolio of a firm f, C_f. We define it as the average $B_{i,j}$ between each pair of products of the firm, weighted with the sales of both products. To have a measure consistent also for firms with a different distribution of sales between products, even the relatedness of a product with itself has to be considered. Therefore,

$$C_f(t) = \sum_i \sum_j Q_{f,i}(t) Q_{f,j}(t) B_{i,j}(t) / \sum_i \sum_j Q_{f,i}(t) Q_{f,j}(t), \tag{11}$$

where $Q_{f,i}$ is the value of good i sold by firm f. The measure C_f is defined in such a way that if the firm is producing two products i and j which are perfectly coherent between them ($B_{i,j} = 1$), the product basket of the firm is as coherent as in the case in which the firm is single-product. On the contrary, if the firm f is producing a set of products that is completely incoherent ($B_{i,j} = 0, \forall i, j \in f$), C_f is equal to $1/D_f$, where D_f is the diversification of the firm, that is, the total number of products. In the next section, we use this measure, $C_f(t)$, to study the relation between the coherence of a firm's portfolio and its performance.

5.3 Coherence and Firm Performance

In this section, we use the measure of coherence that we built in previous sections to investigate how coherent diversification affects the performance of the firm. Note that, in the estimations that follow, we limit our sample only to multi-product firms. We present the simple linear form:

$$Y_{it} = \beta X_{it} + \gamma S_{it} + \lambda C_{it} + \upsilon_{it}, \tag{12}$$

where, as before, Y_{it} is the performance variable and X_{it} is the set of independent variables.[5] S_{it} is the number of different products produced by firms, C_{it} is the coherence defined as in equation 11 and υ_{it} represents the idiosyncratic shock term. We use a fixed effects estimation to estimate equation 12 and the results are reported in table 8.

[5]The vector of variables X_{it} is the same as in previous exercises and defined in section 3.1.

