Firm R&D Inertia

Connie Lee
CERNA, Mines Paristech, PSL University
connie.lee@mines-paristech.fr

Get Latest Version Here

September 2020

Abstract

Firm R&D decisions are likely to have lasting consequences. Here I will document patterns of firm technological position over its life cycle. Using patent data, I build a measure to compare the similarity between an innovative firm's technological contents over time with its technological position when it enters. I find that new entrants are likely to continue patenting in areas similar to their initial invention for multiple years – they exhibit inertia. I then describe how the degree of inertia is affected by initial conditions. The heterogeneous firm size distribution is also explored as is the effect of technology sector concentration.

JEL classification: O31, O33, D22
Keywords: innovation, technological change, firm behavior

1 Introduction

Although entire disciplines in economics have arisen to study firm behavior, there are not many studies that have empirically focused on the technological dimension of firms. Firms are usually characterized by their product and any effects associated with technological changes tend to be grouped into the marginal costs of the product or as an increase in quality or product variety. However, as a firm grows its product line, it also admits a parallel process where it builds up a technological position with associated human and physical capital. Intuitively, this build up of expertise in a technological area will influence the future decisions the firm makes. In fact, one can expect that any decisions firms make will have some persistent effects.

Here I will explore the extent to which these inertial tendencies exist in the development of a firm's technological position, which is what I call R&D inertia. I will focus in particular on the early stages of a new entrant's life cycle to test for the impact of initial conditions and choices made by the entrant, namely its initial originality and whether it has past patenting experience. Using patent data I measure the proximity of the entrant's technological position to its entry position and estimate the persistence of this measure over firm age. I also address the effects of firm size on the firm's technological choices which leads into a discussion on the effects of competition. The overall objective of this study is to better understand how new entrants develop, how the initial conditions affect their development and how they interact with the other firms in their industry.

What I call the technological content of the firm is simply the part of the know-how and capital that is involved in the production and R&D projects in the firm. The position is inferred from the composition of people with different expertises and specialized investments that the firm has made. Technology can enter during the production process of the product or it can compose of a part of the product. Even though the technological content is intrinsically linked with a firm's products. The position of a firm in product space (which the literature often studies with industry classifications) cannot be directly translated into a position in technological space. This is because products are usually characterized by the elements that consumers value and this does not necessarily align with the sophistication of the technological input used to develop and produce the product. Bloom et al. (2013) further discuss the differences between these spaces. Nonetheless there are cases where patents can be directly connected to products. For instance, one will sometimes find a product with a label that says 'patent pending'. Virtual patent markings introduced in the US in 2011 also specify the patents used in products directly. See Argente et al. (2020) for an in depth study of this association.

My main hypothesis is that firms should exhibit a tendency to grow around the initial position(s) it takes on. This could be due to frictions or sunk costs. Frictions can come in numerous forms, for example, labor contracts might have costs associated with firing employees. In addition, the hiring process can be long and constitute a type of sunk cost. Another sunk cost can simply be an investment in equipment to build the product or the effort put into establishing a network of suppliers. Contracts with external entities are a source of friction as well.

The different fixed costs and frictions entrench the firm in its initial position. Without considering consumers' demand imposed on the product, the different ways the firm has invested in developing its first product will naturally impose a direction for future R&D. This is what I want to document. The concept that firm technological development is path dependent is intuitively accepted however it has not been studied much in the literature. In microeconomics, studies do not usually look past one period in time. In macroeconomics, studies often abstract the different product lines by aggregating them into firm level sales or growth or they model product lines that are independent of one another (see Klette and Kortum (2004), Akcigit and Kerr (2018), etc.). These common approaches fail to account for firm inertia in technological space.

I will attempt to capture this inert quality to firm technological life cycles through patent data. Patents are a type of innovation output that allow me to solely focus on the technological content of firms. A patent application is filed to protect a new technological finding. Although it is written up to define the technological content, different firms and people may write about the same thing in different ways. In order to classify filings more systematically, technology codes were introduced to label them. I exploit these technology codes to understand a firm's technological position. To measure a firm's degree of inertia I build the proximity measure developed in Jaffe (1987) over the firm life cycle instead of between two firms. This measure compares the technology codes of a firm's new patent filings each year with the patents that were filed in its first year. A high proximity implies that the firm has not moved much from its initial position which I interpret as a measure of inertia.

My first central finding is that we clearly see firms are patenting closer to their initial position in the beginning of their life cycle and that this measure declines over the firm's life cycle. This is robust to different proximity measures and time periods.

It is then interesting to consider what initial conditions can affect a firm's inertia. Lee (2020) suggested that entrants have become less original in their innovations. Hence, I explore the consequences of the initial originality on the future technological development of the entrant. how does their initial originality affect the technological proximity of future patents? Furthermore, do firms that start with a low initial originality continue to file patents that are low in originality? One might assume that a firm will want to move away from a low originality innovation, however I find evidence that the low originality entrants will on average actually double down on their position and patent in closer proximity than a firm who enters with a high originality. This corresponds with a slow increase in the originality of the firm over time.

Another firm initial condition that has recently been explored in the literature is the founding team (see Choi et al. (2019) and Gompers et al. (2010)). I address this by identifying entrants with team members that have previous experience patenting. With patent data, I can identify the inventors and applicants of the patents and deduct to some degree of accuracy who the initial researchers/engineers are. Moreover, I can find the patenting history of the inventors listed and determine whether they have prior experience patenting. Leveraging this data, I identify whether a new firm entrant has at least one inventor on the team who has experience.

Shifting away from the focus on initial conditions, I also address whether firm size has an effect on inertia. The literature has largely conflated young firms with small firms and large firms with old firms. Although many young firms are small and many old firms are large, not all young firms are small and not all old firms are large. Schumpeter was one of the first economists to explore this topic and his ideas are still referenced in the economics literature. His early work developed the theory centered around young and small firms as the innovative disruptors through the mechanism he called creative destruction. However, there is room for debate given that his later work suggests that larger firms are the main innovators in an economy due to their access to more resources. Here I attempt to disentangle the age and size effects to get a clearer picture. I find that large firms were more inert. And that they surprisingly remain at a higher degree of inertia than the smaller firms even among old firms. This suggests that to grow in size, their R&D efforts were more focused around their initial position than smaller firms. However, this does not translate directly into a conclusion on originality. I find that firm maximum originality is on average higher for large firms however it is not significantly different when we look at average originality.

A better understanding of how firms evolve in terms of their technological content will also help us understand the industry growth dynamics. The innovation output from technological choices can affect firm growth in many ways. By studying the firm life cycle, we also get a better understanding of how young and old firms contribute to the economy differently. In particular, understanding how new firms develop will help us understand the competitive dynamics of an industry. Do older firms have reason to consider new entrants as future competition? Which new firms are the threats?

The literature on competition in economics has long theorized that a motivating incentive for incumbents to innovate is to insulate against the threat of new entrants.

This question motivates my analysis to examine technological sector concentration at the time a firm enters. Like industries, I show that technological fields have their own life cycle dynamics and I suggest that they are also subject to competition. For instance, this would explain why some firms choose to not patent their inventions and keep them as trade secrets. The data shows that technological sectors follow a similar life cycle as industries with a U shaped dynamic of concentration over time.

The evolution of a firm's technological position will also help me better understand competitive dynamics overall. We can reasonably assume that firms that are further apart in technology space are competing less with each other. This has been expounded on in depth in the literature on monopolistic competition in product space (see Hotelling, Salop, and subsequent papers. It is also recognized in the antitrust literature, Shapiro (2012)) however I suggest it is also true for technology space. A difference is that the reaction to competition in product space often means reducing the price on the product - in technology space, I expect it to take another form. It might be to file a patent faster in order to expropriate some space from the competitor, or it may mean changing the technological position to have less competition. I will show some evidence of both reactions.

Overall, the existence of firm inertia means each decision in product research and development has lasting consequences. This implies that initial conditions are especially important for dictating the technical trajectory. By following firms over their life cycle we also get to see how their patenting behavior changes over time. This sheds light on the different roles of young and old firms in the economy and how they affect competitive dynamics.

The rest of the chapter proceeds as follows. Section 2 provides a literature review on the studies concerned with firm path dependency as well as a discussion of the literature on competition dynamics. Section 3 describes the dataset and sample construction choices. Section 4 discusses my econometric approach while Section 5 presents the main results. Section 6 concludes.

2 Literature Review

The concept of firm inertia has not entered the mainstream economics discussion, however it has been examined in the organization and strategic management literature. Here I will briefly review that literature. However, there are few studies that directly address inertia, instead the discussion has primarily turned to the concept of balancing exploration and exploitation within organizations. This in turn resembles some concepts in the innovation literature such as incremental versus radical innovation and product versus process innovation which I will also briefly summarize.

Hannan and Freeman (1984), influenced by ecological theories of natural selection, pose the question of what favors the selection (a.k.a. survival) of firms. They argue that stability in products, processes and policies favor selection and that therefore “high levels of structural inertia in organizations can be explained by selection in ecological-evolutionary processes.” However they also discuss the consequences of excessive inertia. In particular, they note how the investments made in specific types of physical and human capital alongside the public appeal garnered for a specific product or service greatly limits the firm's options for transformation.

Casamatta and Guembel (2010) is a more recent paper that builds on the notion discussed in Hannan and Freeman. They propose that manager incentives are a cause of inertia. With a theoretical model, Casamatta and Guembel show that when managers are concerned about their legacy, they can entrench a firm on one path and generate inertia. Another study that addresses firm inertia is Ruckes and Ronde (2015). They regard inertia as the result of a moral hazard problem due to the sunk cost of finding a first successful innovation. Developing a two period theoretical model, they find that inertia increases firm profits in stable environments while it decreases profits in volatile environments; a stable environment is defined as having a high probability that the optimal project is the same in both time periods.

The organization literature expounds on the effects of inertia on learning through exploration or exploitation. March (1991) is a pioneering paper on this paradigm of organizational learning. He posits that the difference between exploration and exploitation can primarily be characterized by the degree of uncertainty of the returns and that this uncertainty exists due to a 'greater distance in time and space between the locus of learning and the locus for the realization of returns'. In effect, I estimate this distance with the proximity measure I will describe in Section 3. My measure compares the distance of a firm's first year technological content - which is the location previous returns were being realized - with a future year's innovation output - the locus of learning. March concludes that a balance is needed in organizations as too much exploitation leads to inertia while too much exploration drives out efficiencies and prevents economies of scale.

There is a vast literature on the exploration and exploitation dichotomy, albeit a lot is quite qualitative (see Levinthal and March (1993), Sorensen and Stuart (2000), Smith and Tushman (2005), Gupta et al. (2006), among others). Lavie et al. (2010), Beckman et al. (2004), Lavie and Rosenkopf (2006) and Rothaermel and Deeds (2004) study these forms of organizational learning in the context of external alliances. Benner and Tushman (2003) investigate the role of managerial processes. Sadler (2017) uses a network structure to model the collective choice to explore or exploit. He finds different incentives depending on whether the network has a more centralized or decentralized structure.

While exploration and exploitation must be balanced for firm performance, Uotila et al. (2009) suggest that the optimal proportions of the two are highly subject to environmental conditions. Tushman and O'Reilly (1996) and O'Reilly and Tushman (2008) put forth the concept of strategic ambidexterity as a way to balance the two. They define strategic ambidexterity as: the ability of a firm to conduct exploration and exploitation at the same time. They highlight the role of management structures and consider ambidexterity to be a capability to be learned. He and Wong (2004) test the impact of ambidexterity on sales growth rates and find a positive association.

Another form of organizational learning that can potentially provide a balance between exploration and exploitation is experimentation. Koning et al. (2020) empirically analyze the effects of A/B testing in the digital industry. They find that testing increases page views and product features. However they also find partial evidence that start-ups fail faster with A/B testing while large firms scale faster. This seems to imply that testing is a tool to identify misallocation. Thomke (2001) describes the fall in costs of experimentation derived from new technologies such as fast prototyping, computer simulations, etc. He suggests that these new developments make experimentation a viable form of organizational learning now.

Finally, the broadest definition of exploration and exploitation can involve learning about many different types of information. In Zhou and Wu (2010), they focus only on the technological dimension. They find that greater technological capability affects exploitation positively whereas it has an inverted-U shaped relationship with exploration. Their reasoning centers around the trade off between absorptive capacity and structural inertia.[1] They regard absorptive capacity as an element that drives exploration by increasing the firm's receptiveness to new information. Although I focus on confirming firm inertia and the consequences of it, I do not explore the role of absorptive capacity in exiting it. Nevertheless, I will look at the role of previous experience patenting in the firm's founding team, with the suggestion that previous experience may be a weak indicator of more absorptive capacity.

[1]Absorptive capacity is a term coined by Cohen and Levinthal (1990) which refers to the ability of a firm to integrate external information. One major constituent of this ability is captured in the existing technological stock.This is the motivation for many papers in innovation to include a knowledge stock measure in their models.

Many of the studies on exploration vs exploitation have similar counterparts in the product and process innovation literature or the incremental and radical innovation literature. Notably, Manso (2011) uses the exploration and exploitation paradigm to develop a model of the incentives for radical innovation. His results emphasize the importance of tolerance for early failure. A number of papers empirically test his results including Aghion et al. (2013) and Tian and Wang (2014).

