DISCUSSION PAPER SERIES

DP17945

INNOVATION, PRODUCTIVITY AND FIRM PERFORMANCE

Abigail Watt, James McCann and Jeremy Lawson

OCCASIONAL PAPERS

ISSN 0265-8003

INNOVATION, PRODUCTIVITY AND FIRM PERFORMANCE

Abigail Watt, James McCann and Jeremy Lawson

Discussion Paper DP17945
Published 28 February 2023
Submitted 27 February 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

INNOVATION, PRODUCTIVITY AND FIRM PERFORMANCE

Abstract

This paper systematically assesses the impact of innovative activities on US firms' total factor productivity and in turn the implications for firm performance. Initially, we find that firm-level productivity trends in our sample of over 2000 Russell 3000 companies match those of the wider US economy. Applying a hierarchical clustering algorithm to measures of innovation we find that there are clusters of innovation within the Russell 3000 and that these tend to be related to higher productivity. To assess this relationship more robustly we then use a fixed effects panel model to directly test innovation measures as drivers of productivity. We find that research and development (R&D) spending and intangible investment, the size of a firm's patent portfolio and the quality of its patents, increase a firm's level of productivity. An innovation index is then created, ranking the relative importance of these innovation factors for the firm's productivity. The impact of a firm's innovation rankings on its returns is then assessed and we find that there is a strong statistically significant positive impact here - with those higher up in the index on average delivering stronger returns. The pathway of companies through the index overtime is also studied and we find that there is evidence of innovation mattering for firm performance but also of firms' need to continually innovate to stay ahead of the pack.

JEL Classification: O31, O32, O34, D24

Keywords: Innovation, Innovation processes, R&d, Intangible assets, Intellectual property, Factor productivity, Firm performance, Firm size

Abigail Watt - abigail.watt@abrdn.com
abrdn Research Institute

James McCann - james.mccann@abrdn.com
abrdn Research Institute

Jeremy Lawson - jeremy.lawson@aberdeenstandard.com
abrdn Research Institute and CEPR

INNOVATION, PRODUCTIVITY AND FIRM PERFORMANCE

Abigail Watt[1], James McCann[2] and Jeremy Lawson[3]

February 2023

Abstract

This paper systematically assesses the impact of innovative activities on US firms' total factor productivity and in turn the implications for firm performance. Initially, we find that firm-level productivity trends in our sample of over 2000 Russell 3000 companies match those of the wider US economy. Applying a hierarchical clustering algorithm to measures of innovation we find that there are clusters of innovation within the Russell 3000 and that these tend to be related to higher productivity. To assess this relationship more robustly we then use a fixed effects panel model to directly test innovation measures as drivers of productivity. We find that research and development (R&D) spending and intangible investment, the size of a firm's patent portfolio and the quality of its patents, increase a firm's level of productivity. An innovation index is then created, ranking the relative importance of these innovation factors for the firm’s productivity. The impact of a firm's innovation rankings on its returns is then assessed and we find that there is a strong statistically significant positive impact here - with those higher up in the index on average delivering stronger returns. The pathway of companies through the index overtime is also studied and we find that there is evidence of innovation mattering for firm performance but also of firms’ need to continually innovate to stay ahead of the pack.

JEL Classification: O31, O32, O34, D24, L250

Keywords: Innovation, Innovation Processes, R&D, Intangible Assets, Intellectual Property, Factor Productivity, Firm Performance, Firm Size.

[1] Research Economist, abrdn Research Institute, 28 State Street, Boston, MA 02109. E-mail: abigail.watt@abrdn.com

[2] Deputy Chief Economist for Macroeconomic Research, abrdn Research Institute, 28 State Street, Boston, MA 02109. E-mail: james.mccann@abrdn.com

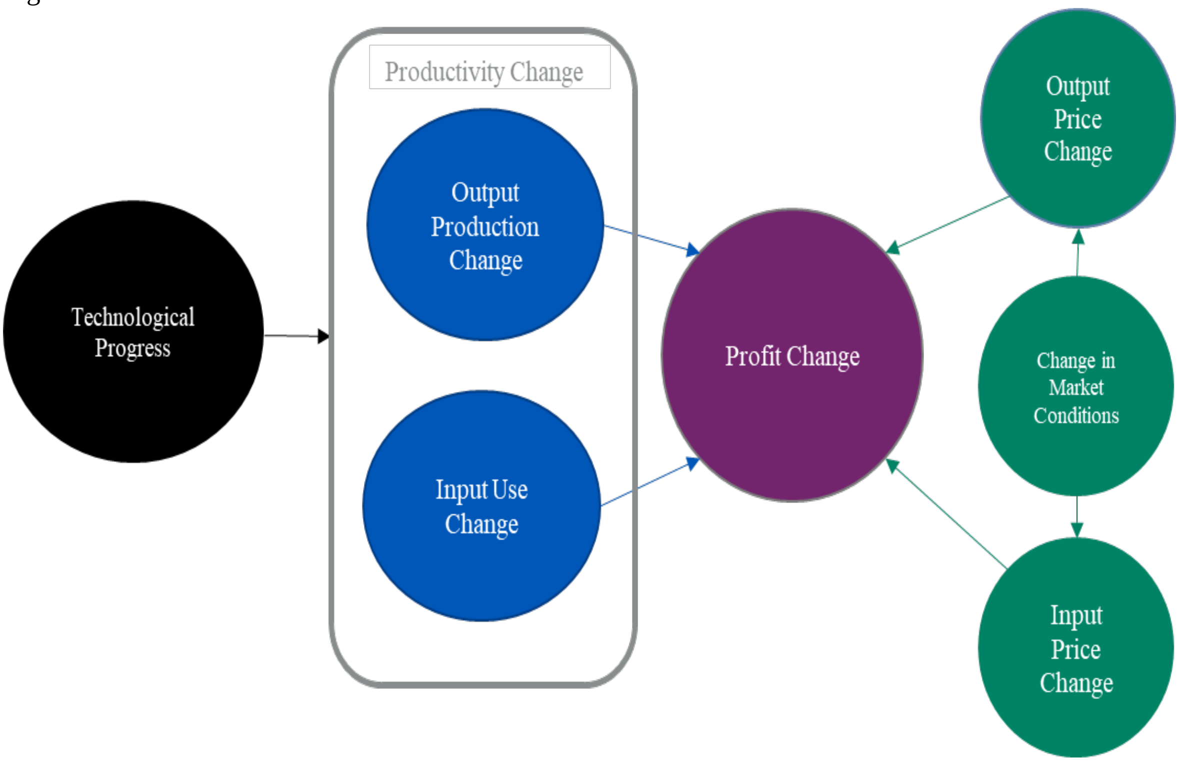
[3] Chief Economist and Head of the abrdn Research Institute, abrdn Research Institute, 280 Bishopsgate, London, EC2M 4RB. E-mail: jeremy.lawson@abrdn.com

1 Introduction

Quantifying the innovative process is challenging but improved data methods have led to the emergence of more firm-level data than ever before. In this research we aim to capture this innovative process and better understand the importance of this for firm-level productivity and firm performance. We also aim to generate a tractable database which can be readily updated.

At the core of this paper is the measurement of firm-level productivity of companies in the Russell 3000, which captures the top 3000 US-listed companies by market capitalisation, between 2000 and 2019 where data is available. Productivity is the channel through which innovation feeds in the generation of returns for the firm and Figure 1 shows the process that underpins the analysis in this paper. In this process technological progress driven by innovation feeds into a change in productivity for the firm by generating a more efficient use of inputs and/or an increase in output leading to an increase in profitability. Outwith this, there numerous factors that can also influence the firm's profitability, hence when identifying the effect of innovation, controlling for other confounding factors is critical.

Figure 1: The Innovation Process



Source: abrdn, as of 2022

There have been many studies of the dynamics of productivity at the firm level and how these relate to the innovative activities undertaken by firms. Hall (2011) provides a good overview of these, finding that many of these studies look across countries and sectors, and considering several types of innovative activity and different estimation methods for productivity itself. Most papers use the Crépon, Duguet and Mairesse (1998) (CDM) framework to assess the link between innovation and productivity.

This framework is an innovation output-oriented approach that starts with a selection equation separating those firms that undertake innovative activities from those that do not. A knowledge production function which relates innovation inputs and other factors to innovation outputs is then formulated. CDM then estimate a Cobb Douglas production function augmented with the innovative activities of firms to capture the impact of innovation on productivity. Many papers using this approach find significant and positive relationship between various measures of innovation and different measures of productivity but not always. Griffith et al (2006) use survey dummies for product and process innovation with mixed results across countries – the effect of process innovation was only significant and positive in France but product innovation was significant and positive for all countries studied bar Germany. Another group of papers use log innovative sales share to measure product innovation – Benavente (2006), CDM (1998), Jefferson et al (2006) – all find a positive and statistically significant effect across the countries studied. Finally, some studies combine log innovative sales with process dummies from surveys. van Leeuwen and Klomp (2006) find that product innovation has a positive and significant effect but process innovation has a negative, significant effect. Janz. Loof and Peters (2009) also find this for Germany.

A key difference between our study and those previously discussed is that we are using a panel of firms across time rather than a fixed cross section of firms. This means that we focus on R&D expenditures, patenting activity and intangible investment across time as our measures of innovative activity rather than survey-based measures of innovation. This is because we do not have surveys that are consistently measured across the universe of listed companies that we use in the study. Given this panel setting, we choose to first quantify firm-level productivity and then regress the innovation measures on productivity. This is also useful because we wish to construct an index through which we can rank companies based on the contribution of innovative activities to their level of productivity.

This paper continues as follows, we first estimate firm-level productivity, then, in the second section, we describe the methods used to compute indicators of innovation at the firm level and then clustering analysis is used to explore whether there are patterns across the measures of innovation and the characteristics of firms in section 3. Following the preliminary evidencing of the clustering effects that are identified, the same data is then analysed in a panel data setting alongside a set of controls to better understand the importance of these measures for the generation of firm-level productivity. Fitting the final piece of the puzzle, the evidence of the importance of these innovation proxies for productivity is then analysed against firm performance.

The key findings of this research paper are: that productivity is increased by innovative activities such as R&D spending, intangibles investment and patenting; that there are clusters of listed US firms which share similar characteristics when it comes to the innovation process; and that the extent to which these activities are carried out can positively impact firms' returns.

2 Total Factor Productivity at the Firm Level

2.1. Background

To understand how to better identify the innovative features of companies and the implications of this for productivity and profitability, we must first estimate productivity. Productivity can be defined as a measure of the efficiency with which a firm transforms inputs into outputs in the production process (Hulten, 2001). Whilst much work has been done to understand how to identify this, there remain several methods commonly used in the literature today. A useful summary and assessment of these methods can be found in Ornaghi and Van Bereven (2011) but here we summarise the key features of those most used. Each method has its own pros and cons but is centred on the estimation of the firm-level production function, equation 1. Where output for firm i at time t, Y_{it}, is a function of inputs - capital, K_{it}, and labour, L_{it} - and productivity A_{it}, with β_k and β_L representing the share of capital and labour in the firm's production process.

$$Y_{it} = A_{it}K_{it}^{\beta_K}L_{it}^{\beta_L} \ (1)$$

Equation 1 is traditionally log-linearised to aid estimation resulting in equation 2 where lower-case letters are the natural log of the variables.