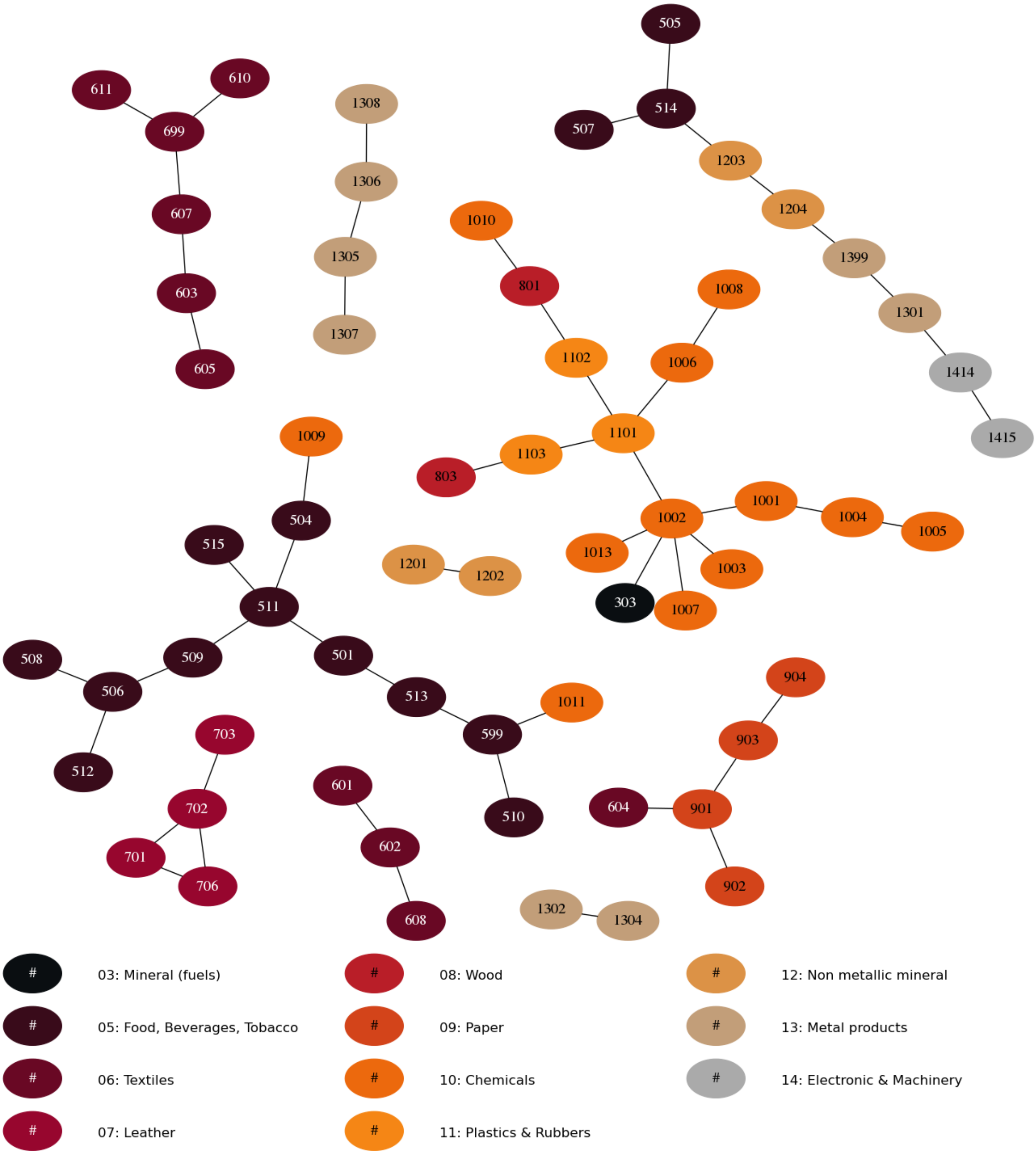


Figure 2: Network of products, defined at 4 digits product classification (64 different products). Two products i and j are connected if either j is the product more related to i (other than i itself) or i is the product more related to j (other than j itself), for the year 2010.

Table 8: **Coherence and firm performance: all manufacturing.**

	(1) I	(2) II
Log Sales	-0.4243***	0.0075
	(0.0251)	(0.0061)
Squared Log of Sales	0.0073***	-0.0003
	(0.0016)	(0.0004)
Log Age	0.0414**	-0.0197***
	(0.0209)	(0.0051)
Investment Intensity	0.0183***	0.0001
	(0.0021)	(0.0005)
Export Dummy	0.0058	-0.0022
	(0.0098)	(0.0023)
R&D Dummy	-0.0147	-0.0026
	(0.0101)	(0.0024)
Log of Selling Expenses	0.1225***	0.0021*
	(0.0046)	(0.0011)
Log Leverage	-0.0068*	-0.0015
	(0.0041)	(0.0010)
Cash Balance	-0.0013	-0.0012**
	(0.0024)	(0.0006)
Profitability Growth	0.6033***	0.2513***
	(0.0508)	(0.0119)
Log number of products	0.0052	-0.0007
	(0.0043)	(0.0011)
Log number of inputs	0.0028	-0.0068***
	(0.0086)	(0.0021)
Log Coherence	-0.0002	0.0124***
	(0.0193)	(0.0046)
Time & Sector Dummies	Yes	Yes
Observations	7286	6742
R^2	0.289	0.138
Number of firms	1468	1397

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Column I - Fixed effects with firm growth, Column II - Fixed effects with profitability

We observe that the coherence of the firms' product portfolios does not affect firm growth, which is consistent with the idea that opportunities are more related to the scope of the firm and not to its coherence. However it does significantly affect the relative profitability of the firm, which is consistent with the idea that related products use similar capabilities and a coherent product basket allows the firm to increase the efficiency of its resources and reduce costs. When a firm decides to diversify its production, the choice of a product that is related to its previous basket allows the firm to best utilize its excess resources and capabilities, hence increasing opportunities without losing efficiency.

Similar to the results concerning diversification, we also expect different sectors to behave differently with respect to the relationship between coherence and firm performance. Therefore, we investigate the relationship between coherence and firm profitability (equation 12) for different Pavitt sectors (reported in table 9). We observe that the effects of coherence on relative profitability vary among the four Pavitt sectors. While the effect of the coherence of the firm's product basket is positive for supplier dominated and science based sectors, it is not significant among firms in scale intensive and specialized suppliers sectors.