Other papers that discuss incremental and radical innovation include Chandy and Tellis (2000) who discuss the effects of firm size starting from an assumption that large firms do not do much radical innovation. They explain the incentives for incremental vs radical innovation in terms of the product S-curve. Ettlie et al. (1984) considers the firm structure, identifying its role in determining whether a firm does more radical or incremental innovation. They suggest that in order to produce a radical innovation, the firm should be uniquely structured for it while increment innovations can result from more traditional structures. This also connects with the firm size dimension as they investigate the food processing industry and find that incremental innovation tends to be found in large firms while radical innovations are found in specialists.

Zhou and Li (2012) explore how the existing knowledge stock interacts with different knowledge integration mechanisms, namely internal knowledge sharing vs external knowledge acquisition) and how that affects radical innovation. Using survey data they find that a firm with a broad knowledge stock is more likely to produce radical innovation from internal knowledge sharing while a firm with more depth in their knowledge stock is more likely to integrate external knowledge to produce radical innovation. This addresses a similar question as Zhou and Wu (2010) with the added dimension of breadth and depth of knowledge stock.

Finally, there is a vast literature that makes a distinction between product and process innovation. This dichotomy matches well with economic models given that a product innovation can be introduced as a higher value or quality product or as a new product line (see Klette and Kortum (2004)), while a process innovation is likely to be captured in the marginal cost parameter.

Product and process innovation can be considered through the lens of inertia in product space.

New products resulting from product innovation can then generate incentives to conduct process innovations. For instance, Utterbeck and Abernathy (1978), a paper also often cited in the exploration and exploitation literature, consider the technological life cycle and suggest that product innovations have a higher payoff at the starting of the technological sector life cycle while process innovation payoffs increase later in the life cycle. This, they suggest, is due to the increased importance in reducing costs.

Some more evidence for this connection can be found in Damanpour and Gopalakrishnan (2001) who find that product innovations are adopted faster than process innovations and that the adoption of product innovations is associated with more process innovation. Similarly Cohen and Klepper (1996) suggest costs are a chief reason for process innovation. Instead of the technological life cycle however, they consider firm size and they suggest that process innovations contribute less to firm growth. However their model finds that large firms have an advantage in process R&D because they have a larger output over which they distribute the R&D costs.

A new product essentially imposes a known demand on process innovation where as the demand for product innovations is less certain. Therefore when a firm has a large demand for its products, the decision to focus on process innovations to take advantage of economies of scale can be optimal. I will provide evidence for a similar life cycle effect however I instead consider complementarities instead of processes. Complementarities are also a way to take advantage of economies of scale from prior inventions.

3 Data Description

I use patent data to measure a firm's technological content. Although many firms do not patent, I would suggest that for the purposes of studying just the technological dimension, patenting firms are the ones that are of primary interest. To be precise, I use the 2017 spring vintage of the Patstat database provided by the European Patent Office (EPO) for the entirety of my analysis. This database has a very detailed and broad coverage of patent publications information. To avoid open economy influences and avoid differences between different countries intellectual property policies, I focus my analysis on the United States.

When a technological invention is made by a firm it can choose to either patent it, publish it with a scientific journal or keep the invention internal as a trade secret. Scientific articles, however, do not protect the invention in any way. They are more a tool to gain in reputation among the scientific community. Trade secrets on the other hand, are hard to keep. Many inventions lose their secrecy once they are introduced into a product since many products may simply be taken apart to learn the technologies (a.k.a. reverse engineering). Hence patents are a viable option and arguably the best. There may also be firms that do R&D but do not produce a technological innovation, ex. firms that study the impact of color on appetite, may be considered to do research however their finding does not necessarily contribute to a technological field. I expect these firms to be similarly affected by inertia however they may be arguably less entrenched by physical capital but perhaps more by human capital.

The Patstat database has very good coverage of granted patents in the US. Since a patent filing is made at a patent office in order to be evaluated for grant this data is nearly population data however the data becomes less reliable if we go too far back when the filings were not digitalized[2]. On the other hand, patent data can also be subject to a truncation issue. Patent applications take time to be added into the database. Patent grants are a particular issue because patents often take many years to be evaluated before a grant decision is made. Furthermore, this grant lag is quite variable. To avoid these issues I choose my dataset with a time buffer, namely I use only patents filed between 1980 and 2014.

In order to study the firm life cycle and effects of inertia, I need to restrict the sample to firms that patent at least twice. Furthermore a number of patents are in fact not associated with a technology code. Since my objective is to measure the technological content of firms, I keep only firms with patents that have associated technology codes. The US patent office uses a classification different from the ones used by the international community and namely different from the one used by the EPO. The EPO uses the International Patent Classification (IPC) established by the Strasbourg Agreement in 1971 and overseen by the World Intellectual Property Organization. There might be some patents lost in the concordance between the USPC and the IPC. In 2013 the USPTO jointly developed a classification system with the EPO called the Cooperative Patent Classification (CPC) system. The CPC code is likely a better match however it is quite

[2]There are digitization efforts that have been done to improve data quality in earlier years. See Akcigit et al. (2018) for example

new and not quite standard to use yet. Therefore I focus my analysis on the patents that have IPC codes while I do some robustness checks with CPC codes. My final dataset covers 88511 firms.

Although patent applications do not change, their technology classification codes may change. The technology codes were introduced by the patent offices to sort documents and facilitate search. As new technological sectors emerge, the patents get rearranged into those sectors. While the manual process before the digitization of the patent offices classified patents under a single technological code, digitization now allows patents to be classified under different technology codes. As I am primarily interested in the technological content of firms; which I infer from the patent classification codes, my measures are subject to change as classification codes change. This is another reason I use CPC codes for a robustness check. The results will show that the patterns are very robust to the classification system chosen.

3.1 Construction of main measures

The primary objective of this study is to compare the composition of firm technology over its life cycle. In particular, to measure the inertia of initial decisions, I will compare a firm's technological content to its technological position in the first year it patented. To do this, I refer to Jaffe (1987) for a proximity measure that can measure the technological distance between two patent portfolios. I will construct this measure with 4 digit IPC codes as is common in the literature when building measures with technology codes. The Jaffe measure is essentially an uncentered correlation. As such, if the definition of the technology codes is too narrow, the measure may return many zeros. In 2013 Bloom, Schankerman and Van Reenen develop a variation on this proximity measure where they introduce a weighting matrix between the technology codes that captures spillovers between technology codes. They build the weighting matrix by calculating the correlation between patent classes that are filed together within a firm.

I modify this measure in a way that I suggests captures complementarity between patent classes. The Bloom Schankerman and Van Reenen weights are constructed by grouping the technology codes at the firm level. This could be mixing the technology codes that are complementary to each other with the technology codes that are filed in different products within a firm. It could be argued that different products within a firm are likely to be complementary products and therefore technology codes between the two are possibly complementary. However this mixes complementarity in the product space with complementarity in the technology space. Their measure also confuses the complementarity versus substitution effects by keeping the values along the diagonal. Here I construct a slightly different weighting matrix where the correlation is built at the patent level. The majority of patents are classified under multiple technology codes and it is the combination of these technology codes that capture the content of the patent. This implies that these technology codes are complementary in this patent. Therefore the measure I build starts at the patent level where I use the set of technology codes within a patent. In addition, to ensure that I am only measuring complementarity and no substitution effects, I remove the values along the diagonal. Ideally I would construct this weighting matrix at the full digit level of the technology codes however when I am dealing with patent level matrices, this is simply not feasible. Therefore I build the measure with IPC codes at the 4 digit level. Unfortunately when I remove measures along the diagonal, I am also removing different full-digit technology codes that might otherwise have been considered complementary. It is difficult to say whether two full technology codes are more complementary or more substitutable if they are in the same IPC 4 digit group. Essentially my measure here assumes that those measures are more substitutable than complementary.

Another patent measure that captures the technological content of the invention is the Trajtenberg et al. (1997) originality and generality measure. The generality measure is however built with forward citations which are subject to serious issues of truncation (see Hall et al. (2001) among others that document this.). Therefore I chose to only look at the originality measure which is built using backward citations. The originality measure is built like a Herfindahl index on the technology codes in a patent's citations. This measure is higher when the set of technology codes in the cited patents are more diverse The assumption in this originality measure is that a patent is more original when it covers a wider set of technology codes, when it has a larger breadth. This might not be an exact match with the average person's impression of what constitutes originality however there is some evidence that an invention is more original when it covers a more diverse set of technological fields (see Angrist et al. (2020) and Beckman (2005)). Furthermore originality is an abstract concept that will at best be captured by a proxy and is this measure is widely accepted in the literature.

The originality measure construction is not as computationally intensive as the proximity measure therefore I can use the full technology codes to build an alternative measure. The originality measure is a patent level measure. Since I am interested in the firm's decisions, I want to aggregate this to the firm level. I do this in two ways - by taking the mean and the max. The mean is the intuitive choice since it is the average originality within the firm. However in the case of technological content, using the maximum can also make sense as it can be argued that the technological borders of a firm are defined by the most original inventions and not the average invention (see Henkel and Roende (2018)).

In Lee (2020) we have already seen that originality was steadily increasing until 2008 afterwards which it has been decreasing. Lee (2020) expounded on the idea that this fall in originality has been due, to some extent, to new entrants who have changed their innovation strategies to increase their likelihood of getting bought out.

3.2 Identifying firm characteristics

The debate over whether young or old firms are more innovative was first expounded on by Joseph Schumpeter who himself seems to have changed his mind suggesting first that young firms are the driving force however then arguing later in his life that large firms are the primary source of innovation. Since Schumpeter there have been many studies tackling this question without reaching a consensus. Part of the reason this might be so confusing is that the firm age and size terms are often used interchangeably and the empirical tests have usually used the small-large distinction. While it is often true that young firms are small and old firms are large, it is not always the case.

Focusing on the firm age dimension is much more straightforward and to the best of my knowledge has not been done for patent originality measures. It is evident that studies that have used a count of patents to measure innovativeness have found that older firms are more innovative since they are likely to be larger and have larger R&D teams as well as more reason to do more R&D on all the product lines they own. Plus older firms have more reasons resources and incentives to do more process innovation as well.With originality, I can aggregate without being directly affected by firm size. Figure 1 displays the average firm originality over time for new entrants and 20 year old firms. Albeit some noise, we see that new firms have historically been the more original firms however they have recently been overtaken by older firms.

In order to define the firm life cycle I need to be able to determine firm age. This information is not explicitly available in Patstat. Instead I apply the assumption that firms that have patented sometime in their life are going to be patenting from the start. This means that I assume no firm enters without patenting out of the firms that do patent. In reality there could be a some firms that exist for a few years without patenting that later choose to patent. With this assumption I can infer the entry year of a firm from the first year it begins patenting. I verify this choice by comparing the founding dates of public firms from the Jay Ritter dataset with the first year a firm begins patenting in my dataset[3]. The match is usually quite good, with the most common discrepancy being only one year. I check a random selection of some of the larger gaps by manually finding the firm's founders and comparing it with the inventors on the patent. They are often a match. This implies that the R&D for these firms does start from the year in my dataset however the firm incorporation sometimes occurs many years after.

A rough timeline of a new entrant's progression could look something like this: in the starting stages, they may begin with a person or group of people with an idea that is set in a particular technological sector. Then they gather more resources such as engineers and physical supplies to conduct the R&D and build the idea into a product. This R&D process establishes a technological position for the firm. With frictions and sunk costs associated with this initial position, firms inherently develop a comparative advantage in that position and therefore have incentives to continue building off it.

It is also useful to determine a firm's primary technological sector and explore the sector dynamics. Like industries in product space, technological sectors are also likely to be heterogeneous and have a life cycle pattern. I assume the main heterogeneity of sectors is their concentration. If a sector is highly concentrated, it is likely to be dominated by a few firms. When the few firms are large, they might disincentivize innovation in the sector because a new firm might expect it to be hard to compete. On the other hand more innovation in the sector, regardless of whether

[3]See https://site.warrington.ufl.edu/ritter/files/2019/05/FoundingDates.pdf

they are concentrated in few firms or not imply knowledge spillovers that will encourage more innovation in the area.

I define a firm's primary technology sector by calculating the number of patents filed in each IPC 4 digit code over the firm's lifetime. Then I designate the IPC code with the most number of patents as the firm's primary sector. There are some cases where two 4 digit IPC codes have the same count of patents, I choose to drop these cases to avoid excess noise in the data. If I were to wrongly classify firms into sectors, they are likely to behave differently than the real firms in that sector and they will simply introduce more noise. Another option is to use 3 digit or 1 digit IPC codes to allow for a broader definition of a technological sector. This decreases the cases where the primary sector is uncertain; however it also means a more aggregated sector definition that might include sub sectors that have very different trends. For example the "A61K" 4 digit IPC code is very different to the "A01B" IPC code. However they would both be grouped into the same sector if I use 1 digit IPC codes. Nevertheless, for tractability in the analysis I will sometimes use 1 digit IPC codes, although I use the 4 digit IPC code when possible.