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \mu_{it}(2)$$

This gives us log total factor productivity (TFP), $\Omega_{it} = \ln(A_{it}) \equiv \beta_0 + \omega_{it} + \mu_{it}$. Where $\beta_0$is the mean level efficiency level across firms, over time, and $\omega_{it} + \mu_{it}$ measures time and firm specific deviations from the mean with this partly observable, ω_{it}, and partly an unobservable component represented unexpected deviations from the mean, μ_{it}.

The simplest and least data intensive methods for identifying firm-level TFP are index methods, such as the Solow residual, where the inputs for a firm are weighted according to their factor shares and the resulting residual output when compared to actual value added is defined as the TFP for the firm.

Contrastingly, approaches which involve the formal estimation of the production function require not just a wealth of data but also high quality data to ensure unbiased estimation. The simplest method to estimate econometrically is via ordinary least squares normally, industry-by-industry, using data for firms across time. The main drawback of this approach is that productivity is likely to be known by the firms when selecting input choices and hence estimated coefficients will be biased and inconsistent. Semi-parametric methods, such as those proposed by Olley and Pakes (1996) and Levinson and Petrin (2003), aim to overcome this endogeneity issue by using a proxy variable for unobserved productivity. However, the proxy variables used may introduce other biases if poorly measured or if they are poor instruments for productivity itself. More recently, Wooldridge (2009) improved these original methods by proposing a more efficient single stage estimation procedure, but this may still suffer from biased coefficient estimates.

In this paper we choose to use the simpler approach of the Solow residual method, even though it differs from the more complex econometric methods in the sense that we are assuming constant returns to scale and do not formally estimate input weights. We think concerns over both data quality and availability for the possible proxy variables required for other more complex methods mean the simpler index method approach is more appropriate.

2.2. Data

The study focuses on US-listed firms from the Russell 3000 Index and data limitations mean the final sample consists of 2194 firms with annual data spanning from 2000 to 2019 -- to avoid any contamination from pandemic related effects. The dataset is imbalanced in nature with firms entering and leaving the sample – it is however ensured that at least five years of continuous data are available for a firm to remain in the analysis. A modified version of the UN (2008) ISIC Rev4 is used to categorise the firms into 25 different sectors, see Table 1. We strike a balance between detail and having enough similar companies in each sector to identify sector effects. We modify this by grouping together classifications D26-27, D28-D30 and D58-D61 and breaking up the services sector to obtain a greater level of granularity, given it is increasingly complex to efficiently categorise. We dropped defence and education companies from the data and did not have a sufficient sample size to include arts & recreation.

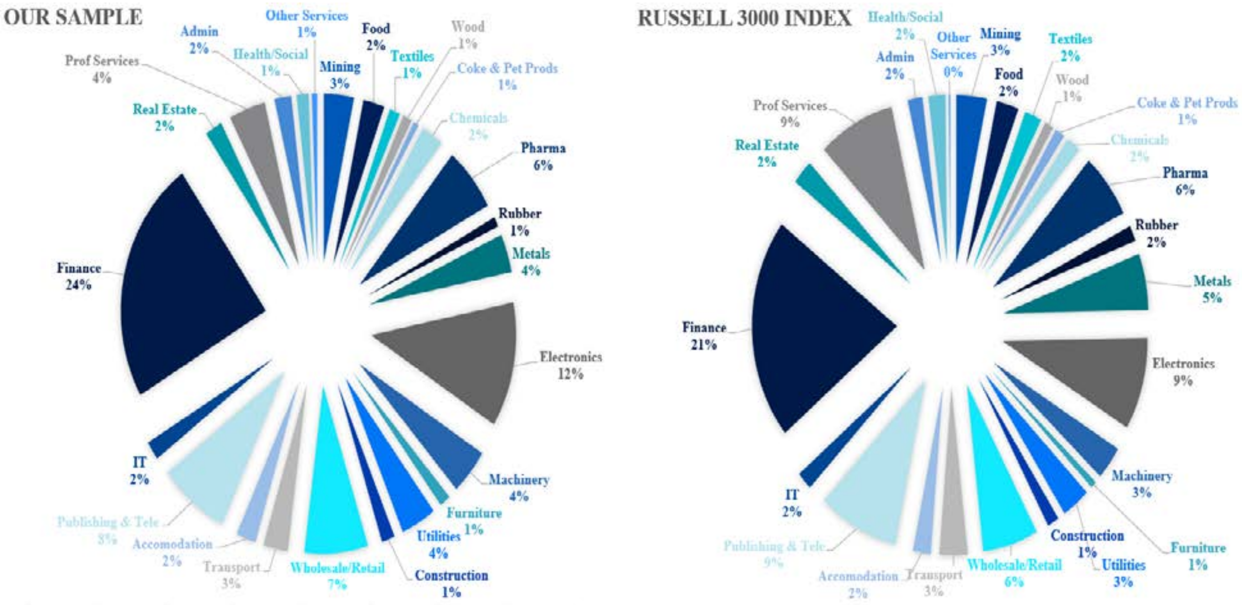
Table 1: Sector definitions

Sector Number	Aggregation of ISIC Rev 4	Our Sector Name
1	D05T09: Mining and quarrying [B]	Mining
2	D10T12: Food products, beverages and tobacco [CA]	Food
3	D13T15: Textiles, wearing apparel, leather and related products [CB]	Textiles
4	D16T18: Wood and paper products, and printing [CC]	Wood
5	D19: Coke and refined petroleum products [CD]	Coke & Pet Prods
6	D20: Chemicals and chemical products [CE]	Chemicals
7	D21: Basic pharmaceutical products and pharmaceutical preparations [CF]	Pharma
8	D22T23: Rubber and plastics products, and other non-metallic mineral products [CG]	Rubber
9	D24T25: Basic metals and fabricated metal products, except machinery and equipment [CH]	Metals
10	D26T27: Computer, electronic, electrical and optical products [CI]	Electronics
11	D28T30: Machinery, transport equipment and equipment n.e.c. [CK]	Machinery
12	D31T33: Furniture; other manufacturing; repair and installation of machinery and equipment [CM]	Furniture
13	D35T39: Electricity, gas and water supply; sewerage, waste management and remediation activities [D-E]	Utilities
14	D41T43: Construction [F]	Construction
15	D45T47: Wholesale and retail trade, repair of motor vehicles and motorcycles [G]	Wholesale/Retail
16	D49T53: Transportation and storage [H]	Transport
17	D55T56: Accommodation and food service activities [I]	Accomodation
18	D58T61: Publishing, audiovisual and broadcasting activities, telecoms [JA]	Publishing & Tele
19	D62T63: IT and other information services [JC]	IT
20	D64T66: Financial and insurance activities [K]	Finance
21	D68: Real estate activities [L]	Real Estate
22	D69T75: Professional, scientific and technical activities [M]	Prof Services
23	D77T82: Administrative and support service activities [N]	Admin
24	D86T88: Human health and social work activities [Q]	Health/Social
25	D94T96: Other service activities [S]	Other Services

Source: UN, abrdn as of 2022

This classification leads to the sample displayed in Figure 2, which closely matches the overall sector split of the investable universe of the Russell 3000 Index itself. We have a slightly higher representation of electronics companies at the expense of professional services, rubber, and metals firms.

Figure 2: Sector breakdown our sample versus the Russell 3000 Index



Source: Refinitiv, Bloomberg, abrdn as of 2022

2.3. Method

Following the Solow (1957) method for calculating firm level productivity, value added is first computed as the sum of the labour and capital inputs for each firm, where capital is taken to be equal to the firm's earnings before interest, tax, depreciation, and amortisation (EBITDA), a measure of profits capturing the share of a firm's income going to capital, and is sourced from the Thomson Reuters Worldscope database (Refinitiv, 2020). The labour input is slightly more complicated to come by. Where available, the reported staff costs data from the Worldscope database is used, but, if unavailable, the imputation method outlined in Gal (2013) is applied - with labour costs calculated as the average wage in a sector, in a given year, multiplied by the number of employees for the firm in the same year. This suffers from measurement error – as firms wage distributions vary from the industry average and more productive firms will deviate from this by more. However, in the face of a lack of data, this imputation is preferred to forgoing observations, and evidence from Gal (2013) suggests that productivity measures based on these imputations are still highly correlated with non-imputed estimates across countries.

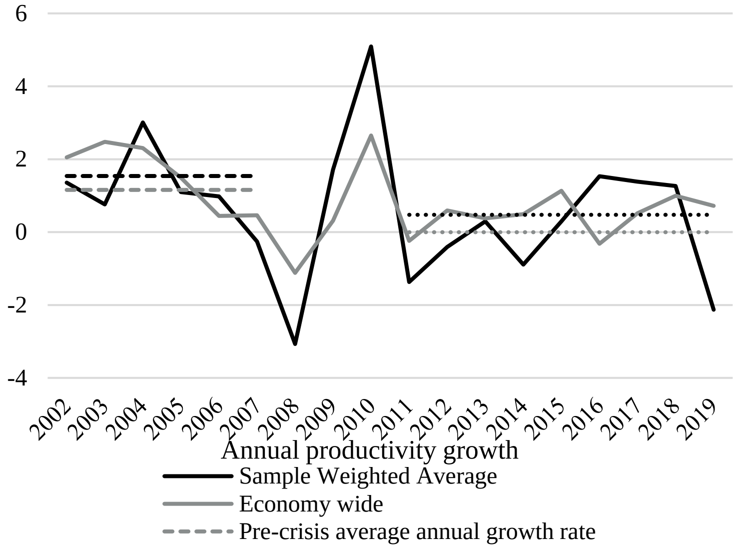
The OECD STAN database is used to impute the labour share for each sector as the share of total labour costs in each industry's value added each year (OECD, 2022). The capital share is then assumed to be one minus the labour share.

With this componentry in place, firm level productivity is identified at the part of value added which is not explained by the relative use of capital and labour inputs by the firm. As all variables are measured in logs the computed measure of productivity is a log level.

2.4. Trends in firm level productivity

The path of annual productivity growth in our sample-weighted average follows the pattern of economy wide multifactor productivity growth over the same period (see Figure 3). The firm-level average is unsurprisingly more volatile than the aggregate economy but the broad trends in productivity growth are similar – with the post-financial crisis period showing a much slower pace of growth than the pre-crisis period as is well documented in the productivity literature (Sprague, 2021).

