Well Begun Is Half Done: Initial R&D Competence and Firm Growth

Kyungran Lee S. Katie Moon Seungjoon Oh*

November 17, 2016

*The University of Hong Kong, the University of Colorado, and Peking University HSBC Business School, respectively. Lee can be reached at kyunglee@hku.hk, Moon at katie.moon@colorado.edu, and Oh at sjoonoh@phbs.pku.edu.cn. We thank Sugato Bhattacharyya, Hursit Selcuk Celil, Lin Chen, Jaewon Choi, Tony Cookson, Diego Garcia, Jayant Kale, Dukhee Lee, Stefan Nagel, Uday Rajan, Brain Waters, Ed Van Wesep, and Jaime Zender for excellent comments and suggestions. We also thank seminar participants at Korea University, Northeastern University, Peking University HSBC Business School, the University of Colorado Boulder, and Yonsei University.

Well Begun Is Half Done: Initial R&D Competence and Firm Growth

ABSTRACT

We examine the effects of initial competence on innovation strategies and growth. Using a detailed project-level data for private firms in the drug sector, we show that R&D competence is persistent, suggesting the large part of the cross-sectional variation in R&D performance is initially determined. We find that firms with high initial competence are more likely to be exploitative in their best segment and grow faster and more successfully. By contrast, firms with low initial competence tend to be explorative, diversifying into multiple segments. Medicare Part D legislation as an exogenous shock to diversification incentives suggests a likely causal relationship.

1 Introduction

A widely circulated practical article on startup and growth defines a startup as a "company designed to grow fast, not just a firm that works on technology, takes venture capital funding, or has some sort of exit".[1] However, to grow (*i.e.,* not to fail), especially to grow fast, is extremely challenging for a startup firm in a highly innovative industry. Holmstrom (1989) considers the particular challenges of innovation in his model and characterizes innovation projects as risky with a high probability of failure and unpredictable as future contingencies are impossible to foresee.

However, existing empirical literature is silent on the probability of failure in innovative projects and the fate of the unsuccessful startup firms in innovative industries. It is mainly because most prior studies use patent data (*e.g.,* Griliches, Pakes, and Hall (1988) and Hall, Jaffe, and Trajtenberg (2001)), which, by definition, registers only on innovation projects that have already succeeded and thus are patentable. Failed or ongoing projects which may better represent the challenges of innovation projects are not observable in patent data.[2] This paper uses detailed project-level new drug development data from BioMedTracker instead, with a focus on ongoing innovation that potentially includes both successful and unsuccessful outcomes. Thus do we attempt to fill the void in the innovation literature regarding how startup firms with initially heterogenous failure probabilities may grow differentially in highly innovative industries.

We begin our analysis by assessing heterogeneous R&D competence across firms. We find that firm competence in developing new drugs, as measured by changes in the progress of individual projects, exhibits remarkable persistence over time; firms with relatively high initial competence (*i.e.,* a low failure rate) are likely to maintain relatively high competence throughout their lifetimes. We focus on this permanent component of R&D competence and examine how this innate competence affects a firm's innovation strategies and growth patterns.

Firm initial conditions are explored in the previous studies, but those studies are very few.

[1] See "Startup = Growth" by Paul Graham at http://www.paulgraham.com/articles.html.

[2] Farre-Mensa, Hegde, and Ljungqvist (2016) overcomes this limitation of patent data by using detailed micro data from the U.S. Patent and Trademark Office (USPTO) on all patent applications filed by startups since 2001. Although their data include both approved or rejected patent applications, whether innovation is always patentable remains as a concern.

Lemmon, Roberts, and Zender (2008) show that the majority of variation in capital structure is driven by an unobserved time-invariant effect, and that initial leverage is the single most important determinant of future capital structure. Maksimovic, Phillips, and Yang (2013) find that the initial difference in firm-specific potential for long-run profitable growth in the early pre-IPO stage explains a firm's public status and acquisition activities subsequent to going public. In this paper, we specifically consider innate firm R&D competence as a fundamental driving component of comparative advantage in producing in a highly innovative industry. More recently, Ayyagari, Demirgüç-Kunt, and Maksimovic (2015) find that the initial size, productivity, and legal form of a start-up in developing countries are persistent and have greater explanatory power in determining growth than the effects from financial institutions.