An interesting observation from table 8 is that the number of different inputs used in production is an important driver of profitability, i.e. using a wider variety of inputs reduces the relative profitability of the firm, even when controlling for the relatedness of the firm product basket. Here, the number of inputs acts as a proxy for complexity of production routines, with a higher number of inputs indicating more complex production processes.

6 Conclusions

In this work, we present the empirical evidence related to the diversification choices of firm and its performance. We looked at two crucial aspects of this relationship. *First*, we estimated the effects of the choice of the firm to diversify on its performance. Critically, and in contrast with most of the previous empirical literature, we did not assume that the effects of this choice would have been the same for different firms. Instead, we re-examine this relationship by considering the potential differences between firms that diversify and not, and the potential differential effect of diversification on both kinds of firms.

Second, we introduce a measure of relatedness between products that is able to capture the overlap between the capabilities required to produce different products. We further use this measure to investigate the relation between coherent diversification and firm performance.

We observe that both the scope and the coherence of the firm affect its performance, but the two aspects interact in an intuitive but non-trivial way, consistent with the predictions of the Penrose-Teece capability-driven theoretical framework of firm dynamics. In this framework, firm diversification is a strategy to tap the unexploited opportunities arising from the intangible and non-tradable resources the firm might have if the firm cannot further expand in the markets in which it is already active in. Hence, one could expect that diversification choice of the firm is not a random choice and that the firm decides to diversify when the growth opportunities in its existing markets are exhausted. A coherent diversification would allow the firm to use its excess capabilities at best, since such capabilities are not product independent. Being this excess resources mostly intangible, as stated in Teece et al. (1994), a further theoretical hypothesis is that coherence would be more important in knowledge-intensive sectors.

The results we present in this work are indeed in line with the predictions we stated above. Our findings show that: i) diversifying firms are not a random set of firms, but they are big and mature firms, often investing in R&D and in the sale process; ii) firms with increased scope attain higher sales growth than they would have achieved if they were single-product firms; iii) non-diversifying (single-product) firms are such because they lack the need or the capabilities to diversify: diversification would not grant them additional growth; iv) the effects of an increase in firm scope to firm profitability are mixed: they are sector and context-specific; v) the coherence of the product basket of the firm does not affect firm growth, however a coherent basket directly increase profitability, in particular in specific sectors.

This dynamics has non trivial implications on the industrial policy of a country aiming at diversifying its production. For this purpose, big and mature firms, with both the required organizational capabilities to

Table 9: **Coherence and firm profitability: sectorwise analysis.**

	(1) I	(2) II	(3) III	(4) IV
Log Sales	0.0306	-0.0047	0.0613***	-0.0492
	(0.0790)	(0.0265)	(0.0192)	(0.0513)
Squared Log of Sales	-0.0034	-0.0013	-0.0013	0.0012
	(0.0052)	(0.0016)	(0.0012)	(0.0032)
Log Age	0.0133	0.0017	-0.0728***	-0.0266
	(0.0642)	(0.0234)	(0.0200)	(0.0566)
Investment Intensity	-0.0043	0.0022	0.0031*	0.0137***
	(0.0058)	(0.0021)	(0.0019)	(0.0049)
Export Dummy	0.0203	-0.0156	-0.0241	0.000
	(0.1352)	(0.0543)	(0.0483)	(0.000)
R&D Dummy	0.000	0.0464	0.000	0.0349
	(0.000)	(0.0877)	(0.000)	(0.0837)
Log of Selling Expenses	-0.0219	0.0063	-0.0186***	-0.0103
	(0.0142)	(0.0045)	(0.0042)	(0.0109)
Log Leverage	-0.2827***	-0.3249***	-0.3448***	-0.1859***
	(0.0330)	(0.0112)	(0.0155)	(0.0415)
Log Cash Balance	0.0404***	0.0241***	0.0073***	0.0343***
	(0.0075)	(0.0030)	(0.0024)	(0.0070)
Profitability Growth	0.0083*	0.0074***	0.0028*	0.0114***
	(0.0044)	(0.0017)	(0.0015)	(0.0037)
Log number of products	-0.0335	-0.0270**	0.0062	0.0286
	(0.0350)	(0.0128)	(0.0088)	(0.0271)
Log number of inputs	-0.0176	-0.0279**	-0.0381***	-0.0003
	(0.0269)	(0.0109)	(0.0097)	(0.0178)
Log Coherence	0.0828**	0.0200	-0.0014	0.0691**
	(0.0420)	(0.0153)	(0.0113)	(0.0337)
Time&Sector dummies	Yes	Yes	Yes	Yes
Observations	1105	3638	1641	1588
R^2	0.200	0.301	0.377	0.093
Number of firms	341	806	359	350