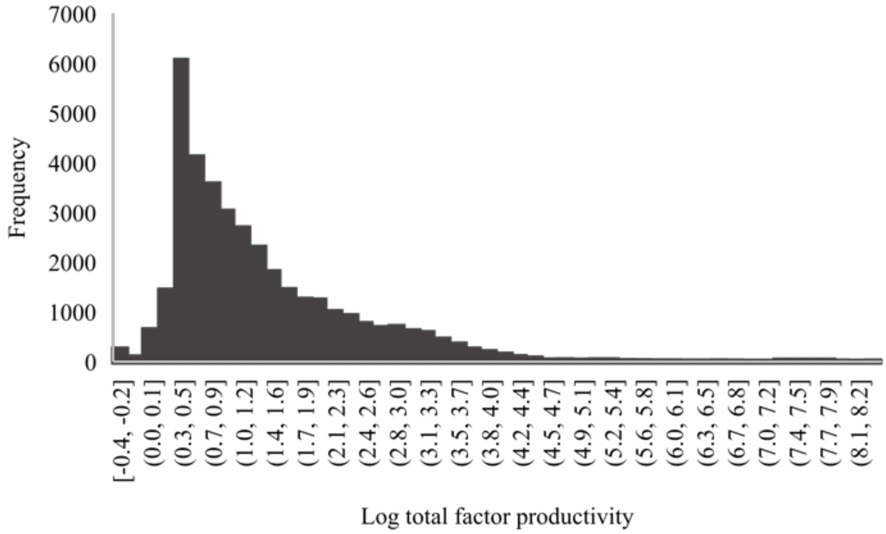
Figure 3: Growth in average firm level productivity versus economy wide productivity growth



Source: abrdn, Refintiv, Haver as of 2022

The distribution of productivity as shown in Figure 4, is positively skewed with a long right-hand tail. This suggests that there is a group of highly productive firms sitting the upper tail of the distribution – in line with the findings of Andrews et al (2015) – who find a similar frontier of productive companies across 23 OECD economies.

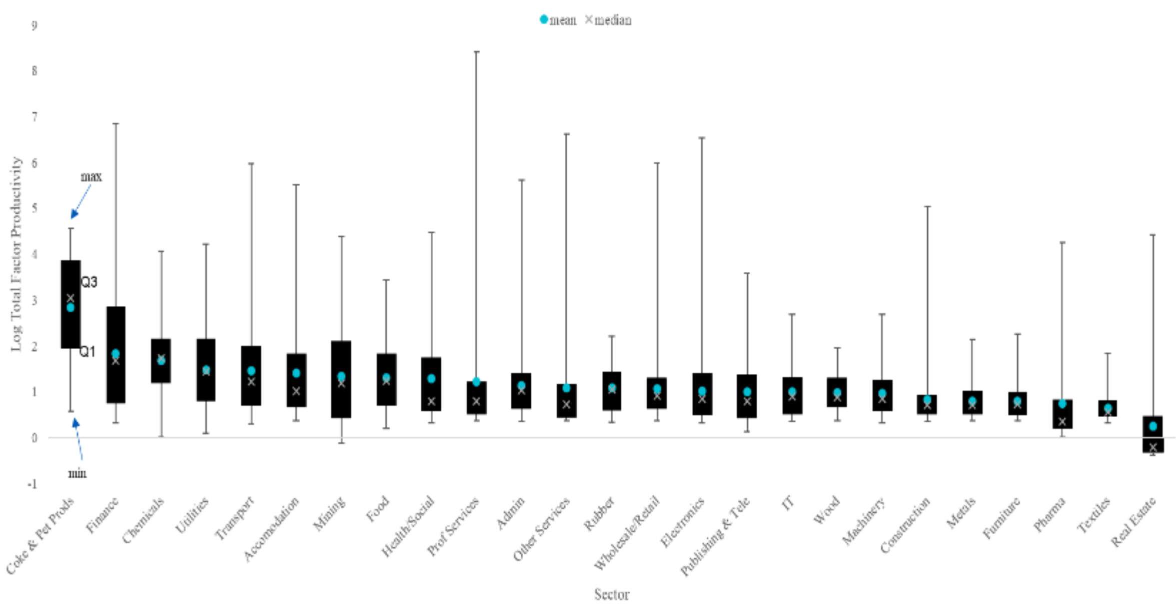
Figure 4: Distribution of firm-level productivity



Source: OECD, Refinitiv, abrdn, as of 2022

But there is evidence that this distribution differs by sector. Figure 5 plots the distribution of productivity in each sector, ranked from the highest average productivity to the lowest. At the upper end of the spectrum the highest average productivity can be found in coke and petroleum producers whilst real estate firms sit at the other end. This could be a function of the difficulty in measuring the input/output process for listed real estate firms, which in the case of this sample are more likely to be real estate investment trusts than property developers, hence they may not be well captured by the traditional production function approach.

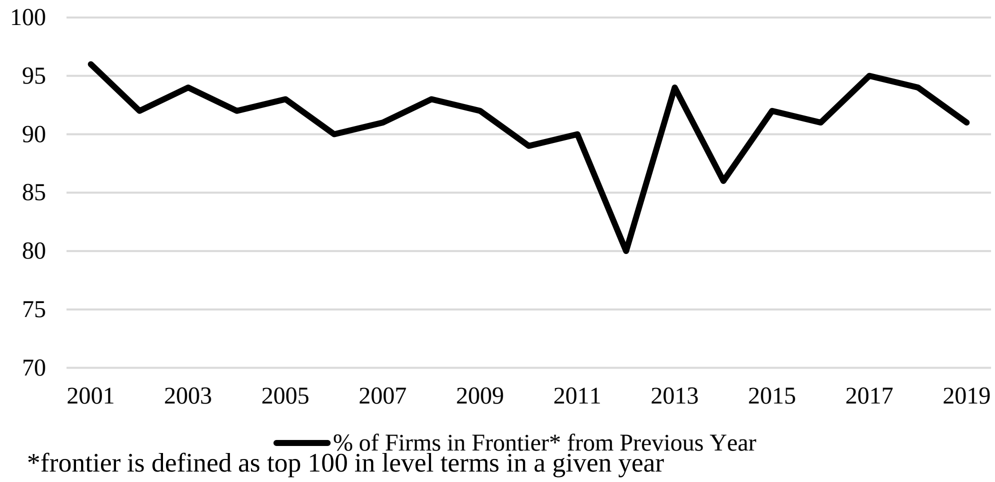
Figure 5: Distribution of productivity by sector



Source: OECD, Refinitiv, abrdn, as of 2022

Figure 5 also shows that positive outliers are a feature of the productivity distribution across the sectors, suggesting that the fat right-hand tail in the overall productivity distribution is not concentrated in a specific sector. There is also persistence in the firms that make up the frontier. Figure 6 shows that, when we consider the top 100 companies with the highest levels of productivity in each year, over 90% of firms remain the same from year to year on average.

Figure 6: Persistence in the frontier



Source: OECD, Refinitiv, abrdn, as of 2022

3 Capturing the Innovation Process

To better understand these trends in productivity and identify the features of those companies at the frontier, we look to better capture the innovation process at the firm level. Quantifying innovation is challenging. Certain aspects of a company's structure such as management or culture cannot be readily quantified.

The OECD's 2018 update of the Oslo Manual, which outlined guidelines for data on innovation, switched their previous four types of innovation (product, process, organisational and marketing) to two key measures:

- Product innovation – "new or improved good or service that differs significantly from the firm's previous goods or services and that has been introduced on the market." – OECD/Eurostat (2018).
- Business process innovation – "new or improved business process for one or more business functions that differs significantly from the firm's previous business processes and that has been brought into use by the firm." – OECD/Eurostat (2018).

Our aim is to include data which measure inputs and outputs of the innovation process to capture the elements of innovation as outlined above. This section will dive into the measures which we have chosen to focus on.

Table 2: Quantifying innovative activities

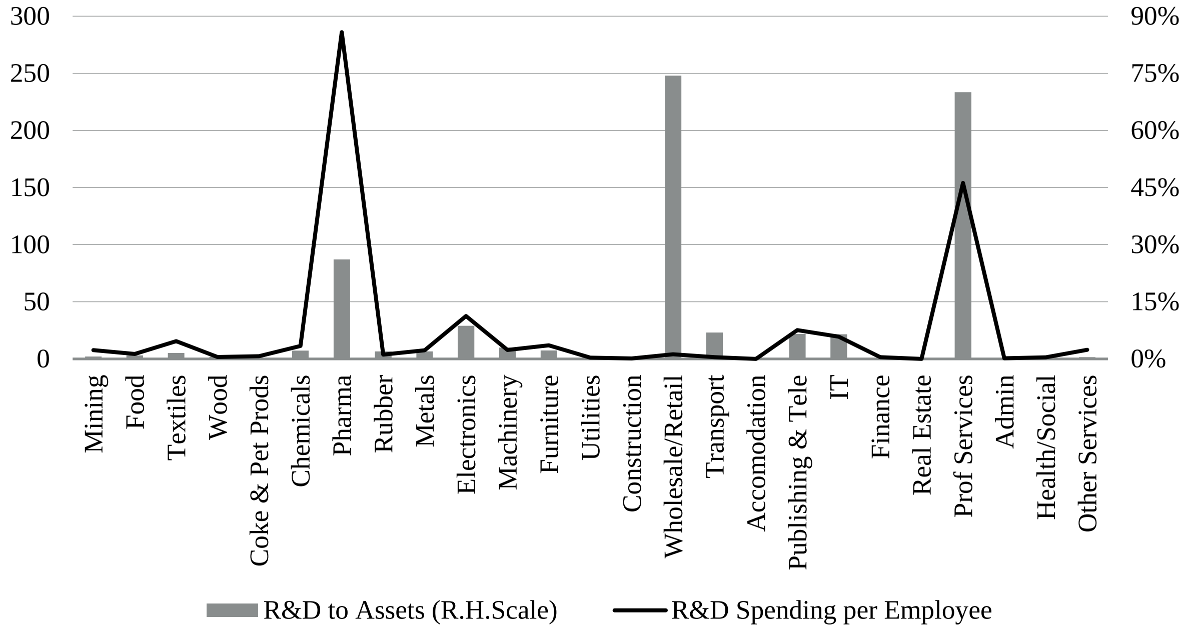
	Variable	Description	Transformation
Traditional Innovation Measures	Intangible Assets Share	Intangible assets divided by total assets	level
	R&D per Employee	R&D expenditure divided by number of employees	log
	R&D to Assets	R&D expenditure divided by total assets	log
Financing Innovation	Sales per Employee	Net sales or revenues divided by number of employees	level
	Debt to Equity	Debt to equity	log
	Interest Coverage Ratio	Earnings divided by interest expense	log
	Gross Profit Margin	(Net sales – Cost of Goods Sold)/Net Sales	log
Patents Database	New Patent Activity*	Number of patents filled per year	log
	Cumulative Patent Portfolio*	20 year rolling sum of filled patents	log
	Patent Scope*	Scope of a patent equals the number of distinct 4 digit subclasses in the patent document, normalised according to the maximum scope value of patents in the same cohort, these cohorts defined by year of filing and technology field.	Normalised
	Family Size*	Number of patent offices which a patent has been filed at normalised by maxvalue exhibited by other patents in same year and tech cohort.	Normalised
	Backward Citations*	Citations per patent normalised according to max value received by patents in that year, and tech cohort. Includes self citations.	Normalised
Controls	Foreign Sales % of Total Sales	International sales divided by net sales or revenues	log
	Number of Employees	Total number of employees in workforce.	level

Note: all data sourced from Refinitv Worldscope database apart from those marked with * which were sourced from the OECD’s patent statistics. We also adjust the data in real terms using the OECD STAN value added deflators on an industry and industry basis.