A similar notion of firm competence is found in the management literature. Hamel and Prahalad (1990) characterize a corporation as a tree the roots of which are its particular competencies and suggest that a firm's core competence is difficult for competitors to imitate because it is a complex harmonization of individual technologies and production skills.[3] Their idea is also consistent with Drucker (2007) who finds firm competence to be based on a set of skills and technologies that include skilled human capital, organizational endowment and culture, and superior technology.[4] Similarly, previous studies of industry life cycles often assume that firms randomly differ in their product innovation expertise (*i.e.,* different capabilities) which influences their success at product R&D (Klepper (1996)). Accordingly, we consider a firm's R&D competence as one of the initial conditions possessed or able to be acquired at the initiation of business that continue to affect the firm's performance and potential throughout the life of the firm.

We capture firm-specific R&D competence based on the extent to which initiated drug projects are suspended, remain active, or successfully proceed to the next phase of the development process. Specifically, we create initial and time-varying (current) annual competence measures that reflect the initial three years and each year, respectively. We find that the current competence measure is highly heterogeneous across firms from the earliest stage and also highly persistent over time. Moreover, R&D competence within the initial three years is

[3]See "Idea: Core Competence", Sep 15th 2008, The Economist, online extra.

[4]See also Henderson and Cockburn (1994), Hallen (2008), Powell and Sandholtz (2012), and Marquis and Tilcsik (2013) for other studies in management on sources of firm competence.

significantly associated with overall future performance with an estimated correlation coefficient of 0.5, thus indicating that our initial R&D competence measure is one of the important initial conditions of firms in highly innovative industries. We especially highlight the role of the initial competence in startup firms' main segment in explaining stark differences in their subsequent growth.

The permanent component of R&D competence not only affects growth such as a firm's public status as in Maksimovic, Phillips, and Yang (2013), but also influences innovation strategies, exploitative versus exploratory. The concept for the two distinct innovation strategies - exploitative versus exploratory - has been developed in the literature (*e.g.,* March (1991), Levinthal and March (1993), and Benner and Tushman (2002)). Benner and Tushman (2002) define exploitative innovation as involving improvements to existing components and advancing the existing technological trajectory, and exploratory innovation as involving a shift to a different technological trajectory. The existing literature makes numerous important contributions to identifying determinants of innovation strategies including managerial myopia in learning (Levinthal and March (1993)), variance seeking (McGrath (2001)), dynamic environment as characterized, for example, by fluctuation in demand and supply, tolerance for early failure (Manso (2011)), and form of equity financing or public market listing (Ferreira, Manso, and Silva (2014) and Gao, Hsu, and Li (2015)) among others. Especially, Ferreira, Manso, and Silva (2014) find that the incentives of public firms are biased towards exploitative innovation, while those of private firms towards explorative innovation. Our paper, distinctively going back to the beginning of a firm, shows that the innovation incentives of startup firms vary between exploitative versus explorative according to their initial conditions with respect to R&D competence.

We consider operating strategies one of the important representations of exploitation versus exploration (*i.e.,* operating in multiple disease groups implies a high level of exploration). We therefore assume that firms with an exploitative innovation strategy would focus on what they have been doing, and that firms with a high level of exploration innovation strategy would diversify into many segments in different industries. Our prediction regarding the relation between initial R&D competence and innovation strategies is that firms with relatively good initial performance in their main segment are more likely to remain focused and develop drugs within the same disease group. It is because a shift to other industries would incur higher opportunity costs to such firms that show high competence in their initial main segment. We find strong evidence that supports for this prediction. This is consistent with Maksimovic and Phillips (2002) and Maksimovic and Phillips (2013), who document that firms with a comparative advantage in producing within an industry have higher growth and attain a larger size in that industry.