Standard errors in parentheses

* ($p < 0.10$), ** ($p < 0.05$), *** ($p < 0.01$)

I - Supplier dominated, II - Scale intensive, III - Specialized suppliers, IV - Science based diversify and tapped opportunities in their own markets, are a necessary condition as much as country-level capabilities. This is not dissimilar from the diversification process in a biological evolving system, which happens when a mature species finds constraints to its own exponential growth due to the interaction with the environment. When we move farther from the simplistic representative firm modelization of macro economies, we see that the internal dynamics of firms is a central element to understand the development of countries.

References

Amemiya, T. (1985): *Advanced econometrics*, Cambridge, M A: Harvard university press.

Ammann, M., D. Hoechle, and M. Schmid (2012): "Is there really no conglomerate discount?" *Journal of Business Finance & Accounting*, 39, 264–288.

Barney, J. B. (1997): *Gaining and sustaining competitive advantage*, Addison-Wesley Reading, MA.

Berger, P. G. and E. Ofek (1995): "Diversification's effect on firm value," *Journal of Financial Economics*, 37, 39–65.

Bernardo, A. E. and B. Chowdhry (2002): "Resources, real options, and corporate strategy," *Journal of Financial Economics*, 63, 211–234.

Bettis, R. A. (1981): "Performance differences in related and unrelated diversified firms," *Strategic Management Journal*, 2, 379–393.

Bottazzi, G. and D. Pirino (2010): "Measuring Industry Relatedness and Corporate Coherence," LEM Working Paper, 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.

Brendel, M., A. D. Lahmann, and B. Schwetzler (2015): "About Estimating Gains from Diversification and Why Firms Self-Select," *Available at SSRN 2581456*.

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.

Bryce, D. and S. Winter (2009): "A General Interindustry Relatedness Index," *Management Science*, 55, 1570–1585.

Campa, J. M. and S. Kedia (2002): "Explaining the diversification discount," *The Journal of Finance*, 57, 1731–1762.

Castaldi, C., M. Cimoli, N. Correa, and G. Dosi (2009): "Technological Learning, Policy Regimes and Growth: The Long Term Patterns and Some Specificities of a 'Globalized' Economy," in *Industrial Policy and Development: The Political Economy of Capabilities Accumulation*, ed. by M. Cimoli, G. Dosi, and J. E. Stiglitz, Oxford University Press.

Chen, S.-S. and K. W. Ho (2000): "Corporate diversification, ownership structure, and firm value: The Singapore evidence," *International Review of Financial Analysis*, 9, 315–326.

Christensen, H. K. and C. A. Montgomery (1981): "Corporate economic performance: Diversification strategy versus market structure," *Strategic Management Journal*, 2, 327–343.

Cimoli, M. and G. Dosi (1995): "Technological paradigms, patterns of learning and development: an introductory roadmap," *Journal of Evolutionary economics*, 5, 243–268.

Cimoli, M., G. Dosi, and J. E. Stiglitz (2009): *Industrial policy and development: The political economy of capabilities accumulation*, Oxford Oxford University Press.

Cimoli, M., G. Porcile, and S. Rovira (2008): "Structural change and the BOP- constraint: why did Latin America fail to converge?" *Cambridge Journal of Economics*, 23.

Cireira, X. and W. Maloney (2017): "The Innovation Paradox," Tech. rep., World Bank Group.

Coad, A. and C. Guenther (2013): "Diversification patterns and survival as firms mature," *Small Business Economics*, 41, 633–649.

Dehejia, R. H. and S. Wahba (2002): "Propensity score-matching methods for nonexperimental causal studies," *Review of Economics and statistics*, 84, 151–161.

Delios, A. and P. W. Beamish (1999): "Geographic scope, product diversification, and the corporate performance of Japanese firms," *Strategic Management Journal*, 20, 711–727.