3.1. Common measures of innovation

First, we capture R&D spending by firms – this is likely to be an indication of an innovative firm as research is a core part of innovating. It is the case that some sectors are more likely than others to undertake this type of activity (see Figure 7), it can be poorly measured in firm accounts and this type of spending can be sporadic – with the effects taking time to be felt in terms of productivity gains. Rather than using the pure level of R&D spending by firms, we use this as a ratio to the number of employees or the total assets of the firm. The aim is to try and account for large firm effects which may skew this data. We want to capture *effective* R&D spending therefore we need to account for larger firms needing to allocate a higher absolute amount and a similar relative share of resources to these research endeavours.

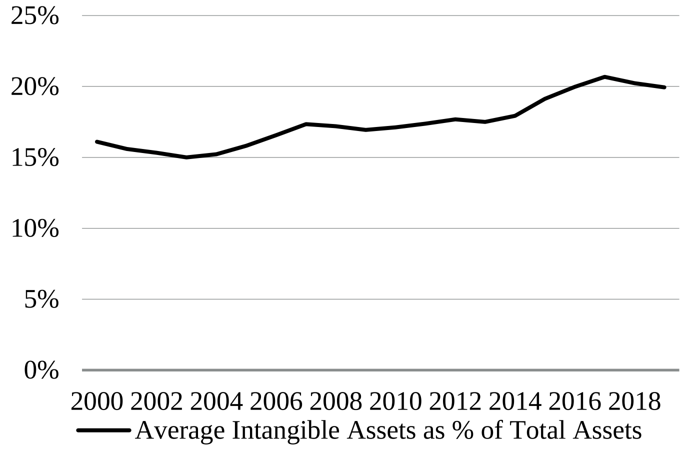
Figure 7: R&D spending per employee and R&D as a share of assets by sector



Source: Refinitiv, abrdn, as of 2022

We also explore innovation through the lens of a firm's intangible asset share. Intangible assets include investment into software, training and other non-tangible goods or services. Marrocu et al (2009) have found intangible investment to be a statistically significant driver of firm-level productivity when assessed across six European countries and we therefore include this in our panel of US firms. Through the digital revolution of the past decades the intangible asset share has risen for firms, see Figure 8.

Figure 8: Intangible asset share overtime



Source: Refinitiv, abrdn, as of 2022

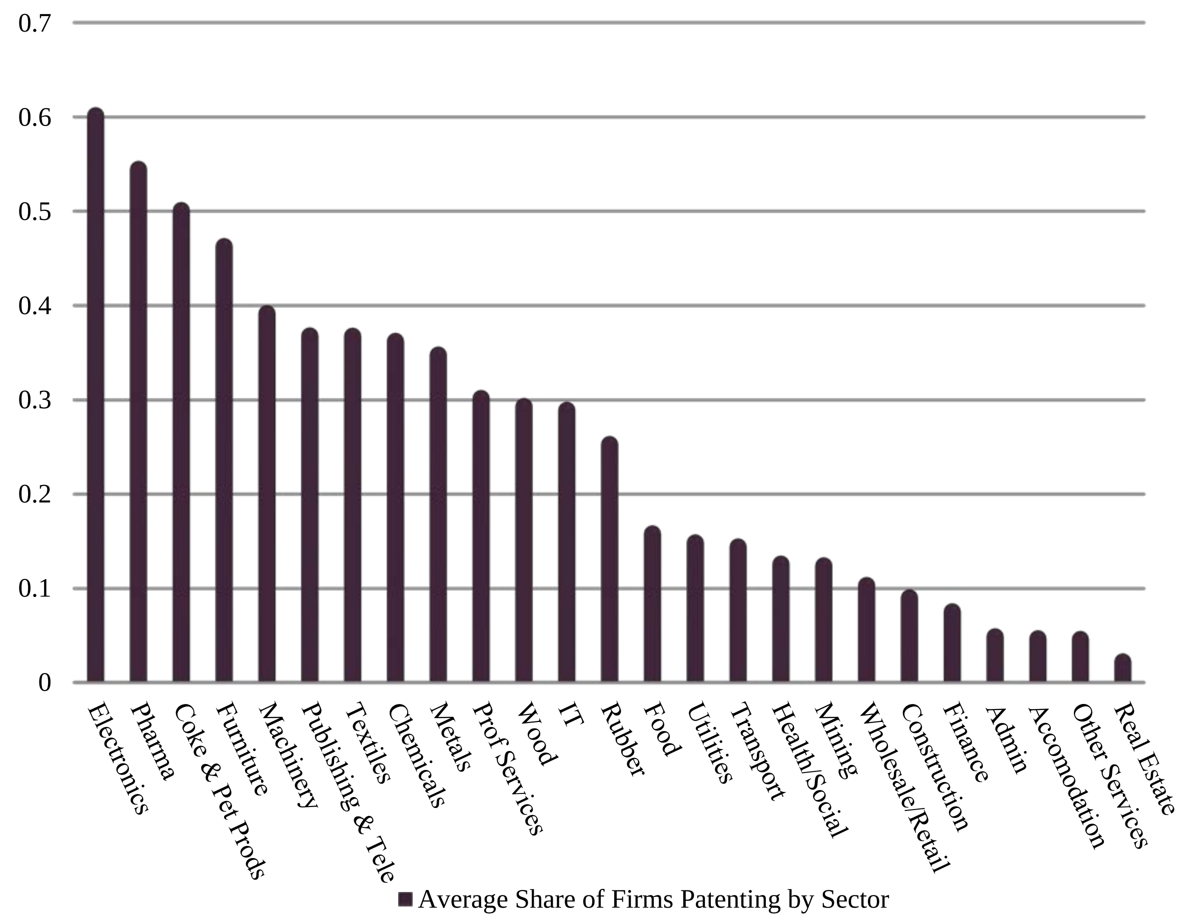
3.2. Patent Indicators

The second set of indicators that we are using to capture the innovation process at the firm level focuses on a firm's patenting activity – patents filed by inventors within firms to protect intellectual property. Hall and Haroff (2012) provide a good overview of the use of patent data in the innovation literature, whilst Griliches (1990) looks at patents as an innovation proxy specifically noting that the outright count of patents alone might not be sufficient. We are conscious that patents may be used as a barrier to entry (Boldrin and Levine, 2013) to defend a company's market share. Hence, we choose to capture not just the outright patenting activity but also the quality of the patenting activity of each firm.

We build out a database of indicators from the OECD's patent data (OECD, 2019) where patents that have been filed by individuals are deanonymized and matched to company names. We use a fuzzy matching process to link OECD company names to the names of the companies in our sample. This process is required due to different naming conventions used through the OECD data for the same company. For example, Procter and Gamble could be included as "P&G Company," "The Procter and Gamble Company" or even with a spelling mistake such as "Teh Procter and Gamble Company." By using fuzzy matching, where the Levenshtein (1966) distance between two strings is calculated based on additions, subtractions, and substitutions, we can try and assign the closest matches to our sample of companies. But in the interest of transparency this process is not perfect and there is scope to both over and underestimate the patent activity of companies. This may be a source of measurement error in the data.

We calculate both the cumulative number of patents assigned to a firm over time, a firm's patent portfolio, and the number of new patents filed from year to year. As with the R&D spending – some sectors are more prolific patentors – Figure 9 shows the average share of firms patenting by sector – electronics companies come out on top, closely followed by pharma and coke & petroleum producers.

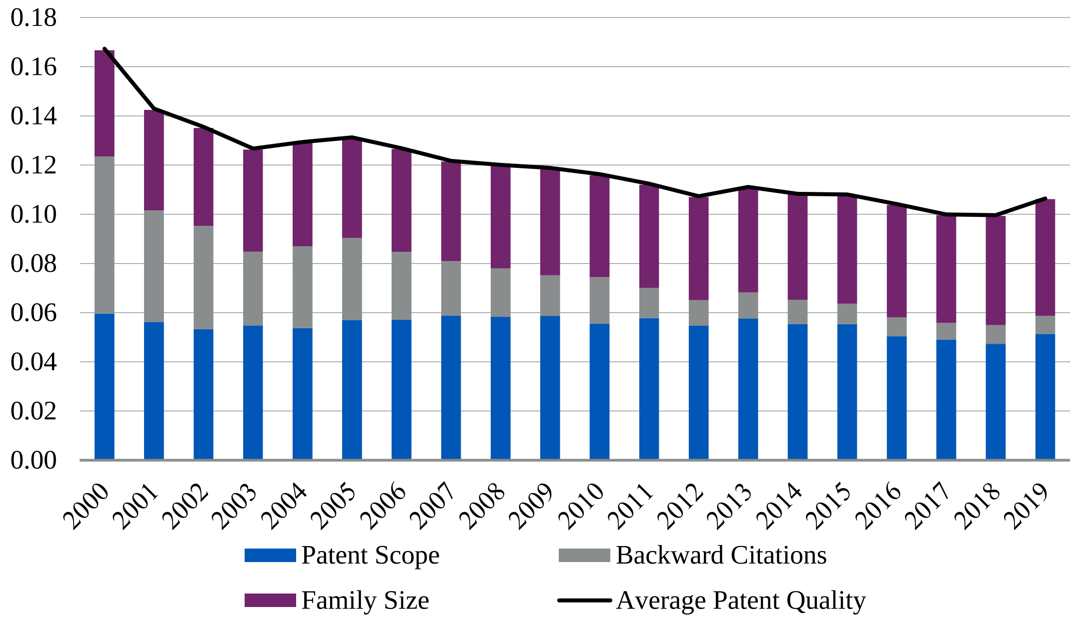
Figure 9: Shares of firms patenting overtime



Source: OECD, abrdn as of 2022

But it is not just the mere act of patenting that we are interested in. The OECD (2019) also captures several measures of patent quality, which we include in our paper. The first measure of interest is patent scope – where the number of technology fields covered in the patent are counted. To aggregate this overtime, we take a firm's average number of technology fields covered in their newly-issued patents each year. The next measure looks at the patent family size, patent families are the same patent being filed at different regional patenting offices. A greater family size should in theory indicate something about the quality of the idea if a company is patenting this at multiple patent offices. The final quality indicator we consider are backward citations that capture the number of previous patents that a patent cites in its own filing, including self-citation. This should tell us something about the research involved in generating the idea being patented. One issue with this data could be that, if an idea is in its infancy, it may have less patents to cite in its own filing. Whilst a high number of self-citations may be sign of defensive patenting, the data does not allow for this differentiation. This could be one avenue for exploration in future research papers. Again, we look at the average number of backward citations for a firm each year. We also follow Squicciarini et al (2013) and normalise all the quality indicators relative to the cohort maximum each year where the cohort is defined by the technology field.