Finally I also group firms into categories by firm size and the concentration of their primary technological sector which is measured by the Herfindahl index. This makes the analysis more tractable and allows me to interact the firm age effects with these measures. For concentration, I group firms by the 25th, 50th, 75th, and 90th quantile that their primary technological sector is in each year. For size categories I define them by groups delimited by the 25th, 50th, and 90th quantiles each year.

3.3 Summary Statistics

Table 1 summarizes the main variables and provides some descriptive statistics. It shows that the average originality is higher for measures built with the full technology code as opposed to the 4 digit code. This is logical since originality is higher when the set of technology codes is higher. This measure, like the Jaffe proximity measure, does not take into account proximity between technology codes. So two technology codes like "G02B 1/02" optical elements made of crystals and "G02B 1/06" optical elements made of fluids in transparent cells are considered completely different despite both being optical elements. This leads to an average firm originality built from the full technology code to be on average higher than the maximum firm originality built from

4 digit technology codes. This displays a disadvantage of using the originality measure - the nominal levels of originality cannot be easily compared to other originality measures. Instead, it is mainly useful to compare the same originality measure with different subsets of the sample or over time. The Pearson correlation between the average 4 digit IPC originality measure and the average full digit IPC originality measure is 0.6669. In comparison the correlation between the average 4 digit IPC originality measure and the firm maximum 4 digit IPC originality measure is 0.8772. The correlation between the average originality built with 4 digit IPC technology codes and the average originality built with the 4 digit CPC technology codes is 0.8279.

Variable Name	Count	Mean	Std	Min	Q25	Q50	Q75	Max
Mean Originality - 4 digit IPC	504266	0.6249	0.2155	0	0.5075	0.6724	0.7873	0.9996
Mean Originality - full IPC	504266	0.8604	0.1358	0	0.8283	0.9012	0.9439	0.9998
Max Originality - 4 digit IPC	504266	0.6799	0.2214	0	0.5787	0.7449	0.8437	0.9998
Mean Originality - 4 digit CPC	588300	0.6333	0.2098	0	0.5242	0.6789	0.7894	0.9997
Jaffe Proximity	378236	0.5286	0.4012	0	0	0.5774	1	1
Adjusted Proximity	378236	0.6074	0.4467	0	0.0517	0.7056	1	2.052
Complementary Proximity	378211	0.0513	0.0967	0	1.716e-3	1.526e-2	5.191e-2	0.7148
Knowledge Stock	403033	29.91	276.7	1	1.276	3.550	10.02	30669

Table 1: Basic summary statistics.

Table 1 also summarizes the different proximity measures. We see that the Jaffe proximity has many zeros and ones. This is to be expected since it does not take into account the different spillovers between technology codes. The spillover adjusted proximity measure is always equal to or larger than the Jaffe proximity because it essentially keeps the Jaffe measures - which would be the ones along the diagonal of the weighting matrix - and adds the off-diagonal spillovers. This also means that the adjusted measure is no longer bounded between 0 and 1. Finally the complementary proximity measure is much smaller in magnitude than the other proximity measures and that is to be expected because it keeps only the off-diagonal elements in the weighting matrix. Since the vast majority of patents are more likely to be filed with full technology codes under the same 4 digit code than under different 4 digit codes, the weights along the diagonal are much heavier than off the diagonal. Since I removed the diagonal elements to avoid confounding with substitution effects, I am left with much smaller weights. I avoid normalizing the weighting matrix to give me a proximity between 0 and 1 because in this way, I can compare between the different proximity measures. The difference between the Jaffe proximity and the adjusted spillovers proximity captures essentially only the spillovers. This measure is not quite the same as the complementary proximity because I build the weighting matrix for the complementarity construction from the patent level. This results in smaller weights in the off diagonal matrix when compared to the firm level spillovers weights. The correlation between the Jaffe proximity measure and the adjusted proximity measure is 0.9540 while the correlation between the Jaffe proximity measure and the complementary proximity measure is only 0.17436. As expected, the correlation between the adjusted proximity and the complementary proximity is in between those two measures at 0.4584.

Also to note, the proximity measures calculate the correlation of a firm's first year technological position with the technology codes of the firm's new patent filings over time. Namely, the new patent filings make up the change in the firm's position. If I were to calculate the firm's full position, I would add the count of technology codes in previous patent filings. However this would clearly give a much higher proximity since the patents filed in the first year would still be counted in the full position.

Finally, Table 1 shows that the standard deviation of the proximity measures is quite high relative to its mean. This might imply that proximity is quite volatile. However the measures in this table are calculated over all the firm-year observations. We will see in the results that they actually follow quite predictable patterns. I also summarize the firm size proxy knowledge stock in Table 1. This gives us a rough idea of the distribution of firm sizes in my sample. While the mean knowledge stock is 29.91, the median is only 3.55 and the maximum is 30669 (IBM holds the title of maximum knowledge stock in my sample). This implies that the distribution is highly skewed. As such I define size quantiles to categorize the firms instead of using nominal values. This avoids having outliers influence the results.

Table 2 compares the size distribution of firms by their primary ipc 1 code. We see the large degree of heterogeneity between industries and again the skewness of the size distribution. At the 95th quantile, it would appear that the C class has the largest firms and that class G is one of the classes with smaller size firms. However, when we look at the largest firms by primary IPC 1 digit codes, we see that G actually has some of the largest firms (including IBM and Microsoft) and that the C class has medium sized firms, becoming sixth ranked of the 8 different 1 digit classes. Furthermore if we look at the originality and proximity measure averages by 1 digit IPC codes, we see that the C class has the highest average originality. This corresponds to a high complementary proximity however it is the lowest in the conventional Jaffe proximity measure. This makes sense since intuitively, originality should be the inverse of proximity. Note however that this relationship is more subtle because proximity is in relation to a firm's first year patenting choices and not to the theoretical technological frontier.

1 digit IPC: Description		Q25	Q50	Q95	Q99.9	Originality	Proximity	Adjusted Proximity	Complementary Proximity
A	Human Necessities	1.197	3.445	53.02	5792	0.604	0.6056	0.7419	0.0953
B	Performing Operations; Transporting	1.444	3.6	53.27	9270	0.5909	0.4853	0.5134	0.0181
C	Chemistry; Metallurgy	1.966	5.152	226.31	4358	0.6743	0.4669	0.6152	0.1016
D	Textiles; Paper	1.197	3.496	79.63	659.2	0.5831	0.5213	0.5523	0.0219
E	Fixed Constructions	1	2.85	40.23	3911	0.5451	0.5718	0.5958	0.0155
F	Mechanical Engineering; Lighting, etc	1.522	3.795	64.96	13483	0.6004	0.4699	0.508	0.0249
G	Physics	1	3	55.43	30658	0.5969	0.5837	0.6377	0.0288
H	Electricity	1.723	4.464	129.78	11767	0.6018	0.5232	0.5847	0.0357

Table 2: Summary statistics by 1 digit IPC technology codes. The quantiles are defined using the discounted knowledge stock as a size proxy. The originality measure is built using 4 digit IPC codes, the proximity measure is the Jaffe (1987) proximity measure, the adjusted proximity measure is the Bloom et al. (2013) measure and the complementary measure is built with the complementarity weighting matrix as described in Section 3.

4 Empirical Strategy

As discussed in Section 1, there are many reasons for a firm to exhibit some degree of inertia in its innovation decisions. It is particularly interesting to study how inertia occurs in new firm entrants to better understand their innovation patterns and how they can affect the sector in later years. In particular, I am interested in how long initial choices and conditions affect a firm's technological position.

To study these patterns and quantify the degree of persistence in initial conditions, my general approach is to estimate functions of the form:

$$P_{i,t} = f(\text{age}_{i,t}, X_i; \beta) + \gamma_t \mathbf{D}_t + \epsilon_{i,t} \tag{1}$$

where $P_{i,t}$ represents the proximity measure or the originality measure described in Section 3 for firm i in year t. $\text{age}_{i,t}$ is the variable for firm age, X_i represents the firm's initial conditions. I also include technology sector dummy variables, D_i^s, in the firm's initial conditions as I assume that firms first choose their primary technological sector and then work on developing an invention in that sector. β and γ are vectors of parameters and D_t is a set of year controls. The year controls are added to capture any overall time trends. For instance, with originality we saw clearly that it was increasing steadily before falling around 2008.

My primary objective is to estimate $f(\text{age}_{i,t}, X_i; \beta)$. This function describes the average proximity or originality of a firm who entered with initial conditions X_i at $\text{age}_{i,t}$. I will look at the effect of different initial conditions based on different hypothesis.

For a baseline, I look at the case of the pure age and technology sector effect and move the rest of the initial conditions variables to the controls. To allow for maximum flexibility for the effect of age on my dependent variable, I separate the age variable into dummy variables and define $f(\text{age}_{i,t}, X_i; \beta)$ as $f(\text{age}_{i,t}, D_i^s; \beta) = (\beta^0 + \beta^s D_i^s) D_i^a$ where D_i^a is the set of age dummy variables defined as $\{d_{i,\tau}\}_{\tau=1}^n = \{\mathbb{1}_{(\text{age}_{i,t}=\tau)}\}_{\tau=1}^n$. I add the technology sector dummies to control for sector heterogeneity however I do not want these terms to affect my average age effect. Therefore I follow Guerts and Biesebroeck (2016) and impose a restriction on the β^s parameters. Namely I add a constraint where the summation of the β^s parameters must add up to zero. $\sum_{s \in \mathcal{G}} \beta_i^s = 0$. Where $\mathcal{G}$ is the set of technology sectors as defined by 1 digit IPC codes. I also add in another set of sector dummies without the age interaction to control for the pure technological sector effects. These decisions make my specific baseline regression model:

$$P_{i,t} = \sum_{\tau} (\beta^0 + \sum_{s \in \mathcal{G}} \beta_i^s d_i^s) d_{i,\tau} + \gamma^s \mathbf{D}_s + \gamma^i \mathbf{D}_i + \gamma^t \mathbf{D}_t + \epsilon_{i,t} \tag{2}$$

where d_i^s are individual vectors in the $\mathbf{D}_i^s$ set and similarly $d_{i,\tau}$ are the set of age dummy variables in $\mathbf{D}_i^a$. The $\mathbf{D}_i^s$ are the 1 digit IPC technology sector controls and $\mathbf{D}_t$ are the year controls.

Age and year are clearly exogenous variables so I do not have any endogeneity issues. The technology sector and other firm characteristics are fixed variables here so I also do not have any endogeneity issues. The sole concern may be measurement error in the variables. Namely if there is an imperfect match between the firm and the patents I might either be missing some patents that the firm has filed or I could have wrongly assigned some patents to a firm. Since my observations are defined by the algorithm-cleaned applicant filing name, it is quite likely that there are a few error like this. If my firm name cleaning was not stringent enough, I might have grouped firms that have a similar name but are not in fact the same, together. On the other hand, if the algorithm was too stringent, I will have missed some applicants that should actually have been grouped together. I have done numerous checks in developing the name cleaning algorithm to minimize these errors however they may still exist.

Assuming the firms are correctly assigned, it is still possible for some innovation measures to experience measurement error. One of the reasons for introducing the weighting matrix in the construction of the proximity measure is to decrease this issue. The conventional Jaffe proximity measure is subject to too many values on the boundaries of zero and one and this effect is exacerbated when a patent has few technology codes. See Section 3 for more details.

The originality measure is also subject to measurement error as was already discussed in section 3 where we compared the values from the 4 digit originality measure with the originality measure constructed from full technology codes. In addition, Hall in Trajtenberg et al. (1997) suggests that originality is naturally biased. She suggests that since originality is based on backward citations and that the set of patents that are available to be cited is increasing over time, originality will increase with time mechanically - since more technology codes that are cited will mean a higher originality and having a larger set of potential patents to cite also increases the likelihood of citing more different technology codes. At the time her article was written, we did see an increasing trend in originality that suggested this was the case, however, with the 2017 Patstat vintage we see that since 2008, originality has in fact been decreasing.

I believe Hall's argument is based on the assumption that patent citations are chosen to list the knowledge and technology in the citation that was used in developing the invention that is to be patented - as we do in scientific articles in academia. However citations serve a slightly different purpose in patents. First although the applicant can choose some citations, the final say in what citations are listed is made by the patent examiner. This already implies that patent citations added by the patent examiner were not known or deemed useful in the development of this invention by the applicants or inventors of the patent. Instead, the citations made by a patent are used to delimit the boundaries of the technological content of the patent. In theory the citations are meant to capture the existing knowledge and inventions that are closest to the new patent application. It is not evident that these boundaries change in any systematic way hence I suggest that this process is not subject to the mechanical bias Hall was suggesting. It might be subject to changes in patenting policy or practices however. For instance, it has been observed that the EPO and the USPTO follow different practices in assigning citations. The year dummies included in the model will capture that effect and any changes that occur in the patent citations practice over time.

It is not evident how measurement error on a firm's entry year, which I use to build the firm's age variable, may affect my model. Although a firm may have been founded before the first year they start patenting, this discrepancy does not necessarily affect the future way the firm develops its technological position. One possibility may be that the firm has an advance on the future R&D projects it does in terms of years measured in my data. This might show up in my data as firms that patent more initially. In terms of the firm life cycle, it might lead to a faster fall in proximity as the life cycle might show up more compressed/shortened. Thus I would expect that if this is an issue, it would bias the coefficients downwards.