In Maksimovic and Phillips (2002)'s model, firms with relatively low comparative advantage are more likely to be diversified and operate in multiple segments. We note that this prediction is likely to be stronger for firms in innovative industries. Success in the highly unpredictable environment of innovation depends on generating sufficient variation that at least some will prove to yield desirable results ex post (McGrath (2001)). Variance-seeking by innovative firms is consistent with the notion that investment in innovation can be viewed as a purchase of real options (Thakor (2013)); under the option-like payoffs of innovation, firms with lower likelihood of in-the-money (*i.e.,* distant from the break-even point) are more likely to take risk in their innovation strategies, as the option value increases with underlying risk. We thus predict, and have strong evidence, that firms with low competence that have consequently been doing poorly in their initial main segment are more likely to enter other industries.

Given that firms that perform poorly in their original industry are likely to explore other industries, we examine into which industries such firms expand. One prediction regarding destination is that those firms will expand into industries previously proven to have higher success rates to be on the safe side. Alternatively, they may expand into industries with less matured projects by competitors and thus higher failure rates. We find that such firms are more likely to enter industries in which the number of incumbents is large (*i.e.,* more competitive industries), incumbent firms are generally less successful in developing new drugs yet, and their products tend to be in the earlier stage of the development process. Such industries can be categorized as the industries in the first (fluid) phase of Abernathy and Utterback (1978)'s industry life cycle model, in which technological and market uncertainties prevail and firms are still in pursuit of product innovations.[5] This finding is consistent with firms that have higher failure rates, potentially due to low comparative advantage

[5]The characteristics of such industries are also consistent with those of the emerging industries in Klepper (1996).

in their original innovation, adopting more explorative innovation strategies not only by increasing the number of operating industries, but also diversifying into relatively younger phase industries.

We next examine the effects of the persistent R&D competence on firm growth including IPO exit and venture capital funding. Going public is considered an important stage in the growth of a company (Pagano, Panetta, and Zingales (1998)), and also the best exit choice, the most successful firms choosing an IPO (Chemmanur, He, He, and Nandy (2015)). Prior studies also find that firms are more likely to go pubic when their expected profitability is highest (Pástor, Taylor, and Veronesi (2009), Chemmanur, He, and Nandy (2009), and Chemmanur, He, He, and Nandy (2015)). We therefore predict a positive link between initial R&D competence and likelihood of going public which implies successful firm growth. Continued or increased venture capital funding is another indicator of successful firm growth pre exit, because venture capital firms and other financial intermediaries are viewed as a mechanism for evaluating prospective entrepreneurs and funding the most promising ones (King and Levine (1993)). Thus, given that R&D competence is persistent, we similarly predict that firms with high initial competence are more likely to receive greater amounts of venture capital funding.

Recalling that exploration is consistent with a value enhancing strategy under the option-like payoffs of innovation, we further examine whether a poorly performing firm's pursuit of an exploitative innovation strategy through diversification has an effect on subsequent firm growth. Although we predict this to increase the likelihood of IPO exit or continued venture capital funding, how to test this effect is unclear, given our finding of a significant relation between initial competence and diversification as well. To disentangle the direct effect of initial R&D competence on firm growth from the potential indirect effect of initial R&D competence on firm growth through diversification, we particularly introduce a mediation model that employs diversification as a mediator. Figure 1 illustrates our mediation model where we hypothesize that initial R&D competence in a firm's main segment influences diversification which, in turn, influences the firm's growth in concert with the direct relation of R&D competence itself to growth.

[Insert Figure 1 Here]

Our results show that both direct and indirect effects of R&D competence on growth are significant. We first find that firms with high R&D competence at their earliest stage are more likely to exit through an IPO earlier or receive more venture capital funding than firms with lower competence (direct effect). Second, we find that firms with relatively low initial R&D competence tend to diversify into other disease groups, thereby also increasing the likelihood to successfully exit through an IPO at the end or receive more venture capital funding pre exit (indirect effect). We further find that the direct effect is much greater than the indirect effect.