Figure 10: average patent quality overtime



Source: OECD, abrdn as of 2022

3.3. Financing Innovation

In addition to direct measures of innovative activities this paper is also interested in understanding the financing of innovation. We therefore explore the financial state of the firms in our sample by considering financial accounting ratios such as debt to equity, interest coverage and the gross profit margins. This enables us to identify which companies are best placed to undertake innovative activities and the costs of innovation too.

3.4. Control Variables

The final set of variables which are important to capture to robustly identify the effects of innovation upon both productivity and performance of firms are control variables. These variables are included to ensure that we do not incorrectly attribute the effect of innovation when in fact there is another confounding factor, such as firm size, which is instead driving higher productivity – as found in Van Ark and Monnikhof's (1996) study of five OECD economies including the US. In this analysis we focus on the number of employees, sales per employee and foreign sales as a percentage of total sales to capture some of these agglomeration effects.

4 Patterns of Innovative Activity

This stage of our analysis assesses if there are clusters of similarity across the companies within our data, and studies the features of the distinct clusters which have been identified. Without formally controlling for these characteristics, we look to understand how the identified clusters correlate to the productivity of the firms in the clusters. This is a preliminary analysis before our more formal modelling of these relationships in section 5.

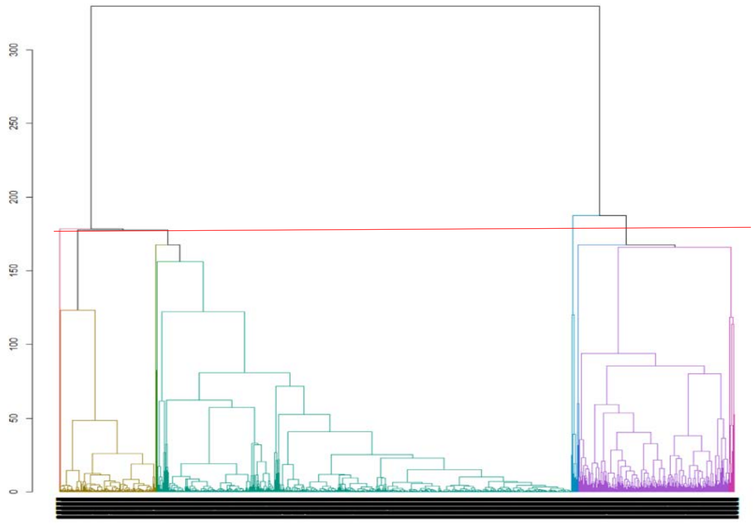
4.1. Method

To do this, we apply a hierarchical clustering algorithm to the normalised measures of innovative activities as well as the control variables. Note, we do not include the firm-level total factor productivity estimates in this analysis – instead, we choose to identify clusters of companies based on their innovative activities and then assess the productivity profile of the clusters themselves.

There are a number of different clustering algorithms that one could choose to apply in this case but broadly these fall into two main categories: 1) agglomerative – starting with every data point in its own cluster then grouping into larger and larger groups on the basis of some measure of dissimilarity between the groups; 2) divisive – starting with the dataset as a whole then splitting into smaller and smaller groups by finding groups that are most dissimilar from one another, see Boehmke and Greenwell (2020) for an introduction to this method and its implementation in R. These algorithms continue until some predetermined stopping threshold is hit. In this paper we choose to use agglomerative clustering with the measure of dissimilarity being Ward's distance calculation which aggregates clusters where the smallest between cluster variance occurs, by measuring the within cluster variance (Ward, 1963; Murtagh and Legrendre, 2014). When we ran the agglomerative clustering algorithm on the whole sample, we found that there was one small outlying group: those companies where the stock of patents was remarkably high relative to the rest of the sample. To have the best chance of identifying clearer clusters we removed this group from the sample and ran our clustering analysis again so that the sample was not being skewed by a small number of outlying companies.

This outlier-corrected sample gives rise to the cluster dendrogram seen in Figure 11, where the splits represent the formation of different clusters and the width between the splits shows the extent of the dissimilarity between the newly formed clusters. For example, in Figure 11 cutting the dendrogram at the red line would identify five clusters, with greatest dissimilarity between clusters 1 (red), 2 (yellow) & 3 (green) versus 4 (blue) & 5 (purple), numbering the branches from left to right, but a greater similarity between 4 & 5 than between 2 & 3.

Figure 11: Dendrogram

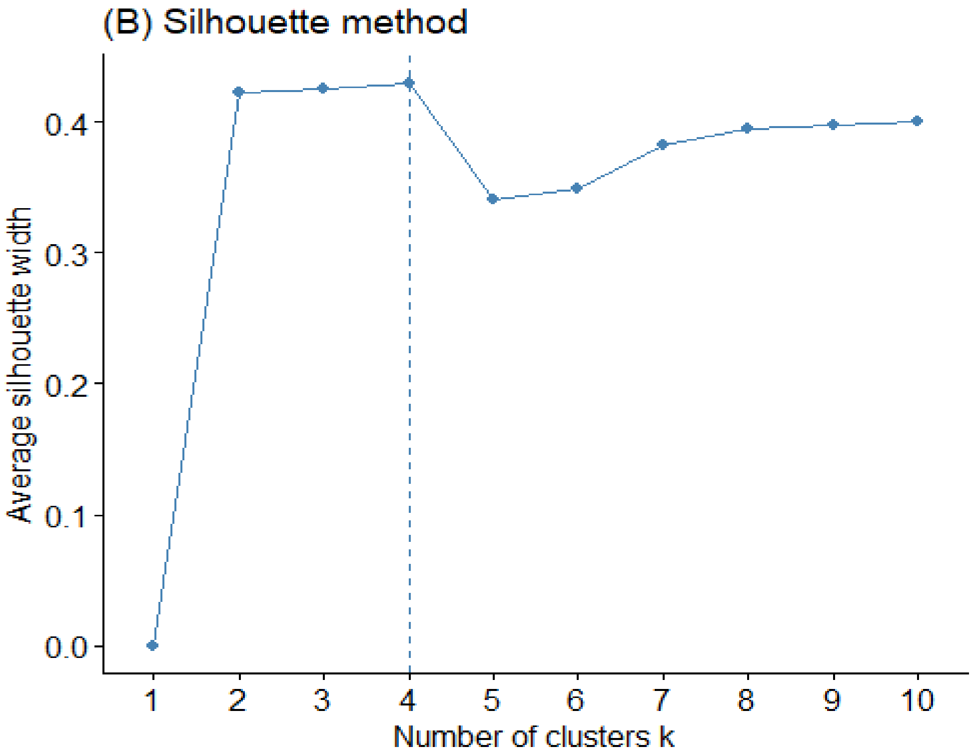


Source: OECD, Refinitiv, abrdn as of 2022

Once the dendrogram has been formulated the next natural question is where to cut the diagram i.e. how many clusters we should identify. It is possible to visually inspect the dendrogram to determine the cut off. But there are also a range of measures to assess statistically what the optimal number of clusters is - the factoextra package in R is useful to assess these (Kassambara, 2020).

We have focused on the silhouette method which measures the average silhouette width as the number of clusters increases. The silhouette width measures how similar each observation is to its own cluster relative to the other clusters and ranges for +1 to -1, where values closer to 1 imply stronger within-cluster similarity and between-cluster dissimilarity. The average value of this is calculated across the clusters as we consider an increasing number of clusters. On this basis there is a clear signal that four is the optimal number as proceeding clusters are less dissimilar to one another than the four. We therefore proceed with four clusters identified on the outlier-corrected sample and then include this group of outliers as the fifth cluster.

Figure 12: Silhouette Plot



Source: OECD, Refinitiv, abrdn as of 2022

4.2. Results

To better understand the output from the cluster analysis, we look at the average of the input data and other measures not included in the cluster analysis, like productivity, across each cluster. Figure 13 summarises the features of the five clusters from their productivity level to their firm-level characteristics and the innovative activities.

Naming the clusters according to their features, the most productive cluster we identify – the 'Patent Machines' – has the highest levels of patent activity, high intangible asset shares and middling R&D spending. This provides tentative evidence that measures of innovative activity are positively correlated to levels of productivity. We then find the 'Innovation all-rounders'. These companies that continue to perform well across a broad range of innovation measures, from R&D spending to patent activity and quality.

Even with the strongest average balance sheet position for the funding of innovative activities, the 'Midtable Mediocrity' cluster, features companies which do not score highly in terms of average patents, R&D spending, or intangible asset shares, leaving their level productivity in the middle of the pack, climbing down we find the 'Innovation Laggards'. On average these companies do not have strong innovation characteristics, nor do they have a particularly robust financial position to fund innovation, and so they sit second to last in terms of their average productivity.

Finally, we find that the 'Inefficient Innovators', despite having the highest average R&D spending of all clusters and middling performance in the patent indicators, have the lowest average productivity of any of the clusters as well as the poorest average financial position. This implies that these companies on average are not reaping the benefits of their investment in innovation. This highlights the issue that we are only including quantifiable measures of innovation and do not account for other aspects which may be more strongly related to process innovation.

Figure 13: Features of Clusters

	Total Factor Productivity	Sales per Emp	Interest Coverage	Debt to Equity	Gross Profit Margin	Intangible Assets	R&D to Emp	R&D to Assets	Patent Scope	Family Size	Backward Citations	Patent Activity	Patent Stock	Number of Employees	Share of Foreign Sales	Cluster Size
Inefficient Innovators																
Innovation Laggards																
Midtable Mediocrity																
Innovation all-rounders																
Patent Machines																

Note: mean level of each indicator is ranked from best to worst performing cluster in each metric, best = green and worst = red.

Source: OECD, Refinitiv, abrdn as of 2022

5 Identifying Innovative Companies

The results of our cluster analysis suggest there may be a positive correlation between productivity and proxies of innovation but clustering does not control for firm-level factors jointly. To do this, we build a panel model whereby we can directly control for sector and firm-level characteristics such as size, given the results of the cluster analysis show the highest-ranking cluster for productivity also features the largest and most geographically diverse companies.

5.1. Data

To have a more balanced and larger sample size for this exercise, we choose to narrow down the variables which we include in the modelling. Firstly, we choose a single measure of firm's ability to finance innovation by focusing on a firm's debt-to-equity ratio. We then narrow the range of innovation indicators to include a single measure of R&D expenditure – R&D to assets – and then include the cumulative patent portfolio, average patent quality (the average across our three quality measures in section 4) and control for firm size with the number of employees.

With only a fraction of firms issuing patents and these being highly concentrated in certain sectors, we try to control for this by constructing a variable which measures the share of firms patenting in each sector each year. We then interact this with patent quality in the final regression.