Finally the technological sector definition is also subject to measurement error. Rather it is an imperfect measure of what I want to capture. In constructing a primary technological sector I have to choose one sector. In Section 3 I already detailed how I dropped the firm observations with uncertain primary technology codes. However, other than a data measurement issue, this definition is also subject to a logical issue since a firm, particularly larger firms, are composed of multiple product lines. In some cases their products and corresponding technologies can be very different (for example in firm conglomerates). This is an issue I discussed in Section 3 when building the weighting matrix at the firm level following Bloom et al. (2013) for the proximity measure. However it also becomes a problem here in defining the primary technological sector a firm is in. It is not obvious how a firm should be assigned a technological class if they are involved in the R&D of very different technologies. For instance, General Electric in my dataset, is classified under the 1 digit IPC code F which is the Mechanical Engineering sector. However at the more granular aggregation of 3 digit IPC codes, General Electric classifies as "C08" (Organic Macromolecular compounds; their preparation or chemical working-up; compositions based thereon) and then returns to the "F" class with "F01D" (Non-positive-displacement machines or engines) at the 4 digit IPC code level. Similarly Intel is classified as "H" (Electricity) at the 1 digit level however it classifies as "G06" (Computing; calculating or counting) and "G06F" (Electric digital data processing) at the more disaggregate 3 digit and 4 digit levels.

This implies that General Electric has a very specialized division on non-positive-displacement machines or engines where they are perhaps one of the leaders and pushing the frontier, however they also have a division that works on organic macromolecular compounds which patents a lot and more broadly than the non-positive-displacement machines or engines group since the primary code 4 digit code does not go under the "C08" section. How does this affect our estimates? In theory if an firm is classified in the wrong category and it is at the extreme end of the distribution of one of the variables, this could bias our estimates. For example if the dependent variable were firm size, then General Electric and Intel would clearly affect the estimates as they would have a differential effect depending on whether they are classified in one sector or another. A large firm that gets classified in one sector will increase the average of the whole sector to offset this, the coefficients on the other smaller firms in the sector will decrease. However my dependent variable is not size, it is originality and proximity. Neither measure is likely to have many outliers and further more, it is not obvious that any one particular kind of firm is more likely to be one of the outliers if they exist. For instance, although General Electric is one of the five largest firms in terms of knowledge stock in my sample, it is not necessarily an aberrant data point in the distribution of proximity or originality. Essentially the size of the firm does not have any weight here, so my model is not sensitive to a few firms that are difficult to classify. Furthermore, none of my explanatory variables are in nominal levels so they are not sensitive to outliers either.

Finally, I also include some fixed firm measures to control for. These are essentially the initial conditions and choices that the firm makes before the start of the observations for the dependent variable.

I assume a firm follows the following timing: A firm/applicant begins with a choice of technological field to do research in. It then builds a team of people (the inventors listed on the patent) to work on the research. These people come into the firm with their own backgrounds and some may have had experience creating a start up or experience in R&D and patenting. At this time the firm also looks for partners to join in the R&D process as well as spending resources to invest in physical capital such as machines and equipment. The outcome of this process is summarized in a patent application filed at the patent office. It is only from then on that I begin to have observations.

When proximity is the dependent variable, the values start in the year after, firm age 1, since the values at age 0 are irrelevant (it would be measuring proximity of the technological position to itself, so all the measures would be equal to one.) With originality, the dataset starts at age 0 (the first year the firm enters). Since the proximity measures start at age 1, I include the originality of the firm's first year as a firm control. To give maximum flexibility to the model I categorize the originality into three groups of low, medium and high where a firm is categorized as a low originality entrant if it's originality is below the 50th quantile; it is labeled as a medium originality entrant if it is between the 50th and 75th quantile, and it is labeled a high originality entrant if its originality is above the 75th quantile. The quantiles are defined by year and therefore the quantile thresholds are changing depending on the firm's entry year. This also means that I do not have an even one quarter of the start ups in the medium or high originality group nor one half in the low originality group although the distribution is not excessively different. In fact, of the 133596 firms, I have 37663 as high originality start up firms, 30473 as medium originality firms and 78840 as low originality start up firms. Lee (2020) discussed what potential factors could affect firm entry originality.

In addition, I can glean some information from the patent data to capture some of the information on the people connected with the firm prior to the patent application. Namely, since Patstat is nearly the entire patent population, I can see whether the inventors of the patent have been involved in a patent previously. As section 3 discussed, the applicant and inventor table in the database is subject to typos, therefore this tracking of the inventors is imperfect, however I would suggest that the majority of the inventors are properly tracked as they are less exposed to errors from name changes or in identifying subsidiaries then the firm applicants are. By tracking the inventors, I can infer who has had prior experience patenting and with this I can create an indicator variable for whether a firm has at least one person with experience patenting before. The prior experience patenting that an inventor has might be a signal that the inventor is more skilled at innovating and therefore might develop more original ideas and inventions. On the other hand, an inventor with experience patenting will have a build up of knowledge stock on the previous work he/she has done. This might influence him/her to patent in areas closer to that knowledge stock which might limit the originality of the research. I will explore the impact of experience explicitly in a section later.

Finally, with the patent application data I can see whether a firm patents its first patent with multiple applicants. Having multiple applicants on an application can mean different things. This could be a measure of the external relationships the firm has and can potentially be a signal of the resources the firm has access to. It may also simply be that the other applicant(s) are the other people involved in the R&D. In the Patstat dataset the people involved in the R&D are listed as inventors while the firm who hires the inventors is the applicant. However some employment contracts may include an allowance for the inventors to share in the intellectual property rights and they would therefore be listed as applicants (or assignee's which is the term the USPTO uses).

Furthermore, this measure may also be confounded with the error introduced by firm subsidiaries. In theory, I want my firm level observation to be the entity that is making the R&D project decisions. For large firms with different subsidiaries, this could mean different things for different firms. Which entity to list as the R&D decision maker will depend on the firm structure. Some firms operate with a very centralized structure while others take a much more decentralized approach. Specifically subsidiaries are still at a level of independence higher than a firm branch and it is reasonable to expect that it is making many of its own decisions. However in very centralized structures, this is less the case, since officially, it is the owners of the firm's/subsidiary's equity who have the most decision making power. For my purposes of identifying entrants, I would ideally group the subsidiaries with their ultimate parent firms to avoid having faux entrants into the dataset. I already do this to the extent that is possible with only names, however some subsidiaries will still be missed. When a subsidiary files a patent, it is likely to include its parent firm as a co-applicant. Therefore by including a dummy variable for whether the firm's first patent included multiple applicants will control for some of these effects. I cannot distinguish them separately, however for the purposes of a control variable I suggest that it is sufficient.

Returning to my principle objective, the existence of inertia in a firm can largely be established with Equation 2. I am also interested in how the degree of inertia differs depending on starting conditions. Namely, I will look at how the initial originality of the firm affects its future behavior and I will look at how previous experience on the founding team affects the development of the firm. In these cases the regression model is:

$$P_{i,t} = \mathbf{X}_i(\beta^0 + \beta^s \mathbf{D}_s)\mathbf{D}^a_t + \mathbf{D}_i + \mathbf{D}_s + \mathbf{D}_t + \epsilon_{i,t} \tag{3}$$

This model includes the starting condition of interest as X_i and it is interacted with the age dummies as well as the age and sector terms. My estimate of interest will be the α coefficients which describe how the firm life cycle dynamics are different depending on the initial originality of the firm or the prior experience of the firm.

Next, I address the long standing debate in the firm innovation literature on whether small firms or large firms are more innovative. To do this, I add firm size into the regression. This will enter in the interaction term with age to disentangle the age and size effects that are often confounded in the literature. Since firm size is highly skewed, I transform the measure into four dummy variables based on its ranking in the size distribution. The groups are delimited by yearly quantile thresholds of 25, 50, 90, and 99. To be exact, this means that I calculate the 25th, 50th, 90th and 99th quantile of the firm size distribution each year. then assign firms to their group each year as well. This is no longer a pre-sample variable (aka initial condition) as I expect the main variation in size to occur when the firms are older.

This means that when I look at the variation in firm life cycles by their size, it is no longer the average of the same set of firms. Firms can switch between size categories as they grow over their life cycle. The regression with firm size can be explicitly written out as:

$$P_{i,t} = \mathbf{D}_{i,t}^{size}(\beta^0 + \beta^s \mathbf{D}_s)\mathbf{D}_t + \mathbf{D}_i + \mathbf{D}_s + \mathbf{D}_t + \epsilon_{i,t} \quad (4)$$

Notably, my precise model captures a slightly different question to firm size as it uses knowledge stock as a proxy which is can be interpreted slightly differently. See Section 3 for the discussion and interpretation of this measure.

The econometric consideration in this regression is the implication of using a dynamic size variable as opposed to a fixed pre-sample variable. To avoid any confounding effects of timing when aggregating by year, I use the knowledge stock proxy, categorized into four groups, lagged by one year for the $\mathbf{D}_{i,t}^{size}$ matrix. This avoids any issue of a firm's proximity/technological choices interacting with the firms knowledge stock over the period of the year. If however, proximity influences the knowledge stock of the firm in the next period there could be an issue of serial correlation. Intuitively, the proximity of a firm's technological position is a measure agnostic to how large it is, however since the two variables are constructed from the same dataset, there is a small possibility that a connection exists that will introduce serial correlation.

To address this, I follow Guerts and Biesebroeck (2016) who suggest a couple different methods. In particular, I apply their method of using the firm's beginning of period and end of period size classifications and split the firm into two weighted by one half each[4]. This allows me to use fixed size measures instead of the dynamic ones which reduces any potential issues of serial correlation. Including a set of observations that use the end of period size categories is also useful for capturing more of the variation in size when firms are older, as opposed to using only the initial size of the firm when it enters since I expect that size variation to be quite small. However, I keep this method mainly as a robustness check as I do not expect the serial correlation to be substantial and it is preferable to use the time-varying lagged size variables.

Next, I also explore the effects of the concentration of the firm's primary technology sector when the firm enters. This poses the question of whether entrants have a different role in overall sector dynamics depending on the competitive situation of the sector. There is a growing literature on

[4]This method dates back to Prais (1958)

how entrants affect other firms in the industry and their role in the economy as a whole.

The general consensus in the firm dynamics and innovation literature is that new entrants are a potential threat to incumbents, and this potential future threat is one of the incentives for existing firms to continue innovating. So my question is, what affects the potency of this entrant threat? It its unlikely that the level of an entry threat is unchanged for all time and all environments. How do we measure the degree of an entry threat? These are the motivating questions for this entire study. I believe it is useful to better understand how firm's develop over their life cycle in terms of their technological position in order to start measuring this degree of "threat" or dynamic competition.

In particular, I next explore how the initial concentration of a firm's technological sector affects the firms degree of inertia and originality. This is done by first building a measure of concentration for the 4 digit IPC technological sectors by year. See Section 3 for more details on how this was constructed. Then I choose to categorize firms into groups by the concentration of their primary 4 digit IPC sector to keep the regression tractable. I assign one group for firms under the 25th concentration quantile, one group for firms between the 25th and 50th quantile, another between the 50th and 75th quantile, another between the 75 and 90th quantile, and lastly a group of entrants who enter in the most concentrated sectors - the 90th to 100th quantile. To build these estimates, I stick with the model described in equation 3 and use this grouping by initial concentration as the X_i measures.

Finally, I include a section on long lived firms - firms that have lived past 40 years old - to investigate what, if anything, they have done differently to the rest of the firms to be able to survive for so long. As a disclaimer, the data quality on long lived firms in their infancy is of worse quality as I have to use data further back in time. The model remains the same as above however I group my sample by the age of the firm they are at at their last-known patenting year. For this, I stop my sample in 2010 to give a larger buffer for firms that may not have exited and simply have a pause in patenting.

5 Results

I first provide different measures to summarize the average age effect. Then section 5.2 estimates the effect of the entrant's initial originality and section 5.3 explores the impact of having prior experience. Next, I introduce size in section 5.4 and examine whether the degree of inertia is different depending on the size of the firm's knowledge stock. Finally, I discuss the implications on the overall technological sector and in particular I test whether the concentration of the sector affects new entry behavior.

5.1 Basic Results

Figure 1 displays the basic results of firm inertia. It plots the β^0 coefficients on the age dummy variables from the model described in equation 2 using the Jaffe proximity measure. This captures the average effect of firm age on proximity to its first year technological position, controlling for the technological sector variation, the firm fixed characteristics and the year effects. We see clearly that the proximity is higher at the beginning of the firm life cycle and declines quite steadily. My focus on the entry year simplifies my estimate to measuring only the persistence of initial conditions. The first age vector was dropped to avoid collinearity therefore the plot starts from age 2. This also means that we have to interpret the coefficients as relative to the first year average.