We recognize that diversification decisions are endogenous. Thus, we instrument for diversification decisions using Medicare Part D legislation which can affect the incentives of firms, especially those performing poorly in their original main industries, to explore specific new industries related to Medicare Part D with anticipation of a positive demand shock in such industries. We find the predicted result that firms with lower R&D competence increase the likelihood of going public or obtaining additional venture capital funding by diversifying into the specific industries where positive demand shocks are expected under the Medicare Part D program, whereas firms with higher competence are un- (or less) affected by the shock remaining focused on comparative advantages in their respective industries.

Our paper contributes on multiple dimensions to an understanding of how firms' initial conditions affect diversification and growth. First, it adds to the literature that explores firm initial conditions, being the first to show that firms' technological competence is inherently given and highly persistent over time. Second, by using new detailed project-level drug development data, we completely cover both successful and unsuccessful R&D activities related to firm-specific technological competence. Our measure is more precise than those in the previous innovation literature, which, in relying mainly on patent filings or citations data, consider only successful innovation. Our data does have a limitation in that, being only for firms that develop new drugs, the sample size is considerably smaller than what can be drawn from patent data. Notwithstanding the smaller sample size consequent to focusing on a specific industry sector, our findings strongly support our predictions. Lastly, we simultaneously estimate our model assuming, in addition to the direct effect of R&D competence on firm growth, an indirect mediation effect through the diversification channel. Our paper is also the first in the literature to document the economic effect of diversification, in which the indirect effect of R&D competence inheres, on firm growth as manifested in

IPO exits and venture capital funding.

Our paper proceeds as follows. In Section 2, we discuss our data and variables. Section 3 presents summary statistics. In Section 4, we discuss R&D competence more in detail regarding its persistence and also by contrasting skill vs. luck. Section 5 discusses our empirical models and presents results for R&D competence, diversification, and firm growth. Section 6 concludes.

2 Data and Variable Construction

2.1 Drug Project Development Data

We obtain our project-level new drug development data for all publicly and privately held firms in the drug industry sector from the BioMedTracker database. BioMedTracker is a real-time database that identifies biotech and pharmaceutical investment opportunities by assessing drug pipelines and future catalysts. The database tracks drug impact events from 1950s using multiple sources including the FDA approval database, news articles, press releases, company filings to the Securities Exchange Commission (SEC), medical conferences, conference calls, direct communication with companies, and the ClinicalTrials.gov database. Although the FDA provides comprehensive information on the approved drugs including the approval date, the FDA approval database does not disclose in-process information for each ongoing project. Differently, BioMedTracker provides information of all drug pipelines which contains the detailed development phase and the final outcome of each project.

The drug development is closely related to the FDA requirements. It is mainly separated into the pre-clinical research on micro-organisms and animals, and the clinical trials (Phase 1, Phase 2, and Phase 3) on humans. In the pre-clinical stages, new compounds are identified in laboratories and companies perform safety tests for Phase 1. Then, an Investigational New Drug (IND) application should be submitted to the FDA with the information on the effect of the active ingredients and toxicities. Once the IND is approved, the development moves to the clinical phases. The clinical phases include three steps - Phase 1, Phase 2, and Phase 3. In Phase 1, safety and dosing issues are reviewed with healthy volunteers. In Phase 2, the effectiveness of a drug is tested with a small number of people who have a certain disease or condition. Phase 3 conducts a large-scale trials for safety and effectiveness of a drug with several hundreds to 3,000 people. At the end of Phase 3, a New Drug Application (NDA) should be submitted for the FDA's review and final approval. Finally, the FDA thoroughly reviews all of the submitted data with an NDA and considers approving a new drug for marketing.