5.2. Method

To model the impact of innovation on firm productivity we use a panel model estimated with time and sector fixed effects. To account for the fact that patenting tends to be concentrated by sector we augment the regression with the patent share variable outlined in the previous section.

$$\Omega_{its} = \alpha + X\_innovation'_{its}\beta_{inn} + X\_financing'_{its}\beta_{fin} + X\_control'_{its}\beta_{cont} + \delta_s + \gamma_t + \varepsilon_{it} \quad (3)$$

$$\varepsilon_{it} = \lambda_i + \eta_{it} \quad (4)$$

$$E(X_{it}\eta_{it}) = 0 \quad (5)$$

$$E(X_{it}\lambda_i) = 0 \quad (6)$$

Here, total factor productivity, Ω_{its}, for firm i in sector s at time t is predicted using three sets of regressors – innovation, financial and controls - which vary over time and across firms. The model includes sector fixed effects, δ_s, and time (within) effects, γ_t (Wooldridge, 2010). Note that the sector effects will influence the estimated constant, α, across companies in the model and can be estimated using a series of dummies which identify each sector. The time effect will shift the constant across time periods and can be estimated with a series of dummies which identify each period.

In the basic panel model, it is assumed that rather than a composite error across time, t, and firm, i, the error takes the form of an idiosyncratic component, η_{it}, which varies across both time and firms, as well as an unobserved non-time varying component, λ_i, which captures deviations from the mean at the firm level. The benefit of this structure versus the assumption made in simpler pooled ordinary least squares (OLS), where only the composite error is considered, is that we can directly account for, and model the unobserved heterogeneity at the firm level within the data.

However, doing so comes at the cost of the introduction of bias into the estimation with the likelihood that the moment condition assumed in equation (6) is unlikely to hold. This is because assuming that firm-level effects are uncorrelated with the regressors may be problematic.

To overcome this, we use the fixed effects estimator. In this, sources of unobserved heterogeneity, λ_i, and all time invariant omitted variables are removed by applying the within transformation in which both the dependent and independent variables are demeaned prior to estimation with OLS. For unbiased estimation with the fixed effects estimator, the idiosyncratic error must be uncorrelated with the regressors as outlined in equation (8).

$$\tilde{Y}_{it} = \tilde{X}_{it}'\beta + \tilde{\lambda}_i + \tilde{\eta}_{it} \tag{7}$$

Where $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$, $\tilde{X}_{it} = X_{it} - \bar{X}_i$, $\tilde{\lambda}_i = \lambda_i - \bar{\lambda}_i = 0$, $\tilde{\eta}_{it} = \eta_{it} - \bar{\eta}_i$ with bars indicating the group mean.

$$E(X_{it}\eta_{it}) = 0 \tag{8}$$

5.3. Results

The regression results require careful interpretation, but some are particularly encouraging.

We found that R&D to assets positively influenced firm-level productivity and that this was a highly significant result. This is in line with the earlier cluster analysis findings around R&D and productivity and suggests that those firms undertaking R&D on average have higher levels of productivity. There is however a lack of high-quality data tracking the R&D activities of companies given this level of detailed reporting is not required by accounting standards. This could mean that more nuanced understanding of the different sectoral impacts of R&D is not well captured as some sectors are more likely to report this type of spending than others.

Higher intangible asset shares are also a statistically significant positive driver of levels of productivity. This suggests that human capital investment via training, software, advertising, and product/process innovation is an important driver of productivity. Again, this line may well be underreported in company accounts, but for those where this is reported we find this relationship holds. It also important to acknowledge that quantifying the value of intangibles is difficult, therefore we are conscious not to overstate these findings.

There is a clear positive and statistically significant relationship between patenting activity, as captured by the patent portfolio of a firm, and the level of productivity. This is an encouraging result given that patent activity is an easily quantifiable activity that can be tracked overtime for firms.

Table 3: Drivers of total factor productivity

	Log TFP
Intercept	1.396 ***
	(0.039)
Debt to Equity	-0.0920
	(0.051)
R&D to Assets	0.003 ***
	(0.000)
Intangible Asset Share	0.426 ***
	(0.025)
Patent Portfolio	0.059 ***
	(0.003)
Patent Quality	-1.175 ***
	(0.180)
Patent Quality * Pat Share	2.823 ***
	(0.415)
Pat Share	0.355 **
	(0.125)
Employees	0.002 ***
	(0.000)
Number of Observations	33652.000
R Squared	0.227

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

regression contains sector and time fixed effects

Assessing the influence of patent quality on productivity proved to be challenging. To quantify patent quality, we averaged across the three patent quality measures – patent scope, patent family size and backward citations. We then interacted this with the patent scope variable to account for differing degrees of patent activity across sectors. This means that when interpreting the influence of patent quality, the extent of patenting in the broader sector matters. The positive and significant coefficient on the interaction term contrasts with the negative significant coefficient on the patent quality variable alone. This implies that the positive effect of higher quality patenting is amplified as more firms in the sector are patenting, whilst the reverse is true for those sectors where patenting is rare. This is interesting and suggests that the competition within high patenting sectors leads to higher quality patents having a higher overall impact on the productivity of the firm.

Table 4: Contribution of patent quality to productivity by level of patent quality and share of firms patenting

		Share of Firms Patenting										
		0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Average Patent Quality	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.05	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1
	0.1	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.2
	0.15	-0.2	-0.1	-0.1	-0.1	0.0	0.0	0.1	0.1	0.2	0.2	0.3
	0.2	-0.2	-0.2	-0.1	-0.1	0.0	0.0	0.1	0.2	0.2	0.3	0.3
	0.25	-0.3	-0.2	-0.2	-0.1	0.0	0.1	0.1	0.2	0.3	0.3	0.4
	0.3	-0.4	-0.3	-0.2	-0.1	0.0	0.1	0.2	0.2	0.3	0.4	0.5
	0.35	-0.4	-0.3	-0.2	-0.1	0.0	0.1	0.2	0.3	0.4	0.5	0.6
	0.4	-0.5	-0.4	-0.3	-0.1	0.0	0.1	0.2	0.3	0.4	0.6	0.7
	0.45	-0.5	-0.4	-0.3	-0.2	0.0	0.1	0.2	0.4	0.5	0.6	0.8
	0.5	-0.6	-0.5	-0.3	-0.2	0.0	0.1	0.3	0.4	0.5	0.7	0.8
	0.55	-0.7	-0.5	-0.3	-0.2	0.0	0.1	0.3	0.4	0.6	0.8	0.9
	0.6	-0.7	-0.6	-0.4	-0.2	0.0	0.1	0.3	0.5	0.7	0.8	1.0
	0.65	-0.8	-0.6	-0.4	-0.2	0.0	0.1	0.3	0.5	0.7	0.9	1.1
	0.7	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0	1.2
	0.75	-0.9	-0.7	-0.5	-0.3	0.0	0.2	0.4	0.6	0.8	1.0	1.3

Source: OECD and abrdn as of 2022

If we consider the financial position of companies, measured by the company's debt-to-equity ratio, this is found to be borderline significant with a small negative coefficient. The intuition here is that companies that have a higher debt burden, or who are using debt financing to a greater degree, are less likely to be able to fund activities which would boost their overall level of productivity. But debt financing may not be outright negative for productivity; a firm which uses the raised capital to increase its efficiency may well boost its productivity but still have a higher debt-to-equity ratio. Therefore, the result is not a strong one, as it is only significant at the 10% level.

Finally, it is worth noting that the number of employees that a firm has, a measure of firm size, is highly significant and positively related to the firm's productivity. This suggests that there are positive spillovers from economies of scale in which large firms on average have a higher level of productivity. In addition to this, we find that both the time and sector fixed effects are jointly significant.

To try and assess the impact of process innovation through governance and management structures, which can also influence the productivity of firms (Bartz-Zuccala, et al., 2018; Kremp & Mairesse, 2004), we augment the above regression with governance scores from Refinitiv (2022). These scores capture management structure and compensation as well as shareholders rights and corporate social responsibility policies. The measures of management structure are a more appropriate proxy for process innovation, but we test the importance of aggregate scores regardless. With these scores only available for a smaller subset of our sample (700 companies) , we choose to use the larger regression for the next stages of our analysis. But we do find that governance is a positive and statistically significant driver of productivity.

Table 5: Productivity regression augmented with governance scores

	Log TFP
Intercept	1.635 ***
	(0.038)
Debt to Equity	0.05
	(0.216)
R&D to Assets	-2.300 ***
	(0.157)
Intangible Asset Share	0.462 ***
	(0.030)
Patent Portfolio	0.032 ***
	(0.002)
Patent Quality	0.988 ***
	(0.179)
Patent Quality * Pat Share	-1.786 ***
	(0.488)
Pat Share	0.675 ***
	(0.152)
Governance Score	0.266 ***
	(0.026)
Employees	0.000 ***
	(0.000)
Number of Observations	12042
R Squared	0.382

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.
regression contains sector and time fixed effects

5.4. Forming an Innovation Index

To rank companies' relative effective innovation efforts, we use the fitted contribution of the innovation measures to firm-level total factor productivity and rank companies each year based on this. A higher rank implies a stronger contribution of innovative activities to total factor productivity.

$$InnIndex_{ist} = \frac{P_{ist}}{100(N+1)} \tag{9}$$

In equation 9, the ranking of firm i, in sector s, at time t, is given by the company's percentile rank, P_{ist}, in the distribution of contributions to productivity from the innovation indicators, $\widehat{\Omega \mathrm{inn}_{ist}}$, i.e. the fitted value for productivity when concerning innovative activities alone, relative to the N number of scores.

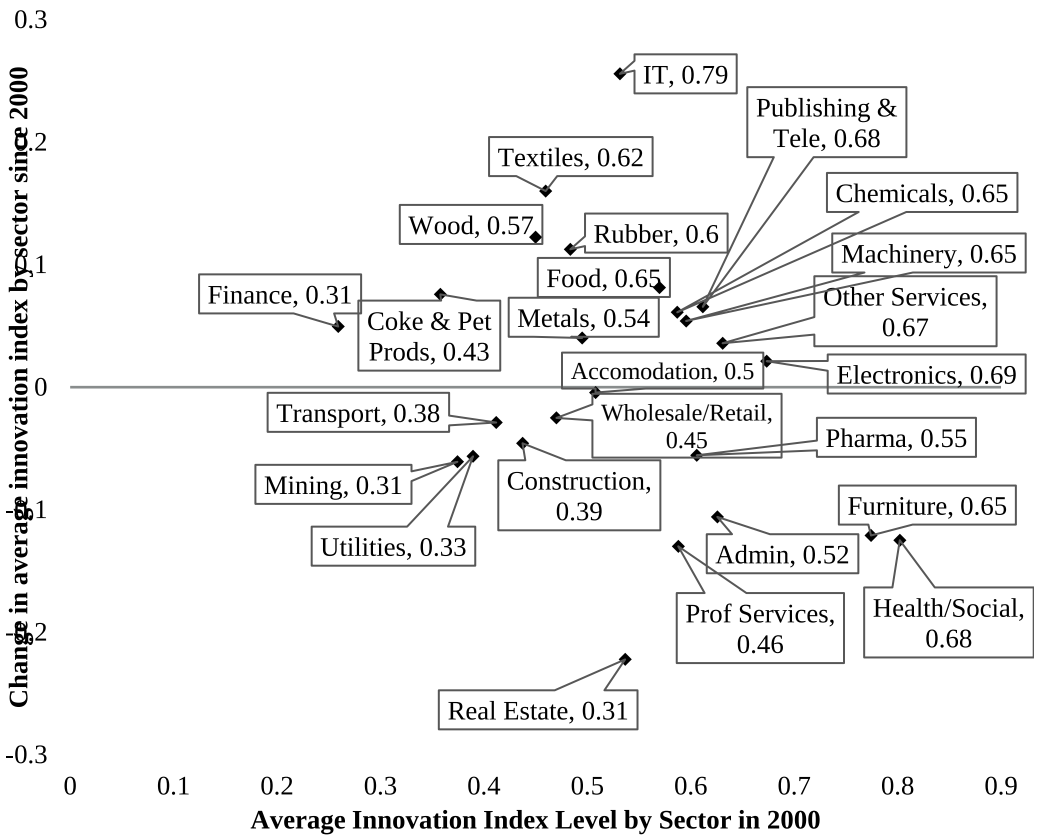
We also construct a similar index to rank the different companies based on their ability to fund innovative activities by ranking the companies each year based on the contribution of debt to equity to a firm's productivity.

$$FinIndex_{ist} = \frac{P_{ist}}{100(N+1)} \tag{10}$$

In equation 10, the ranking of firm i, in sector s, at time t, is given by the company's percentile rank, P_{ist}, in the distribution of contributions to productivity from the indicators for financing innovation, $\widehat{\Omega \mathrm{fin}_{ist}}$, i.e. the fitted value for productivity when concerning financing activities alone, relative to the N number of scores.