We can compare figure 1 to 2 which uses the adjusted proximity measure proposed by Bloom, Schankerman and Van Reenen[5]. The two have very similar trends. However if we focus only on the proximity in complementary fields, the decline is much less clear. Figure **??** plots the complementary proximity measure described in Section 3. This measure removed all direct overlap between technology fields so we expect the magnitude to be much lower. The pattern is quite noisy and looks largely flat for the period I am studying. It seems to fall later in the life cycle, however it is not a clear trend.

While these regressions have confirmed that firms do exhibit inertia, it is not clear how the entrants' proximity evolves relative to other firms. In particular, I am interested in evaluating how long it takes for a given firm to arrive at a technological position that is equally distanced

[5]For the purposes of conciseness, figure 2 and all remaining figures will be in the appendix.

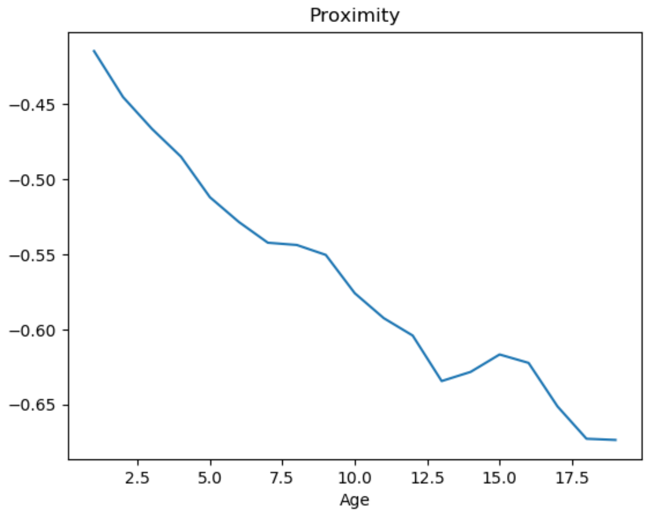

Figure 1: This figure plots the coefficients from the model described in equation 2. The line represents the average Jaffe proximity by firm age controlling for industry, year and firm fixed characteristics. The proximity measure compares the entrant's technological position to its first year technological position. For firm controls I include the firm's initial conditions, namely, its initial size quantile, whether it has previous experience, whether it first patented with other applicants and what the firm's initial originality groups is.

to its initial position as the incumbent firms in the sector. To do so, I construct another set of proximity measures that compare the position of the incumbent firms to the technological position of each new entrant in the same sector over time, where I define an incumbent as a firm that is over twenty years old. This measure of incumbent proximity is essentially a baseline of the technological evolution in the sector. Note that it is not quite a measure of the technological evolution that would occur without entrants as the threat of entrants can be a competitive motivation for incumbents to innovate which could lead to endogenous effects.

To measure when entrants reach a technological position that is similarly distanced to its initial position as the incumbents in the sector, I calculate the difference between the entrants age-varying proximity and the incumbents proximity. This is used as a new dependent variable in the model described by equation 2.

Figure 3 plots the result of this regression using the Jaffe proximity measure. As expected, the average difference in proximity between the entrants and the incumbents is higher in the early years of the firm and declining over time.

The trend is very similar when using the spillovers adjustment to calculating the proximity measure. We can also look only at the incumbent's proximity to new entrants in their sector. Although we have already established in figure 3 that the entrants have a higher proximity to their initial inventions than incumbents, we surprisingly see in figure 4 that the trend is also decreasing for incumbent proximity over the entrants life cycle. This means that new entrants and incumbents are not completely independent, otherwise we would expect this figure to be flat. Instead, we see that incumbents also patent in closer proximity to the patents new entrants are filing and that this proximity declines over time. The fact that it is not flat implies that the technological sector follows an overall evolution and that new entrants also follow the trends. However I cannot distinguish whether this is coming from knowledge spillovers from the incumbent firms, knowledge spillovers from the entrant or due to a parallel exogenous process like changes in university curriculums or government research agendas or another effect.

Section 5.3 will explore the impact of having prior experience patenting. This experience is likely to partially come from incumbent firms, therefore this indicator may be a way to test spillovers from incumbent firms to new entrants. I will return to this issue there.

Finally, I am interested in understanding the consequences of inertia. In particular, how does it correspond to the firm's overall originality. Figure 5 shows the pattern for average originality built from 4 digit IPC codes. The originality is at first decreasing and then starts to increase after around ten years. However the estimates are quite volatile. On the other hand, figure 6 shows that the average of the maximum firm originality is increasing over the firm life cycle and this trend is very robust. It is arguable whether the mean firm originality or the maximum firm originality is the best indicator of firm innovation. While the mean is the default in the innovation literature, it can be reasoned that the maximum originality is the better indicator since it is the invention pushing the innovation frontier. If we take the maximum firm originality to measure firm innovation, then figure 6 would suggest that older firms are the more innovative ones. However if we take the average firm originality as the measure, then the age effect is less clear.

5.2 Initial Originality

So far we have looked at firm proximity measured in relative terms to the initial position. We have also studied originality on its own. However we have not examined the two together, namely we have not looked at whether and how the originality of the firm affects its degree of inertia.

To investigate this, I group new entrants into low, medium, and high categories depending on their initial originality. The low originality entrants are defined as the entrants that fall into the bottom 50th quantile of the originality distribution in their entry year. The medium group consists of the firms between the bottom 50th and 75th quantile and the high originality group is defined as firms with initial originalities in the top 25th quantile. Figure 7 plots the estimates from equation 3 using initial firm originality categories in the interaction term. We see that the low originality entrants tend to continue innovating in close proximity to their initial position while the high originality entrants innovate the furthest from their initial position.

However when we compare the firm's proximity with the incumbents in the sector (see figure 8) we see that low initial originality entrants are the least inert in relative terms - the difference for them falls the fastest. This can be consistent with figure 7 when we consider the behavior of the other firms in the industry as well. As we saw in figure 4, there tends to be trends in the overall evolution in technological content with the average new entrant also following these trends. However the average low originality entrants is likely to be a laggard to these trends, therefore incumbent firms' are less sensitive to their entry and we expect the pattern for incumbent firms to be more flat. This would correspond to a faster fall in the proximity differences for low originality firms.

Figure 9 then shows the trends in the complementary proximity by initial originality groups over the firm life cycle. We see that the order when it comes to complementarity is inversed. The high initial originality entrants are patenting the most in complementary areas. This difference stays quite persistent and we see that gap between medium and high originality entrants increasing over time. A high originality entrant may be confident in its initial invention and therefore may be more comfortable with expanding into complementary fields. In contrast, a low originality entrant may recognize that their initial position is less original and therefore needs to put more effort into solidifying that initial position before exploring complementary areas. Indeed we see that the low originality entrants slowing increase their complementary as they age however it remains much lower than the high originality entrants. We might also expect to see an increase for medium and high originality entrants, however 9 shows their average declining over the firm life cycle. Since these proximities are constructed in comparison to the first year's measure we only capture the proximity to the first year. As an entrant explores new areas it may find new inventions that lead it to continue exploring new areas which lead it farther from its initial position as it has spent less time enforcing its initial position. The fact that a new entrant enters with a high originality may also be a signal for their propensity to explore. They may in general have less of an affinity to remain in the same technological areas. In contrast low originality entrants, who survive, appear to be more entrenched in their initial positions and therefore tailor their future R&D decisions to build off it. This indicates that some firms specialize more in exploration while others specialize more in exploitation.

Lastly, we can examine the average firm originality trends. Figures 10 and 11 confirms that high initial originality firms remain at a high level of originality for quite long. This is true for both their average originality and their maximum originality. We see however that the medium and low initial originality firms increase their originality over time and by the time they are 20 years old, their originality levels have largely converged. Note that originality is a measure with an upper bound of one, therefore the high originality firms are unlikely to increase their originality indefinitely.

In Lee (2020) we saw that startup patenting patterns have changed over time and in particular that their initial originality has been decreasing. Here I have explored the ramifications of this overall fall in initial originality and documented some facts about the dynamic consequences.

5.3 Initial Experience

This section will explore the impact of having prior experience patenting. Experience has been recognized by many papers as a driving factor in firm success (see Gompers et al. (2010), **?**, etc.). Experience can be viewed as a signal for skill, perhaps a higher absorptive capacity and intrinsically a proxy for knowledge stock. Here I explore its impact on firm inertial tendencies.

I find in figure 12 that a new entrant who has at least one person on the team with previous experience patenting is likely to continue patenting in closer proximity to its initial position over time. This implies that experienced entrants associate a value with strengthening their initial position.

As mentioned in section 5.1, experience is likely to partially come from incumbent firms. Thus this indicator may be a way to measure spillovers effects from incumbent firms to new entrants. On the other hand, an inventor who works in an existing firm could simply continue inventing in that firm if it were relevant to the firm. Instead, the action of leaving the firm and starting a new firm implies that the innovation is more radical and potentially of less value to the incumbent firms.[6]

This is in fact what we see. Figure 13 compares the differences in proximity between the entrants and the incumbents by experience history. We see clearly that the behavior is very different between the two and that they diverge in time. Experienced entrants do not decrease their proximity much relative to the incumbents. Since we have seen that the proximity is falling within the experienced entrant, this means that the proximity of incumbents is falling even faster.

In terms of complementarity, we again see a large gap between the experienced and inexperienced entrants (figures 14 and 15). Although the gap is large, the trend is quite similar, complementary proximity to the first year is increasing at first then declining. When removing the proximity of the incumbent firms however, we see that the complementarity is increasing quite steadily.

Figures 16 and 17 show the trends for originality. The maximum originality is increasing while the average originality is inconclusive; there is a lot of noise and little difference between the experienced and inexperienced entrants. This implies that the average invention developed by the experienced entrants is not highly original. As we saw in figure 12, experienced entrants associate a value with more proximity. This implies that they are developing follow on innovations that are not necessarily very original, which would bring the average originality of the firm

[6] The average entry originality from an experienced entrant is 0.6212 while the average entry originality for an entrant with no experienced members is 0.5769. This indicates that experienced entrants do tend to enter with a higher originality.

down. However we see that the maximum originality of the experienced entrants is increasing, implying that while a large portion of their research is in incremental innovations on their initial invention, they also make an effort to develop original innovations.

5.4 Size and Age Interactions

As noted earlier, there is an ongoing debate about the differential role of firm size on innovation. This is further confused with the effects of firm age. This section attempts to address this issue and disentangle the age and size effects. As described in section 3, I define the size of the firm by its discounted count of patents and then group them into quantiles by their ranking in the size distribution of their primary sector each year. As has been vastly documented in the literature, the firm size distribution is highly skewed, therefore I delimit the groups by the 25th, 50th, 90th and 99th quantiles. This is a time varying classification of firm size which means that the ranking of the firms is changing and they may be moving between quantile groups over their life cycle.

Figure 18 displays the estimates on proximity by age and size. We see that the largest firms are the most inert. Particularly, looking at the figure when the firm is 10 or 20 years old shows that the large firm is still in closer proximity to its initial position than the other firms. This implies that there is a certain degree of entrenchment for firm size to grow. The largest firms still have the highest proximity averages, meaning that their patents are very concentrated around their initial position. This is even more clear since firm size is calculated based on number of patents and in general a higher count of patents increases the likelihood of patenting in more technological classes which would correspond to a lower proximity.

This can help us understand the dynamics of competition. In particular, we can infer from this figure which new entrants become future competitors. When an incumbent is considering its incentives to innovate, it will evaluate the threats new entrants pose. I suggest that the entrants that pose the largest threat are the ones that increase in size as they get older. This would be the top 99th quantile in the last few years of my period. Perhaps surprisingly, these firms are not the ones doing a high amount of exploratory research, instead figure 18 shows that these firms are the most inert. They have really strengthened their initial position. I then check the average initial originality of these groups by the size category they are in when they are 20 years old. I find that the smallest firms had an average originality of 0.56, the second group had an average originality of 0.55, the third had an averaged of 0.57, the fourth had an average of 0.58 and the largest firms had an average initial originality of 0.60. This suggests that although the largest firms have a high degree of inertia, they are inert around an initial position that is highly original.

The proximity for firms in the 90th to 99th quantile is also quite high for young firms however this is no longer the case as the firm gets older with this group's average proximity becoming the lowest for firms that are 20 years old. This implies that the older firms in this group have substantially changed their technological position. This may be due to competitive reasons. We can perhaps consider the firms in the 99th quantile as monopolies while the firms in the 90th to 99th quantile are perhaps more in a oligopoly situation.

Finally, we see that the correlation between the firms proximity and size is not linear. In particular, we also see that the smallest firms are quite inert. Their proximity is consistently above the proximity of medium sized firms albeit below the largest firms. This seems to suggest an inverted U relationship where it is the medium sized firms that are changing their position the most. Since these quantiles were defined in terms of the technological sector's size distribution, we can infer some competitive implications. In particular, these estimates match with the inverted U relationship described in Aghion et al. (2005). Since my categorization of a technological sector is at the 4 digit IPC level, there may be room for both neck-and-neck competition in some more narrow technological fields as well as leader and follower relationships in others. If the medium sized firms that I identify are in a neck-and-neck situation, Aghion et al. (2005) argue that the incremental profits from innovating will be higher which will encourage more innovation in order to escape competition. In addition to the quantity of innovation, I suggest that the direction of innovation may also be a way to escape competition. Innovating away from the position where firms are concentrated can alleviate the competition as well. This seems to be the situation I see here with respect to medium sized firms. In section 5.5 I will explore the effect of concentration in the technological sector some more directly.