Our sample primarily comes from these project-level new drug development data for the sample period from 1985 to 2014 according to the data availability from the BioMedTracker database. We note that the data coverage before 2000 in the BioMedTracker database may not be complete. This is because one of the important sources for the phase information of the database is the ClinicalTrials.gov website, but the website was created and publicly available from February 2000. Our results are robust to the further refinement of the sample period from 2000 to 2014. Our final sample size is 3,851 annual firm observations with 799 unique firms.[6] Our sample is a subset of firms in the biotech and pharmaceutical industries that only focus on new drug discovery. The relevant SIC codes for firms in our sample are either 2834 or 2836. Our sample also excludes firms that develop, manufacture, or sell generic drug products. Although our sample is only for firms with new drug development, the number of unique private firms in our sample reaches approximately 65% of the number of private firms in prior studies that use patent data.[7]

2.2 Variable Construction

From the BioMedTracker database we primarily obtain the information on the status of drug development phase including the date of an advance to the next phase or suspension in the middle of the development. We measure firm-specific R&D competence based on whether each project within a firm in a given year is suspended, stays, or advances to the next phase.[8] Distinct from the widely used patent data, our new drug development data includes not only

[6]The number of unique firms in the BioMedTracker database during our sample period is 1,582. Among them, 533 are identified as public firms in their first observations. Among 1,049 private entities, we drop universities, hospitals and firms whose ages are greater than 10 in their first observations. We also drop 19 firms with reverse mergers during the sample period.

[7]For example, Gao, Hsu, and Li (2015) also consider private innovative firms using the patent data and the CapitalIQ database for all industries. Their sample contains 1,221 unique private firms that stay private (829) and also go pubic (392) during the sample period from 1997 to 2008. The total number of firms in our sample for the period after 1997 is about 65% of the number of firms in their sample.

[8]The identification of suspension events became more accurate after the passage of the FDA Amendments Act of 2007 (FDAAA). Section 801 of FDAAA requires firms to submit the results of clinical trials within 12 months after the completion date.

successful projects but also all failed and ongoing projects.[9] Using this unique advantage of the data, we construct our measure of R&D competence that reflects the numbers of events for the advance to the next phase and the suspension of a project. More specifically, we assign -1 for the suspension event, 0 for no change in phase, and +1 for the advance event, and then scale the sum of these scores by the total number of projects in the pipeline in a given year. Thus our competence measure ranges from -1 to +1. It is important to note that a suspension or an advance to the next phase of a project is less likely to be affected by rival firms within industry and thus the firm-specific success rates potentially capture each firm's competence.

This R&D competence measure, how frequently a firm advances, carries on, or suspends its drug development projects, reflects the firm's potential to grow in the drug industry sector. Whether a firm successfully develops a new drug and finally gains a FDA approval to market the drug is vital to a firm's success and continuation. In the middle of clinical trial phases, however, trials can be suspended at any time for a number of reasons. Firms may voluntarily suspend or terminate their clinical trials at any time when they believe that their drugs present any unacceptable risk to participants including significant side effects or their trial results do not show the expected effectiveness. Furthermore, if the trial results are not successful, firms cannot proceed to the next trial phase. Also, regulatory agencies can order a suspension if they believe that the clinical trials are not being conducted in accordance with applicable regulatory requirements or the trials present an unacceptable safety risk to participants. Suspended projects can be resumed with a new clinical trial design or a modification of the existing trial design when the issues are in the experimental design, but we note from the data that resuming a suspended project is not a common event.

According to a study on the cost of developing a new drug by the Tufts Center for the Study of Drug Development, the clinical phase transition probabilities from Phase I to Phase II is 59.52%, Phase II to Phase III is 35.52%, Phase III to a FDA submission is 61.95%, and a FDA submission to final approval is 90.35%. Overall, the clinical approval success rate

[9]The NBER patent database (Hall, Jaffe, and Trajtenberg (2001)) provides patent filing and citation information that is widely used as an innovation measure in the literature. We note, however, that more matured firms are likely to patent their ideas. We match startup firms in our sample to firms in the NBER patent data. Only 114 firms out of 360 firms (32%) in our sample with founding years prior to 2007 are matched to the firms in the patent data. Also, the correlation coefficient between our competence measure and the number of patents or citations is negative but insignificant.