Figure 14 shows the average innovation index score by sector in 2000 relative to the change in this average since 2000 for each sector and the labels show the average value in 2019. The most innovative sectors in 2019 are the tech, electronics, publishing & telecom, and health& social sectors. Many of these come out on top because our measures of innovation focus on patenting activity and R&D, both of which would be classically associated with these industries. The health & social sector ranking so highly may seem surprising, but it includes several medical testing firms and some companies that are closely related to the pharmaceutical sector, which have high rates of investment in intangible assets and patenting activity. The sector however has fallen down the average rankings since 2000 and is clearly eclipsed by the strong rise in US technical innovation through the past two decades, with information technology (IT) moving most convincingly up the index.

Figure 14: Average innovation index level and change by sector



Source: OECD, Refinitiv, abrdn, as of 2022

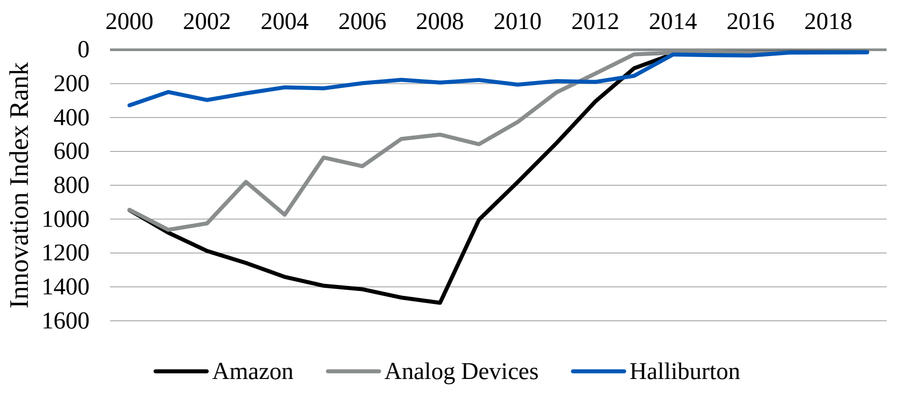
Table 6 provides some summary statistics of those companies which sat at the top of the innovation index in 2019. Even after controlling for sector fixed effects in our regression analysis, there is still a high concentration of the top companies in the electronics and the publishing & telecoms sectors. Note here that the publishing sector includes software publishers and therefore captures some of the large tech companies which have risen to become dominant players in the US equity market. The other interesting feature of the firms at the top are the drivers that put them there. For most, it is the size of their patent portfolio which pushes them to the top of the index. Whilst this may be a feature of the accumulation of portfolios over a longer period these companies continue to expand their patent libraries at a faster rates than their peers. Some of the companies at the top are there because of intangible investment shares but patenting activity stands out as the key driver.

Table 6: Top 20 in 2019 Innovation Index

Rank	Company	Sector	Main Driver	Rank 5 Years Ago	Rank 10 Years Ago
1	Microsoft Corp	Publishing & Tele	Patent Portfolio	1	2
2	QUALCOMM Inc	Electronics	Patent Portfolio	2	4
3	Intel Corp	Electronics	Patent Portfolio	3	3
4	Apple Inc	Electronics	Patent Portfolio	5	12
5	AT&T Inc	Publishing & Tele	Patent Portfolio	7	7
6	Medtronic PLC	Electronics	Patent Portfolio	8	5
7	Texas Instruments Inc	Electronics	Patent Portfolio	6	8
8	Broadcom Inc	Electronics	Patent Portfolio	4	6
9	Boston Scientific Corp	Electronics	Patent Portfolio	9	9
10	Corning Inc	Metals	Patent Portfolio	11	18
11	Applied Materials Inc	Electronics	Patent Portfolio	10	10
12	Oracle Corp	Publishing & Tele	Patent Portfolio	12	11
13	Amazon.com Inc	Wholesale/Retail	Patent Portfolio	22	752
14	Halliburton Co	Mining	Patent Portfolio	18	406
15	Analog Devices Inc	Electronics	Intangible	28	213
16	Eaton Corp PLC	Electronics	Patent Portfolio	13	16
17	Pfizer Inc	Pharma	Patent Portfolio	15	14
18	Northrop Grumman Corp	Electronics	Patent Portfolio	20	20
19	Illinois Tool Works Inc	Machinery	Patent Portfolio	17	21
20	Stanley Black & Decker Inc	Electronics	Intangible	14	13

Looking at the journey of these companies, in the last two columns of Table 6, the ranks have been stable since 2014 but we can find movement if we look further back. Amazon, Analog Devices and Halliburton are the ones that improved the most over the past 10 years when considering the top 20 companies today. Figure 15 shows the rapid rise of Amazon through the index rankings after the financial crisis, this is in sharp contrast to the slow, gradual rise of Halliburton over the past two decades or the above-average rankings of Analog Devices through the earlier part of the sample and then its upward shift after 2014.

Figure 15: Change in rank over time for Amazon, Analog Devices and Haliburton



Source: OECD, Refinitiv, abrdn, as of 2022

Now that we have a measure that ranks companies on their relative stance of innovation, a way to understand companies' ability to finance this, and track both overtime, we are interested in understanding how this translates into company performance.

6 Innovation, Productivity and Firm Performance

Understanding innovation and the relation to productivity is critical but we are particularly interested in how this then feeds into the performance of companies and whether those companies, which are highly innovative, generate higher returns than those that perform poorly on these metrics.

To do this, we take the innovation index ($InnIndex_{its}$) and financing index ($FinIndex_{its}$), which are outlined in section 5.2, and regress these on the annual return of each company (r_{its}), see equation 11. To account for style biases in markets through our sample period we use a dummy variable, $Style_{its}$, where 1 indicates a 'growth' company and 0 indicates a 'value' company. We use the Russell value and growth indices to allocate companies to either style depending on which index they feature. This helps to control for the effect of style biases, but the ideal metric would be continuous rather than binary to account for different degrees of value or growth characteristics.

In this regression we also include the number of employees, to account for firm size, and time and sector effects to control for sector and cyclical trends in firm returns. The fixed effects estimator is used, with the moment conditions as outlined in equations 12, 13 and 14.

$$\mathrm{r}_{its} = \alpha + InnIndex'_{its}\beta_{inn} + FinIndex'_{its}\beta_{fin} + Style'_{its}\beta_{style} + X\_control'_{its}\beta_{cont} + \delta_s + \gamma_t + \varepsilon_{it} \quad (11)$$

$$\varepsilon_{it} = \lambda_i + \eta_{it} \quad (12)$$

$$E(X_{it}\eta_{it}) = 0 \quad (13)$$

$$E(X_{it}\lambda_i) = 0 \quad (14)$$

6.1 Results

Table 7 is a summary of the regression output of equation 7. Firstly, the time and sector fixed effects are also significant in this regression as well as the control for firm size, number of employees, which is significant and negatively related to returns. Growth companies are found to have higher returns on average in our sample, with this result being statistically significant. Encouragingly, we also find that both the innovation index and the index of financing innovation are statistically significant and positively associated with returns. This means that firms that have a higher contribution from innovation factors to their productivity on average have higher returns.

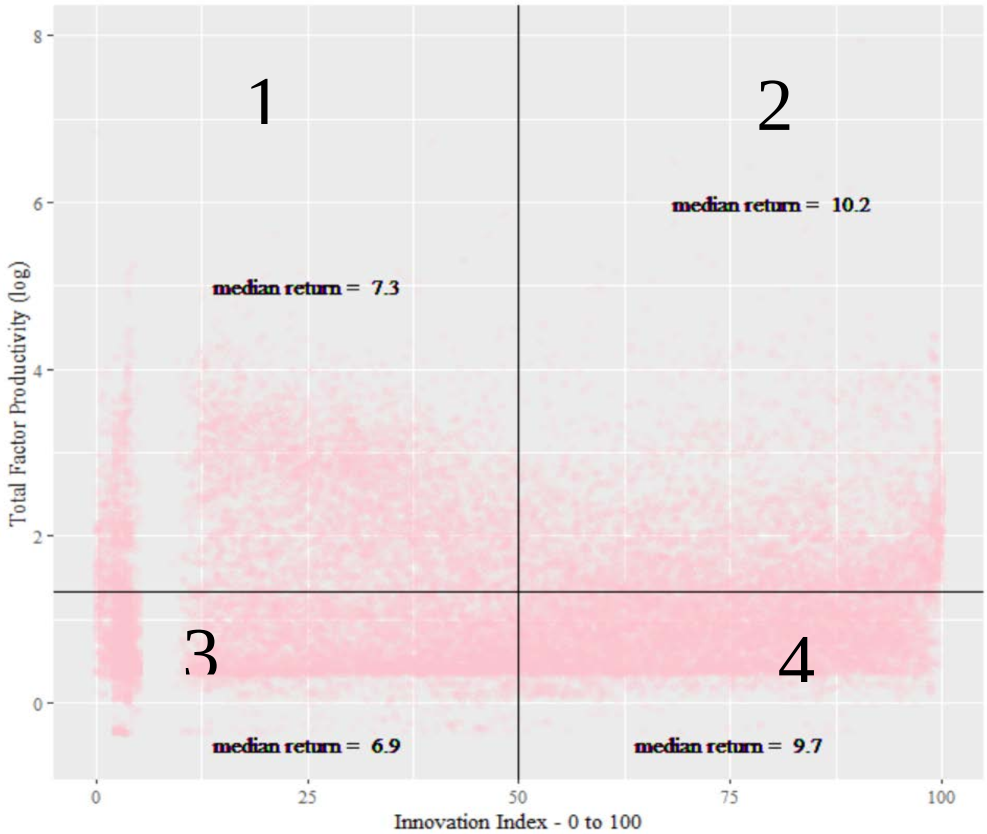
Table 7: Drivers of company returns

	Returns
Intercept	8.336 ***
	(1.768)
Finance Index	4.113 ***
	(0.914)
Innovation Index	2.257 *
	(0.941)
Employees	-0.010 **
	(0.003)
Style	4.786 ***
	(0.485)
Number of Observations	25756
R Squared	0.217

*** p < 0.001; ** p < 0.01; * p < 0.05.
regression contains sector and time fixed effects

Figure 16 visualises the relationship between the innovation index, a firm's level of total factor productivity, and returns. The plot is split into four quadrants by cutting the sample between those which have above- or below-average levels of productivity and sit at the top of the innovation index, above 50, or at the bottom, below 50. We then cross-check to see how companies in each quadrant perform by calculating the median return of the companies in each one. Encouragingly, the highest median return can be found in the top right quadrant (labelled 2 in the diagram) where productivity and innovation are the highest. The second highest median return is in the bottom right quadrant which features companies that perform well on the innovation measures but still have below-average productivity.