The complementary proximity figure (figure 19) shows a more linear relationship with size and complementarity. The smallest firms have a lower degree of complementary proximity with the trend remaining relatively flat. The largest firms do not have a clear pattern. Since they are categorized in the top 1 percentile, the number of firms in the group is much smaller and therefore more sensitive to individual firm variation. Furthermore, recall that the firm size categories are changing over time, this means that a firm categorized in the 90th to 99th quantile when 10 years old may have moved to the 99+ quantile by the time it is 15 years old. This resorting between groups over time will add even more noise.

On the other hand, the medium sized firms have a higher complementary proximity than the smallest firms, however they do not appear to be increasing their complementarity. The young firms in the 90th to 99th quantile are increasing their complementary proximity however the old firms in this group have the lowest degree of proximity for that firm age group.

In comparison, the originality results in figure 20 and 21 suggest that in the early years the largest firms are the least original while the second largest firms are the most original. This changes as the firms age. Again the pattern for the largest firms is too volatile for the average originality however we do see that they are consistently below the rest of the firms in terms of maximum originality. Since the size classes are lagged in the model, this suggests that the largest firms are less original because of their size. They do not experience competitive pressures that push them to innovate in new areas. In the older firms we see that it is the smaller firms who appear to be pushing the boundaries of maximum originality. However this is less clear when we look at average originality. When considering average firm originality, the trend over time and by size seems to remain flat.

Using only the firm's first year size to assign it to a fixed size class, the picture changes a bit. With a fixed size class, there is no resorting between quantiles over time. Figure 22 shows that when we look only at the firm's initial size, it is the smallest firms who are the least inert. On the other hand it is the largest firms who have the lowest degree of complementary proximity over time (figure 23). In terms of originality, the smallest firms again have the highest average for maximum originality while the large firms have the lowest. However average firm originality is a different picture where we see the firms who start in the 90th to 99th size quantile increasing their average originality much more than the other firms.

5.5 Sector Concentration

The section above on firm size has already discussed some possible effects of competition on firm inertia. Here I will accompany those results with a couple more figures that include the degree of concentration explicitly calculated.

Figure 26 shows the estimates for average proximity for firms grouped by the concentration level of their primary technology sectors. We see that the young firms in the highest concentrated sectors are the least inert. This is no longer the case for the older firms in this grouping however. The young firms in the highest concentrated grouping may be in the early stages of the technological sectors life cycle. Figure 25shows the average Herfindahl index for the sectors over their life cycle. Each sector starts with few firms at the beginning then see their Herfindahl index fall as entry increases into the sector. Eventually there are some dominant firms who beat the competition that leads to exit and more concentration in the sector. The young firms in concentrated sectors have the lowest proximity to their initial position.

The effect of concentration is less evident on the rest of the firms. If we examine concentration on the complementary proximity measure (see figure **??**, we see that the firms in the lowest concentration sectors have the lowest degree of complementary inertia while the firms in the medium concentration sectors have the highest degree of complementary proximity. This again suggests an inverted U relationship where it is the firms in the medium concentration sectors that are competing the most. By expanding into complementary fields, they are escaping the more direct competition in their initial positions.

In relation to the firms' originality (figures 28 and 27, we see that within the young firms it is the firms in the medium concentrated sectors who have the highest level of originality both in terms of average originality as well as maximum originality. This is consistent with the high degree of complementarity that we see the firms in the medium concentrated sectors doing. The results are less clear for older firms.

6 Conclusion

In this paper, I provide evidence for the existence of firm inertia in technological space. I then investigate the factors that affect the degree of firm inertia and what this means for overall firm originality. I suggest that a better understanding of these dynamics will help us understand the dynamics of competition.

I focus on young firms in general, with most of my study on firms from 0 to 20 years old. In particular, this allows me to define some fixed firm characteristics from the initial conditions and analyze their effects. I find that firms with a high initial originality are the least inert while low originality firms are the most inert. This ordering is inversed when we look at complementary proximity suggesting that the high initial originality firms are expanding into complementary fields. This translates into firm originality over time where we see that the initial high originality firms remain in the high originality group in terms of both the maximum and the average.

I then estimated the importance of prior experience for new entrants. The experienced entrants were more inert in both the traditional sense of proximity as well as the complementary measure. This has an ambiguous effect on firm average originality however we clearly see that the maximum firm originality is increasing for experienced entrants.

Next I evaluated size effects which show that larger firms are the most inert. The average age by size estimates suggest a weak inverted-U relationship where the smallest and largest firms are more inert and the medium sized firms are in a neck-and-neck situation where they need a lower degree of inertia to escape the competition. I explore this competition aspect further by explicitly calculating the Herfindahl index for 4 digit IPC technology sectors. The pattern seems to also imply an inverted U shaped relationship when we look at the firm decisions to expand into complementary fields.

Finally I compare all my results on firm proximity with firm originality to explore the effects of firm proximity on its patenting originality. While the results on mean firm originality are often inconclusive, the results on maximum firm originality have some clear outcomes. Overall, maximum firm originality is increasing as the firm ages. However large entrants have a low maximum originality, experienced entrants have a high maximumm originality and high initial originality entrants remain at the high degree of originality.

Notably, I do not address the question of what is the right degree of inertia in a firm. The section on prior experiences shows that experienced entrants have a higher average level of inertia. The fact that this is the case suggests there are benefits to inertia. Going back to Hansen and Freeman's observation about firm selection processes. It is possible that some degree of inertia is good for the firm. The norm in the literature on innovation economics is to encourage more innovation and more original innovation. This might not necessarily be the best for the firm, however the effect on overall welfare in the economy is a bigger question.

Ideally I would like to analyze how firm inertia affects dynamic competition in the sector. For the moment, this analysis stops short of that however I document that experienced entrants and entrants with an initial high originality appear to be contributing the most in terms of innovation. When exploring the firm size and concentration effects it appears to be the middle sized firms and firms in medium concentrated industries that have the lowest proximity to their initial position and expand the most in terms of complementary technology fields in reaction to competition. When evaluating the threat from new entrants that incumbents face, it appears that it is the new entrants who are initially more original and who have a high degree of inertia that are the most viable threats to incumbents.

7 Bibliography

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: An inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701–728.

Aghion, P., Reenen, J. V., and Zingales, L. (2013). Innovation and institutional ownership. *American Economic Review*, 103(1):277–304.

Akcigit, U., Grigsby, J. R., Nicholas, T., and Stantcheva, S. (2018). Taxation and innovation in the 20th century. *NBER Working Paper 24982.*

Akcigit, U. and Kerr, W. R. (2018). Growth through heterogeneous innovations. *Journal of Political Economy*, 126(4):1374 – 1443.

Angrist, J., Azoulay, P., Ellison, G., Hill, R., and Lu, S. F. (2020). Inside job or deep impact? extramural citations and the influence of economic scholarship. *Journal of Economic Literature*, 58(1):3–52.

Argente, D., Baslandze, S., Hanley, D., and Moreira, S. (2020). Patents to products: Product innovation and firm dynamics. *FRB Atlanta Working Paper No. 2020-4.*

Beckman, C. (2005). The influence of founding team company affiliations on firm behavior. *Academy of Management Journal*, 49(4):741–758.

Beckman, C. M., Haunschild, P. R., and Phillips, D. J. (2004). Friends or strangers? firm-specific uncertainty, market uncertainty, and network partner selection. *Organization Science*, 15(3):259–275.

Benner, M. J. and Tushman, M. L. (2003). Exploitation, exploration, and process management; the productivity dilemma revisited. *Academy of Management Review*, 28(2):238–256.

Bloom, N., Schankerman, M., and Reenen, J. V. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393.

Casamatta, C. and Guembel, A. (2010). Managerial legacies entrenchment and strategic inertia. *Journal of Finance*, 65(6):2403–2436.

Chandy, R. K. and Tellis, G. J. (2000). The incumbent's curse? incumbency, size, and radical product innovation. *Journal of Marketing*, 64(3):1–17.

Choi, J., Goldschlag, N., Haltiwanger, J., and Kim, J. D. (2019). Founding teams and startup performance. *Working paper.*

Cohen, W. M. and Klepper, S. (1996). Firm size and the nature of innovation within industries: The case of process and product r&d. *The Review of Economics and Statistics*, 78(2):232–243.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1):128–152.

Damanpour, F. and Gopalakrishnan, S. (2001). The dynamics of the adoption of product and process innovations in organizations. *Journal of Management Studies*, 31(1):45–65.

Ettlie, J. E., Bridges, W. P., and O'Keefe, R. D. (1984). Organizational strategy and structural differences for radical versus incremental innovation. *Management Science*, 30(6):682–695.

Gompers, P. A., Kovner, A., Lerner, J., and Scharfstein, D. S. (2010). Performance persistances in entrepreneurship. *Journal of Financial Economics*, 96(1):18–32.

Guerts, K. and Biesebroeck, J. V. (2016). Firm creation and post-entry dynamics of de novo entrants. *International Journal of Industrial Organization*, 49:59–104.

Gupta, A. K., Smith, K. G., and Shalley, C. E. (2006). The interplay between exploration and exploitation. *The Academy of Management Journal*, 49(4):693–706.

Hall, B. H., Jaffe, A. B., and Trajtenberg, M. (2001). The nber patent citations data file: Lessons, insights and methodological tools. *NBER Working Paper*.

Hannan, M. T. and Freeman, J. (1984). Structural inertia and organizational change. *American Sociological Review*, 49(2):149–164.

He, Z.-L. and Wong, P.-K. (2004). Exploration vs exploitation: An empirical test of the ambidexterity hypothesis. *Organization Science*, 15(4):481–494.

Henkel, J. and Roende, T. (2018). High risk or low cost - dichotomous choices of rd strategy by startups in markets for technology. *Working Paper*.

Jaffe, A. (1987). Characterizing the 'technological position' of firms, with application to quantifying technological opportunity and research spillovers. *Research Policy*, 18:87–97.

Klette, J. and Kortum, S. (2004). Innovating firms and aggregate innovation. *Journal of Political Economy*, 112(5):986–1018.

Koning, R., Hasan, S., and Chatterji, A. (2020). Experimentation and startup performance: Evidence from a/b testing. *HBS Working Paper 20-018*.

Lavie, D. and Rosenkopf, L. (2006). Balancing exploration and exploitaton in alliance formation. *The Academy of Management Journal*, 49(2).

Lavie, D., Stettner, U., and Tushman, M. L. (2010). Exploration and exploitation within and across organizations. *the Academy of Management Annals*, 4(1):109–155.

Lee, C. (2020). Buyouts and innovation incentives: The case of the great recession. *Working Paper*.

Levinthal and March (1993). The myopia of learning. *Strategic Management Journal*, 14:95–112.

Manso, G. (2011). Motivating innovation. *Journal of Finance*, 66(5):1823–1860.

March, J. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1):71–87.

O'Reilly, C. and Tushman, M. (2008). Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. *Research in Organizational Behavior*, 28:185–206.

Prais, S. (1958). The statistical conditions for a change in business concentration. *Review of Economics and Statistics*, 40(3):268–272.

Rothaermel and Deeds (2004). Exploration and exploitation alliances in biotechnology: A system of new product development. *Strategic Management Journal*, 25:201–221.

Ruckes, M. and Ronde, T. (2015). Dynamic incentives in organizations: Success and inertia. *The Manchester School*, 83(4):475–497.

Sadler, E. (2017). Innovation adoption and collective experimentation. *Working paper.*

Shapiro, C. (2012). Competition and innovation - did arrow hit the bull's eye? *"The rate and direction of inventive activity revisited. - Chicago, Ill. [u.a.] : Univ. of Chicago Press, ISBN 0-226-47303-1*, pages 361–404.

Smith, W. and Tushman, M. (2005). Managing strategic contradictions: A top management model for managing innovation streams. *Organization Science*, 16(5):522–536.

Sorensen, J. B. and Stuart, T. E. (2000). Aging, obsolescence, and organizational innovation. *Adminstrative Science Quarterly*, 45(1):81–112.

Thomke, S. (2001). Enlightened experimentation. *Harvard Business Review*, 79(2):66–75.

Tian, X. and Wang, T. Y. (2014). Tolerance for failure and corporate innovation. *The Review of Financial Studies*, 27(1):211–255.

Trajtenberg, M., Henderson, R., and Jaffe, A. (1997). University versus corporate patents: A window of the basicness of invention. *Economics of Innovation and New Technology*, 5:19–50.

Tushman, M. and O'Reilly, C. (1996). Ambidextrous organizations: Managing evolutionary and revolutionary change. *California Management Review*, 38(4):8–30.

Uotila, J., Maula, M., Keil, T., and Zahra, S. A. (2009). Exploration and exploitation and financial performance: Analysis of sp 500 corporations. *Strategic Management Journal*, 30:221–231.

Utterbeck, J. and Abernathy, W. J. (1978). Patterns of industrial innovation. *Technology Review*, 80(7):40–47.