Delios, A., D. Xu, and P. W. Beamish (2008): "Within-country product diversification and foreign subsidiary performance," *Journal of International Business Studies*, 706–724.

Dosi, G. (1984): *Technical Change and Industrial Transformation – The Theory and an Application to the Semiconductor Industry*, London, Mcmillan.

——— (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 N. Mathew (2017a): "The cost-quantity relations and the diverse patterns of "learning by doing": Evidence from India," *Research Policy, forthcoming.*

Dosi, G., M. Grazzi, and D. Moschella (2017b): "What do firms know? What do they produce? A new look at the relationship between patenting profiles and patterns of product diversification," *Small Business Economics*, 48, 413–429.

Dosi, G., R. R. Nelson, and S. Winter, eds. (2000): *The Nature and Dynamics of Organizational Capabilities*, Oxford University Press: Oxford.

Dosi, G., K. Pavitt, and L. Soete (1990): *The Economics of Technical Change and International Trade*, New York: New York University Press.

Dosi, G., D. J. Teece, and S. Winter (1992): "Toward a Theory of Corporate Coherence: Preliminary Remarks," in *Technology and Enterprise in a Historical Perspective*, ed. by G. Dosi, R. Giannetti, and P. Toninelli, Oxford, UK: Clarendon Press.

Dosi, G. and M. Tranchero (2018): "The Role of Comparative Advantage and endowments in Structural Transformation," LEM Working Papers 2018/33, S. Anna School of Advanced Studies.

Dubofsky, P. and P. Vandarajan (1987): "Diversification and measures of performance: Additional empirical evidence." *Academy of Management Journal*, 30, 597–608.

Geringer, M. J., P. W. Beamish, and R. C. DaCosta (1989): "Diversification strategy and internationalization: Implications for MNE performance," *Strategic Management Journal*, 10, 109–119.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010): "Multiproduct firms and product turnover in the developing world: Evidence from India," *The Review of Economics and Statistics*, 92, 1042–1049.

Gomes, J. and D. Livdan (2004): "Optimal diversification: Reconciling theory and evidence," *The Journal of Finance*, 59, 507–535.

Graham, J. R., M. L. Lemmon, and J. G. Wolf (2002): "Does corporate diversification destroy value?" *The Journal of Finance*, 57, 695–720.

Hausmann, R., J. Hwang, and D. Rodrik (2007): "What you export matters," *Journal of economic growth*, 12, 1–25.

Hausmann, R. and D. Rodrik (2003): "Economic development as self-discovery," *Journal of development Economics*, 72, 603–633.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

Henderson, R. and I. Cockburn (1994): "Measuring Competence? Exploring Firm Effects in Pharmaceutical Research," *Strategic Management Journal*, 15, 63–84.

Hidalgo, C. A., B. Klinger, A.-L. Barabási, and R. Hausmann (2007): "The product space conditions the development of nations," *Science*, 317, 482–487.

Hirschman, A. O. (1958): *The strategy of economic development*, Yale University Press.

Jose, M. L., L. M. Nichols, and J. L. Stevens (1986): "Contributions of diversification, promotion, and R&D to the value of multiproduct firms: A Tobin's q approach," *Financial Management*, 33–42.

Khanna, T. and K. Palepu (2000): "Is group affiliation profitable in emerging markets? An analysis of diversified Indian business groups," *The Journal of Finance*, 55, 867–891.

Lamont, O. A. and C. Polk (2001): "The diversification discount: Cash flows versus returns," *The Journal of Finance*, 56, 1693–1721.

Lang, L. H. and R. M. Stulz (1994): "Tobin's q, Corporate Diversification, and Firm Performance," *The Journal of Political Economy*, 102, 1248–1280.

Leonard-Barton, D. (1992): "Core capabilities and core rigidities: A paradox in managing new product development," *Strategic management journal*, 13, 111–125.

Maddala, G. S. (1986): "Disequilibrium, self-selection, and switching models," *Handbook of Econometrics*, 3, 1633–1688.

Maksimovic, V. and G. Phillips (2002): "Do conglomerate firms allocate resources inefficiently across industries? Theory and evidence," *The Journal of Finance*, 57, 721–767.

Markides, C. C. (1995): "Diversification, restructuring and economic performance," *Strategic Management Journal*, 16, 101–118.