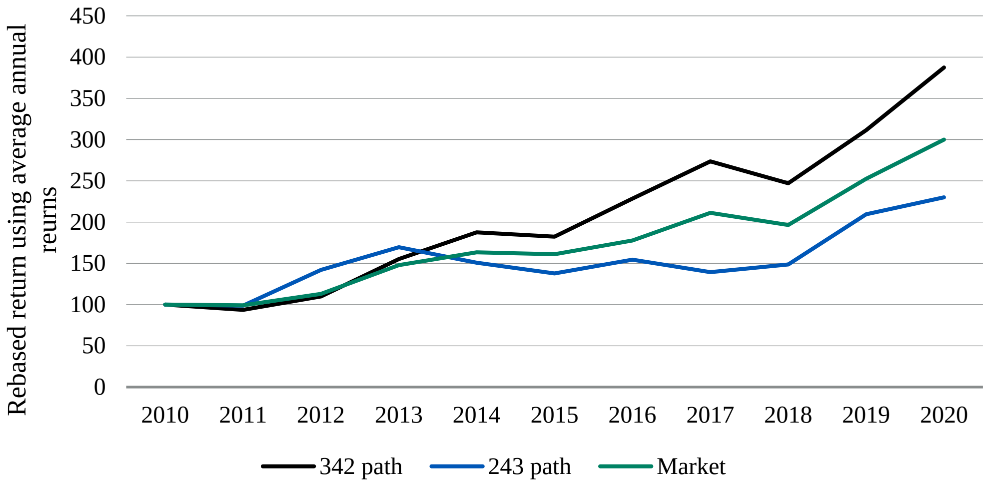
Figure 16: Innovation, productivity and company returns



Source: OECD, Refinitiv, abrdn, as of 2022

If we track companies' journeys through this plot, an interesting pathway is three to four to two. This sees companies go from quadrant 3 – below average productivity, below average innovation – to quadrant 4 – improving innovation, not yet realised in productivity – to quadrant 2, the realisation of higher innovation into higher productivity. Another interesting pathway is the reverse one: those companies that start off in quadrant 2 and then overtime become less innovative and less productive ending in quadrant 3. If we compare these two pathways to the performance of the aggregate Russell 3000 Index from 2010 onwards, as in Figure 17, there is a clear outperformance in the three, four, two companies and an underperformance in the two, four, three companies. This emphasises the importance of not just becoming more innovative but maintaining the pace of innovation when it comes to obtaining superior returns. An extension of this work could formally model the transitions between these quadrants using a multinomial logit framework, but this is beyond the scope of this paper.

Figure 17: 342 pathway delivers superior returns



Source: OECD, Refinitiv, abrdn, as of 2022

7 Conclusions

The aim of this paper is to study the interaction between innovation, firm productivity and returns for US-listed companies. We found that the trends in US-listed firms' productivity matched those of the aggregate economy over the past two decades, but also that there is a tail of highly productive firms across sectors rather than highly concentrated within a narrow number of sectors.

With a clear understanding of the productivity dynamics, we then found evidence that clusters of firms exist with shared innovative features and that the clusters of firms displaying these characteristics tend to be more productive on average than the rest of the sample. This was further evidenced when we modelled the impact of innovation activities and the funding ability for innovation - whilst controlling for firm size, sector, and time effects – and found that innovative activities are a positive and statistically significant driver of firm productivity.

The creation of an innovation index, which systematically ranks listed US companies across time based upon the share of productivity explained by innovation measures, allows us to understand the rapid rise we have seen in certain US stocks since the 2000s and consider how we might identify the rise of companies in the future. We assess the explanatory power of this index by regressing the index upon firm returns and find that a higher ranking in the innovation index is associated with strong firm performance on average, even after controlling for style biases in the market through our sample. A key and critical finding is that, whilst some firms have managed to rise through the innovation index and maintain their strong innovation position, others have fallen back through the index and their returns have suffered relative to the broader market. Therefore, it is not sufficient for companies to have a short-term boost in their innovative activities to guarantee outperformance, –instead firms need to continually innovate to remain at the top of the pack.

8 References

Andrews, D., Criscuolo, C. & Gal, P. N., 2015. Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries, Paris: OECD.

Bartz-Zuccala, W., Mohnen, P. & Schweiger, H., 2018. The Role of Innovation and Management Practices in Determining Firm Productivity. Comparative Economic Stuies, 60(4), pp. 502-530.

Benavente H., J. M. (2006). The role of research and innovation in promoting productivity in Chile, Economics of Innovation and New Technology 15: 301-315.

Boehmke, B. & Greenwell, B., 2020. Hands-On Machine Learning with R. 1 ed. Boca Raton: Taylor & Francis.

Boldrin, M., and Levine., D.K., 2013. "The Case against Patents." Journal of Economic Perspectives, 27 (1): 3-22.

Crepon, B., Duguet, E. and Mairesse, J., 1998. Research, Innovation and Productivity: An Econometric Analysis at the Firm Level. Economics of Innovation and New Technology, 7(2), pp. 115-158.

Gal, P. N., 2013. Measuring Total Factor Productivity at the Firm Level Using OECD-ORBIS. OECD Economics Department Working Papers No. 1049.

Griffith, R., E. Huergo, J. Mairesse, and B. Peters, 2006. Innovation and Productivity in Four European Countries, Oxford Review of Economic Policy 22: 483-498.

Griliches, Z., 1990. Patent Statistics as Economic Indicators: A Survey. Journal of Economic Literature, 28, pp. 1661-707.

Hall, B. H., 2011. Innovation and Productivity. NBER Working Paper Series, Issue 17178.

Hall, B.H., and Haroff, D., 2012. Recent Research on the Economics of Patents. NBER Working Paper Series, Issue 17773.

Hulten, C. R., 2001. Total Factor Productivity: A Short Biography. New Developments in Productivity Analysis. National Bureau of Economic Research , pp. 1-54.

Janz, N., H. Loof, and B. Peters (2003). Firm Level Innovation and Productivity – Is there a Common Story Across Countries? Mannheim, Germany: ZEW Discussion Paper No. 03-26.

Jefferson, G. H., H. Bai, X. Guan, and X. Yu (2006). R&D Performance in Chinese industry, Economics of Innovation and New Technology 15: 345-366.

Kassambara, A., 2020. R Documentation. [Online] Available at: www.rdocumentation.org/packages/factoextra/versions/1.0.7 [Accessed 1 August 2022].

Kremp, E. & Mairesse, J., 2004. Knowledge Management, Innovation and Productivity: A Firm Level Exploration Based on French Manufacturing CIS3 Data. NBER Working Papers, Issue 10237.

Levenshtein, V.I., 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady. 10 (8): 707–710.

Levinsohn, J. & Petrin, A., 2003. Estimating Production Functions Using Inputs to Control for Unobservables. Review of Economic Studies, 70(2), pp. 317-341.

Marrocu, E., Paci, R. & Pontis, M., 2009. Intangible Capital and Firms Productivity. Centre for North South Economic Research, Volume 16.

Murtagh, F. & Legendre, P., 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? Journal of Classification, 31, pp. 274-295.

OECD/Eurostat., 2018, Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris/Eurostat, Luxembourg. https://doi.org/10.1787/9789264304604-en

OECD, 2019. Patents by main technology and by International Patent Classification (IPC). OECD Patent Statistics (database), https://doi.org/10.1787/data-00508-en

OECD, 2022. STAN Industry ISIC Rev. 4. STAN: OECD Structural Analysis Statistics (database). https://doi.org/10.1787/data-00649-en

Olley, G. S. & Pakes, A., 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. Econometrica, 64(6), pp. 1263-1297.

Ornaghi, C. & Van Beveren, I., 2011. Using Proxy Variables to Control for Unobservables When Estimating Productivity: A Sensitivity Analysis. Available at SSRN: https://ssrn.com/abstract=1984573 or http://dx.doi.org/10.2139/ssrn.1984573

Refinitiv, 2020. Refinitiv Worldscope Fundamentals Database , s.l.: Refnitiv.

Refinitiv, 2022. Refinitiv ESG Scores Methodology. Available at: https://www.refinitiv.com/en/sustainable-finance/esg-scores#t-score-range

Solow, R. M., 1957. Technological Change and the Aggregate Production Function. Review of Economics and Statistics, 39, pp. 312-320.

Squicciarini, M., Dernis, H. & Criscuolo, C., 2013. Measuring Patent Quality: Indicators of Technological and Economic Value. OECD Science, Technology and Industry Working Papers, Issue 03.

Sprague, S., 2021. The U.S. Productivity Slowdown: an Economy-Wide and Industry-Level Analysis. BLS Monthly Labour Review. Bureau of Labour Statistics. Available at: https://www.bls.gov/opub/mlr/2021/article/the-us-productivity-slowdown-the-economy-wide-and-industry-level-analysis.htm#:~:text=The%20U.S.%20productivity%20slowdown%3A%20an%20economy-wide%20and%20industry-level,growth%20of%20the%20late%201990s%20and%20early%202000s.

United Nations, 2008. International Standard Industrial Classification of All Economic Activities – Revision 4. Department of Economic and Social Affairs Statistical Division. Statistical Papers. Series M, No. 4, Rev.4. Available at: https://unstats.un.org/unsd/demographic-social/census/documents/isic_rev4.pdf

Van Ark,B., and Monnikhof, E., 1996. Size Distribution of Output and Employment: A Data Set for Manufacturing Industries in Five OECD Countries 1960-1990. OECD Economics Department Working Paper, No. 166.

Van Leeuwen, G., and L. Klomp (2006). On the contribution of innovation to multi-factor productivity growth, Economics of Innovation and New Technology 15: 367-390.

Ward, J., 1963. Hierarchical Grouping to Optimize an Objective Function. Journal of American Statistical Association, Issue 58, pp. 236-244.

Wooldridge, M. J., 2009. On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables. Economics Letters, 104(3), pp. 112-114.

Wooldridge, J. M., 2010. Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge : MIT Press.