Zhou, K. Z. and Li, C. B. (2012). How knowledge affects radical innovation: Knowledge base, market knowledge acquisition, and internal knowledge sharing. *Strategic Management Journal*, 33:1090–1102.

Zhou, K. Z. and Wu, F. (2010). Technological capability strategic flexibility and product innovation. *Strategic Management Journal*, 31(5):547–561.

A Figures

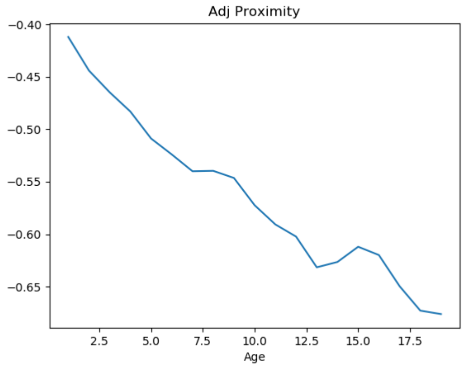

Figure 2: This figure plots the coefficients from the complementary proximity regressions. This represents the average complementary proximity by firm age controlling for industry, year and firm fixed characteristics. The included controls are the firm's initial size quantile, whether it has previous experience, whether it first patented with other applicants and what the firm's initial originality groups is.

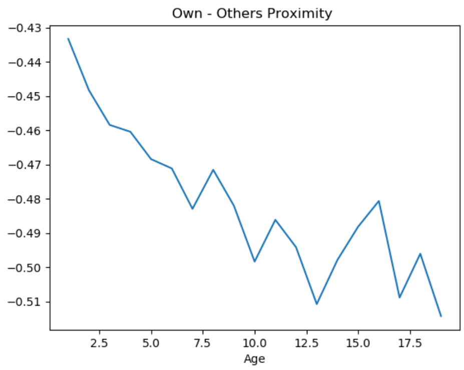

Figure 3: This figure plots the coefficients from the regression with the proximity difference between the entrant and the incumbent as the dependent variable. The additional variables are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it has previous experience, whether it first patented with other applicants and what the firm's initial originality groups is.

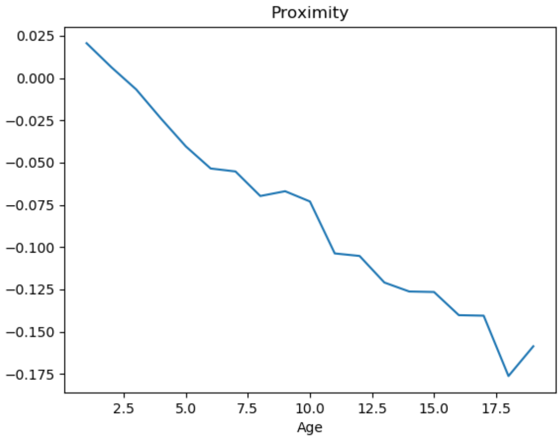

Figure 4: This figure plots the average proximity of the incumbent firms to the entrants initial position by firm age. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it has previous experience, whether it first patented with other applicants and what the firm's initial originality groups is.

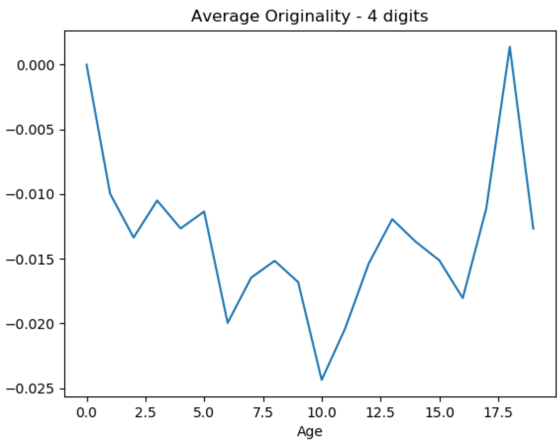

Figure 5: This figure plots the average firm mean originality. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it has previous experience, whether it first patented with other applicants and what the firm's initial originality groups is.

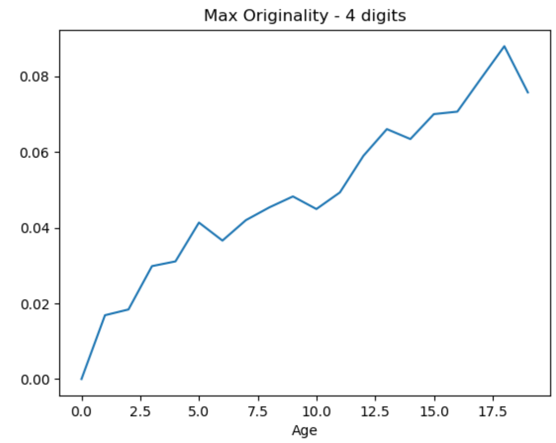

Figure 6: This figure plots the average firm maximum originality. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it has previous experience, whether it first patented with other applicants and what the firm's initial originality groups is.

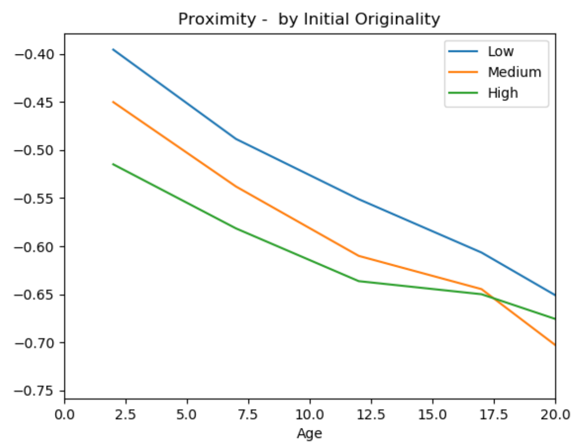

Figure 7: This figure plots the average entrant Jaffe proximity by initial originality category. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it has previous experience and whether it first patented with other applicants.

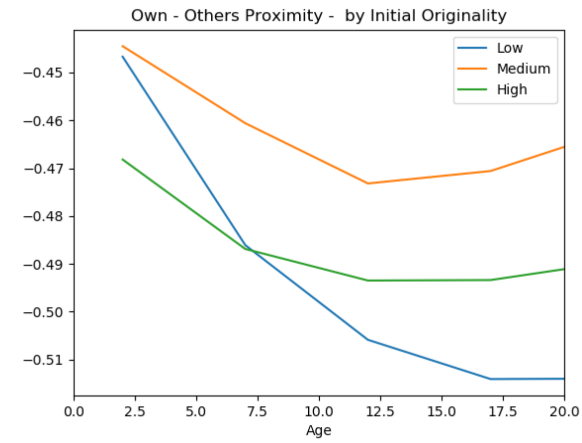

Figure 8: This figure plots the average entrant Jaffe proximity with the incumbent trends removed by initial originality category. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it has previous experience and whether it first patented with other applicants.

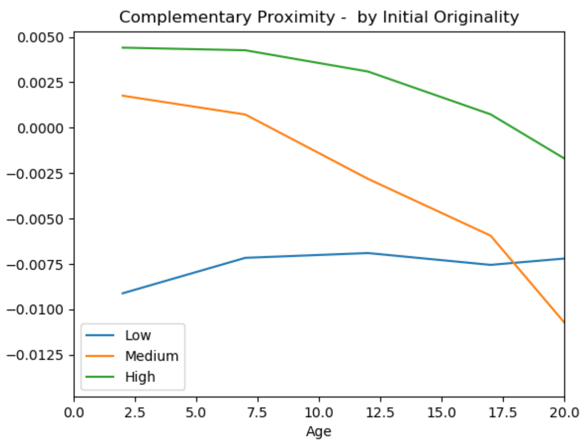

Figure 9: This figure plots the average entrant complementary proximity by initial originality category. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it has previous experience and whether it first patented with other applicants.

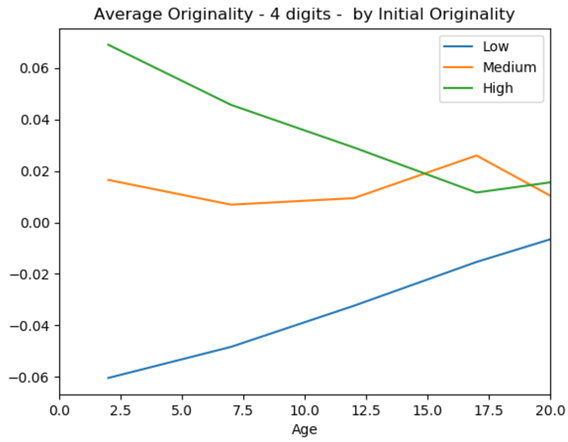

Figure 10: This figure plots the average entrant mean originality. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it has previous experience and whether it first patented with other applicants.

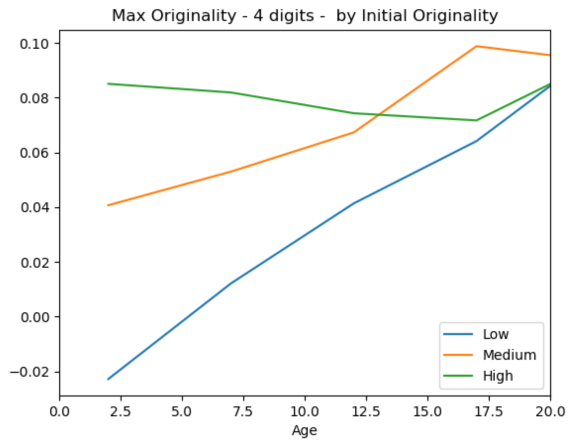

Figure 11: This figure plots the average entrant maximum originality by initial originality category. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it has previous experience and whether it first patented with other applicants.

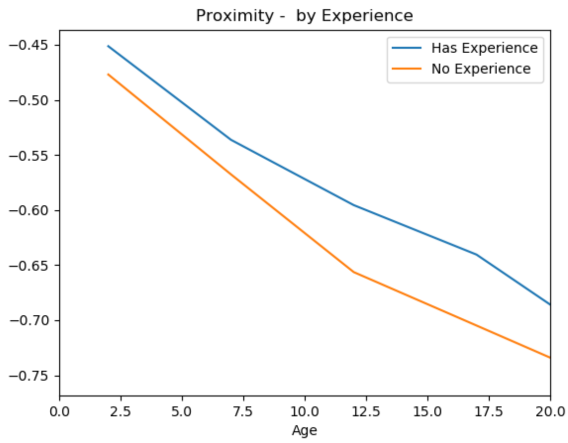

Figure 12: This figure plots the average entrant Jaffe proximity by whether it has previous patenting experience. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it first patented with other applicants, and its initial originality category.

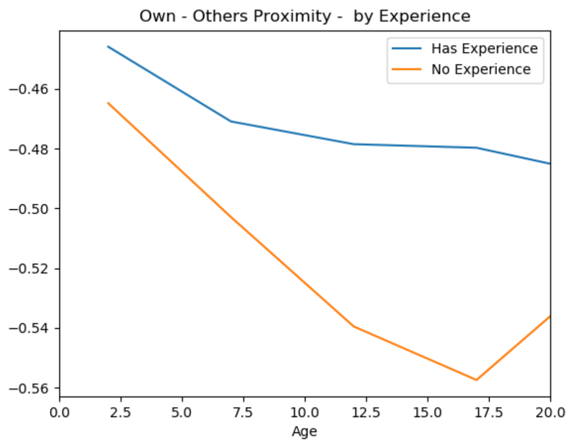

Figure 13: This figure plots the average entrant Jaffe proximity with the incumbent trends removed by previous patenting experience. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it first patented with other applicants, and its initial originality category.

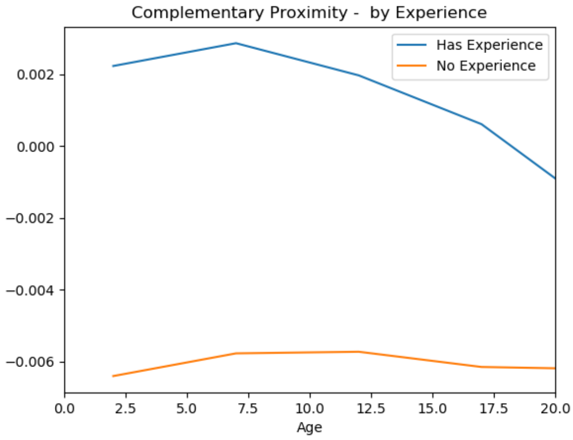

Figure 14: This figure plots the average entrant complementary proximity by previous patenting experience. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it first patented with other applicants, and its initial originality category.

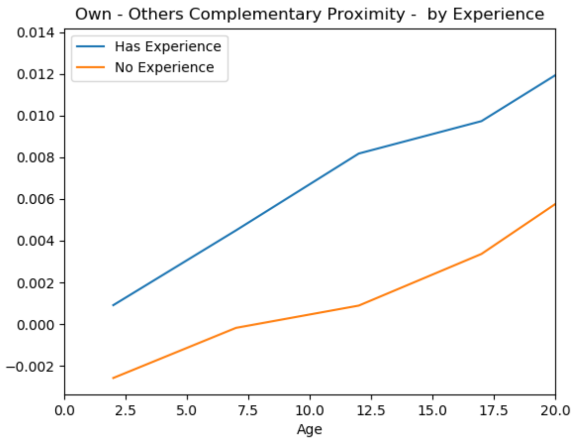

Figure 15: This figure plots the average entrant complementary proximity with incumbent trends removed by previous patenting experience. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it first patented with other applicants, and its initial originality category.