is only 11.83% and historically the rates have declined significantly. Drug development is a long process with more than 10 years of the average development time from synthesis to an approval and with 16 to 31 months of the average phase transition time. The estimated total cost associated with new drug development is $2.9 billion, and the cost is also increasing over time. Due to the increases in the costs, continuing the drug development without suspensions and proceeding to the next phase quickly are critical to the success and continuation of a startup firm in the drug industry sector. The average number of projects that a firm in our sample has is only 2.2 and the median is 1 (the detailed summary statistics are presented in Table 2). This also indicates how a firm can quickly advance its new drug development without any suspension can be a relevant measure of firm R&D competence.[10]

Our R&D pipeline data spans 30 years and 21 different disease classifications, which allows us to effectively consider a firm's initial R&D competence within their very first years using our hand-collected founding year data, and also main versus non-main segment competence using the specific disease code information. We use the average of the first three non-missing R&D competence values to construct a firm's initial competence. We define a firm's major segment (disease group) based on where the firm has the largest number of projects. Then we construct a firm's main R&D competence using the average R&D competence of the firm's main segment. Throughout this paper, we focus on a firm's initial R&D competence in its main disease group. Figure 2 shows distribution of the initial main R&D competence measure. The variable ranges from -1 to +1 by construction and is highly concentrated at zero and roughly shows a bell shape.

[Insert Figure 2 Here]

We also importantly consider a firm's diversification, as we interpret diversification to multiple disease groups as one of the important representations of explorative innovation strategies. Our diversification measure is based on how a firm creates its project portfolio across different disease groups. It is the number of all different disease groups where a firm

[10]We first assume that failure rates across phases are identical as 50% and thus assign -1 for a suspension event and +1 for an advance event regardless of the phases. As robustness, instead of using -1 and +1, we consider other weighting scheme that takes account of phase-specific difficulties. For this, we use the phase-specific success rate estimates in the study by the Tuft Center available at `http://csdd.tufts.edu/files/uploads/Tufts_CSDD_briefing_on_RD_cost_study_-_Nov_18,_2014..pdf` or in DiMasi, Hansen, and Grabowski (2003). For example, when the success rate from Phase II to Phase III is 36%, we assign -0.72 (=-36/50) for a suspension event, 0 for no change, and +1.28 (=64/50) for an advance event. Our results are robust to using any set of phase-specific success rates in the prior studies.

has new drug projects in a given year. Alternatively, we also measure a firm's diversification by considering relative project shares of each disease group. It is calculated as one minus the sum of squared project share of each disease group in a given year. The project share of each disease group is the number of projects in the disease group divided by the number of total projects over all disease groups. For example, a firm with one project in each of two different disease groups has diversification measure of 0.5 ($1 - (0.5)^2 - (0.5)^2 = 0.5$).

Our industry classification is based on the 21 different disease classifications that the BioMedTracker database identifies. The main industry of a firm is set as an industry where the firm has the largest number of projects in a given year. If there are more than one industries with the largest number of projects, we randomly select one industry. Table 1 provides the list of our 21 industry groups. Oncology is the most common disease area in our sample (29.52%), and then Neurology (14.89%), Cardiovascular (7.78%), and Endocrine (7.82%) follow.

[Insert Table 1 Here]

We consider two different measures that potentially capture overall firm growth - IPOs and venture capital funding. Going public is considered as an important stage in the growth of a firm and the best exit choice where the most successful startup firms choose to have.[11] Also, VC funding is an important indicator of growth. VC funding is an important external financing that is provided by venture capitalists to seed early stage startup companies generally with a novel technology or business model. The previous VC literature suggests that VC funding increases the performance of a firm and thus the likelihood to exit private (*e.g.,* Lerner (1995) and Kortum and Lerner (2000)). Our VC funding data comes from the VentureXpert provided by the SDC database. We consider either a firm-level dummy variable of VC funding that equals one if the firm has ever received funding from VC investors or Log(1+VC funding), defined as the log of one plus the amount of VC funding in million dollars. Our IPO data comes from the SDC New Issues database. We construct a firm-year level dummy variable of an IPO event that equals one if a firm goes public in a given year.