Markides, C. C. and P. J. Williamson (1996): "Corporate diversification and organizational structure: A resource-based view," *Academy of Management Journal*, 39, 340–367.

Mathew, N. (2017): "Drivers of firm growth: micro-evidence from Indian manufacturing," *Journal of Evolutionary Economics*, 27, 585–611.

Matsusaka, J. G. (2001): "Corporate Diversification, Value Maximization, and Organizational Capabilities," *The Journal of Business*, 74, 409–431.

McKelvie, A. and J. Wiklund (2010): "Advancing firm growth research: A focus on growth mode instead of growth rate," *Entrepreneurship theory and practice*, 34, 261–288.

McMillan, M. S. and D. Rodrik (2011): "Globalization, structural change and productivity growth," Tech. rep., National Bureau of Economic Research.

Michel, A. and I. Shaked (1984): "Does business diversification affect performance?" *Financial Management*, 18–25.

Miller, D. J. (2004): "Firms' technological resources and the performance effects of diversification: a longitudinal study," *Strategic Management Journal*, 25, 1097–1119.

Nelson, R. R. (1994): "The co-evolution of technology, industrial structure, and supporting institutions," *Industrial and corporate change*, 3, 47–63.

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.

Palich, L. E., L. B. Cardinal, and C. C. Miller (2000): "Curvilinearity in the diversification–performance linkage: an examination of over three decades of research," *Strategic management journal*, 21, 155–174.

Park, C. (2003): "Prior performance characteristics of related and unrelated acquirers," *Strategic Management Journal*, 24, 471–480.

Pavitt, K. (1984): "Sectoral Pattern of Technical Change: Towards a taxonomy and a theory," *Research Policy*, 13, 343–373.

Pehrsson, A. (2006): "Business relatedness and performance: A study of managerial perceptions," *Strategic Management Journal*, 27, 265–282.

Penrose, E. (1959/1995): *The Theory of the Growth of the Firm*, Oxford University Press.

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.

Prebisch, R. (1950): "The Economic Development of Latin America and its Principle Problem," *UN Economic Commission for Latin America.*

Robins, J. and M. F. Wiersema (1995): "A resource-based approach to the multibusiness firm: Empirical analysis of portfolio interrelationships and corporate financial performance," *Strategic Management Journal*, 16, 277–299.

Rumelt, R. P. (1974): *Strategy, structure, and economic performance*, Harvard Business School Press, Boston, MA.

——— (1982): "Diversification strategy and profitability," *Strategic Management Journal*, 3, 359–369.

Santarelli, E. and H. T. Tran (2013): "Diversification Strategies and Firm Performance: A Sample Selection Approach," Working Paper No. 896, Dipartimento Scienze Economiche, Universita' di Bologna.

Tacchella, A., M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero (2012): "A New Metrics for Countries' Fitness and Products' Complexity," *Scientific Reports*, 2.

Teece, D. J. (1982): "Towards an economic theory of the multiproduct firm," *Journal of Economic Behavior & Organization*, 3, 39–63.

——— (2000): "Firm capabilities and economic development: implications for newly industrializing economies," in *Technology, Learning & Innovation: Experiences of Newly Industrializing Economies*, ed. by L. Kim and R. R. Nelson, Cambridge University Press, Cambridge, 105–128.

Teece, D. J., G. Pisano, and A. Shuen (1997): "Dynamic Capabilities and Strategic Management," *Strategic Management Journal*, 18, 509–533.

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.

Villalonga, B. (2004a): "Diversification discount or premium? New evidence from the business information tracking series," *The Journal of Finance*, 59, 479–506.

——— (2004b): "Does Diversification Cause the "Diversification Discount"?" *Financial Management*, 5–27.

Vonortas, N. S. (1999): "How do participants in research joint ventures diversify?" *Review of Industrial Organization*, 15, 263–281.

Zaccaria, A., M. Cristelli, A. Tacchella, and L. Pietronero (2014): "How the Taxonomy of Products Drives the Economic Development of Countries," *PloS ONE*, 9, e113770.

Zuckerman, E. W. (2000): "Focusing the corporate product: Securities analysts and de-diversification," *Administrative Science Quarterly*, 45, 591–619.