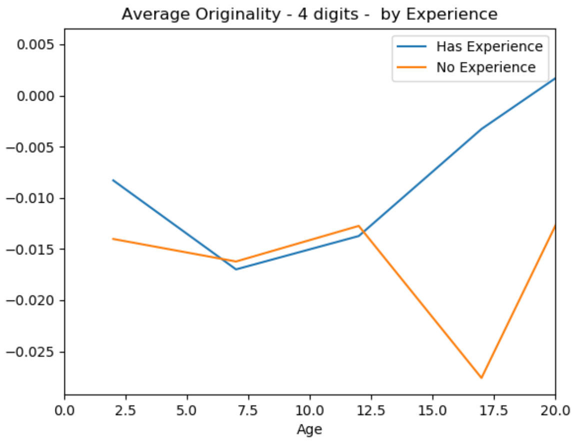

Figure 16: This figure plots the average entrant mean originality by previous patenting experience. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it first patented with other applicants, and its initial originality category.

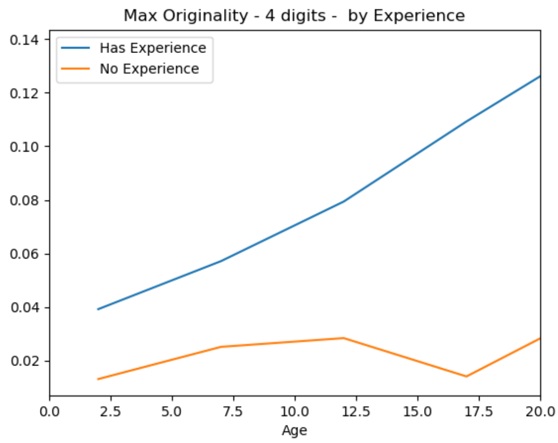

Figure 17: This figure plots the average entrant max originality by previous patenting experience. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, whether it first patented with other applicants, and its initial originality category.

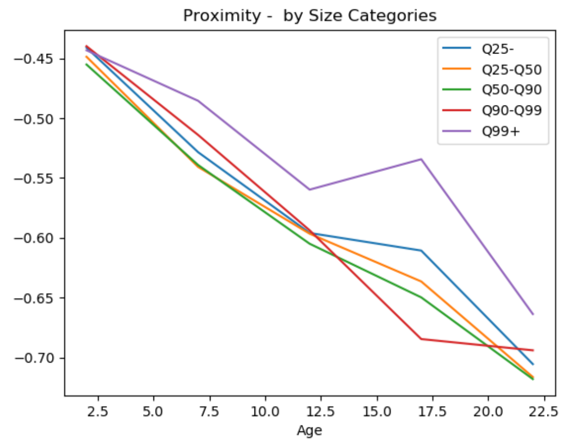

Figure 18: This figure plots the average entrant Jaffe proximity by dynamic size categories. The additional controls are year, the firm's primary 4 digit IPC code, its initial originality category, whether it has previous experience and whether it first patented with other applicants.

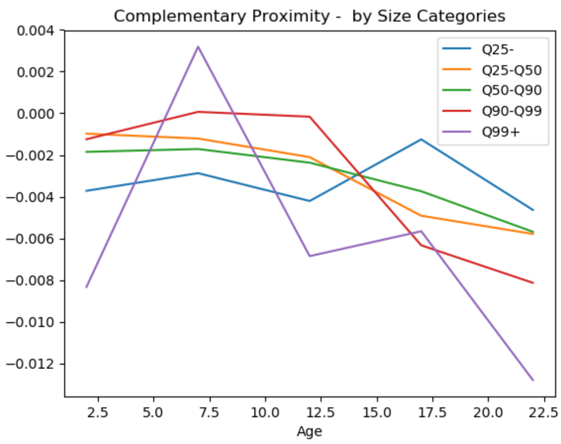

Figure 19: This figure plots the average entrant complementary proximity by dynamic size categories. The additional controls are year, the firm's primary 4 digit IPC code, its initial originality category, whether it has previous experience and whether it first patented with other applicants.

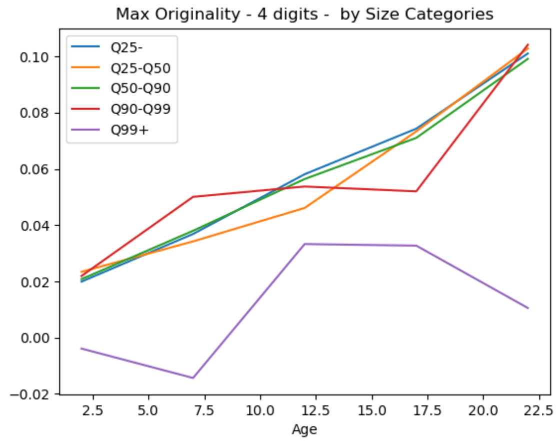

Figure 20: This figure plots the average entrant max originality by dynamic size categories. The additional controls are year, the firm's primary 4 digit IPC code, its initial originality category, whether it has previous experience and whether it first patented with other applicants.

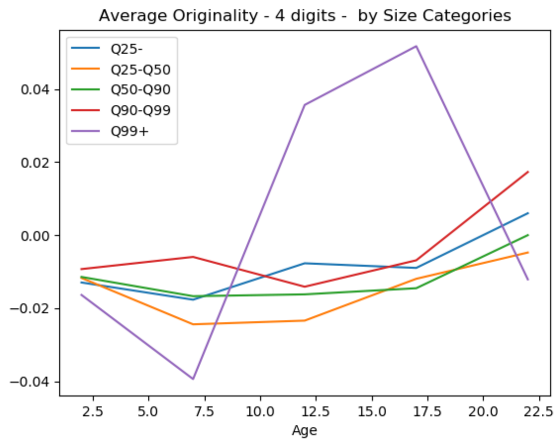

Figure 21: This figure plots the average entrant mean originality by dynamic size categories. The additional controls are year, the firm's primary 4 digit IPC code, its initial originality category, whether it has previous experience and whether it first patented with other applicants.

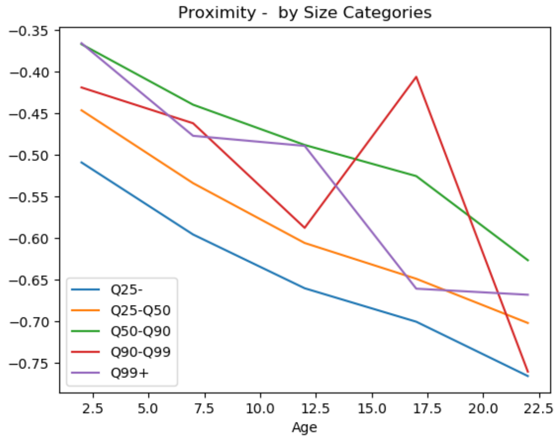

Figure 22: This figure plots the average entrant Jaffe proximity by its fixed initial size category. The additional controls are year, the firm's primary 4 digit IPC code, its initial originality category, whether it has previous experience and whether it first patented with other applicants.

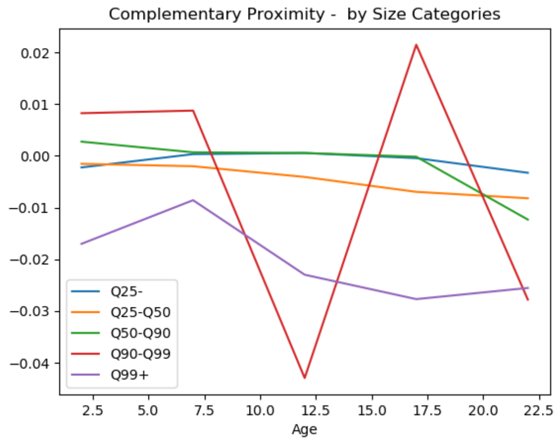

Figure 23: This figure plots the average entrant complementary proximity by its fixed initial size category. The additional controls are year, the firm's primary 4 digit IPC code, its initial originality category, whether it has previous experience and whether it first patented with other applicants.

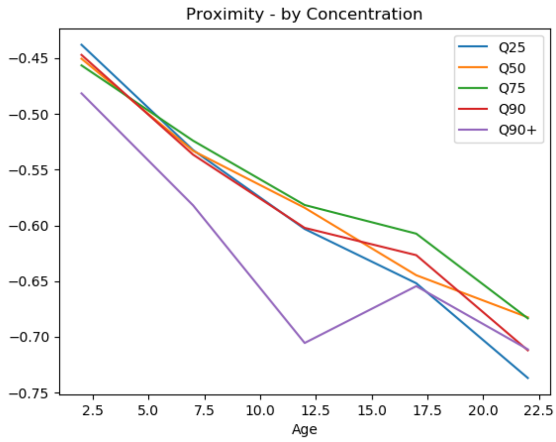

Figure 24: This figure plots the average entrant Jaffe proximity by its dynamic sector concentration quantile. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, its initial originality category, whether it has previous experience and whether it first patented with other applicants.

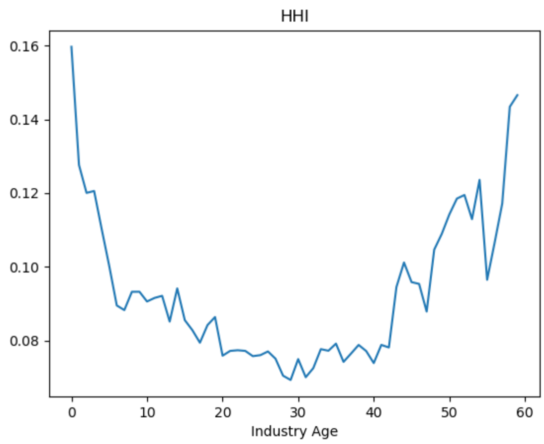

Figure 25: This figure shows the median 4 digit technology sector Herfindahl index where the Herfindahl index is calculated based on firm discounted knowledge stock measures on patent counts.

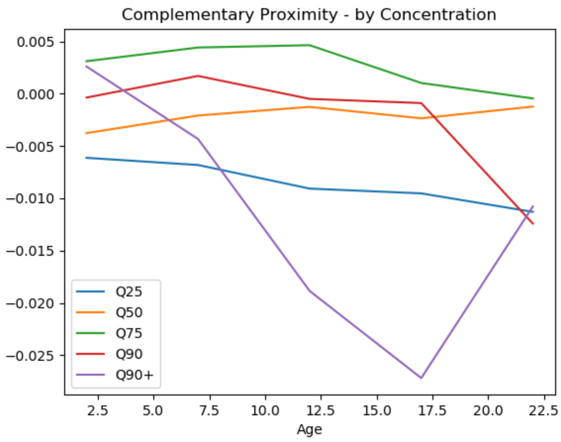

Figure 26: This figure plots the average entrant complementary proximity by its dynamic sector concentration quantile. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, its initial originality category, whether it has previous experience and whether it first patented with other applicants.

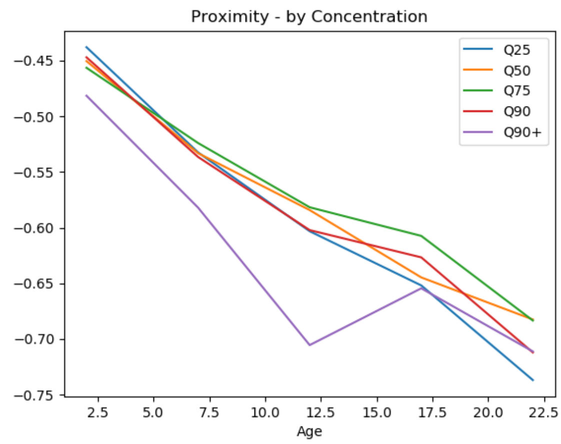

Figure 27: This figure plots the average entrant mean originality by its dynamic sector concentration quantile. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, its initial originality category, whether it has previous experience and whether it first patented with other applicants.

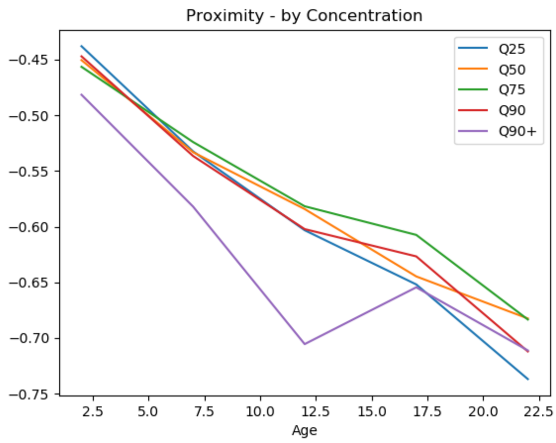

Figure 28: This figure plots the average entrant max originality by its dynamic sector concentration quantile. The additional controls are year, the firm's primary 4 digit IPC code, its initial size quantile, its initial originality category, whether it has previous experience and whether it first patented with other applicants.