We control other relevant firm-specific characteristics that may potentially affect firm R&D competence, diversification and firm growth. Those control variables include the num-

[11] In our later analysis, we consider both IPOs and acquisitions as successful exits of startup firms. Among 799 unique firms, 48 firms have exits via acquisitions. Among 48, 17 (35%) are post-IPO acquisitions.

ber of phase 0 projects, the composition of R&D projects (early vs matured), the presence of partner organizations, and the firm age. We classify early- or matured-stage projects based on their phase information. The early-stage includes phase 1, phase 2, and phase 3 projects, and the matured-stage includes phase 4 (NDA/BLA) and phase 5 (Approved). The composition of R&D projects potentially captures growth opportunities of a firm. In the drug industry sector especially, firms often team up with so called partners or sponsors to be supported technically, financially or with regulatory expertise. Therefore, the existence of partner organizations can affect the success rate of R&D projects. The BioMedTracker database provides the names of partner organizations for each project, if exists. We create a dummy variable of the presence of partner organizations for each project and then calculate the fraction of the partnered projects over all projects. We expect that firms with more partnered projects are likely to show better performance and thus more likely to grow successfully. Lastly, firm age may capture any differences in the development stage based on tenure. These firm-specific control variables help to mitigate remaining concerns with our R&D competence measure. For example, remaining concerns include possibilities that firms with a small number of total projects likely show higher initial competence, and that firms may delay entering markets until they have successfully matured projects and thus higher initial competence.

We also control the effects of the industry characteristics including the total number of firms with new drug development projects in a given industry, the industry's overall failure rate, and the industry's composition of R&D projects alongside industry fixed effects. We consider the number of all firms with new drug development in each industry as one of the proxies for R&D competition. We also consider the overall failure rate within industry, which potentially indicates the difficulty of developing new drugs for the specific disease group. Lastly, the industry's composition of R&D projects is the fraction of matured projects within industry. Although our main focus is on private firms, these three industry variables are based on our extended sample that takes into account all privately and publicly held companies that intend to market their drugs in U.S. This is more accurate way of calculating industry conditions than using either private or public firm sample. The Appendix describes all our variables more in detail.

We depict the time trends of the important variables described above to show any significant economic changes over time in the drug industry sector. Panel (a) in Figure 3 shows the time trends of the firm-specific variables and Panel (b) shows the time trends of the industry-specific variables. In Panel (a), the average number of total projects shows a gradual U shape. The number was the lowest around 2005 and has been increasing after that. Differently, the average numbers of matured projects and projects with partners continuously decrease over time. The decreasing number of matured projects possibly indicates the increasing difficulties or complexities in developing new drugs over time.

[Insert Figure 3 Here]

In Panel (b), the sample includes both publicly- and privately-held companies in the drug industry sector. The number of competitors (total firms) is increasing over time, indicating an increase in the overall industry size and also an increase in competition within industry. The industry R&D competence measure representing industry wide performance slightly decreases over the sample period. There is a significant drop in R&D competence around 2008 and 2009 possibly due to the financial crisis, but the measure is overall stable in other times. This indicates that it becomes more difficult to successfully develop new drugs over time in general. As in Panel (a), we find that the industry-wide percentage of matured projects also declines.

3 Summary Statistics

Table 2 presents summary statistics of the variables used in our analyses. The sample consists of 3,851 firm-year observations of 799 private firms in the drug industry sector during our sample period from 1985 to 2014. The main focus of our analyses is on diversification and firm growth determined by initial R&D competence of a firm. Our diversification measure with the number of different industries is 1.38 on average and the median firm in our sample is not diversified. 14.2 % of our sample firms went public at the end. 52.2 % of firm year observations have VC funding and the average amount of VC funding is 3.812 million dollars.

[Insert Table 2 Here]

One possible concern regarding our initial competence measure is that firms may have different risk-taking preferences and thus some firms could initially start with multiple projects simply anticipating luck, which does not necessarily indicate true firm competence. The summary statistics in Table 2 mitigate this concern in our competence measure. Our summary statistics show that firms maintain 2.2 projects on average, where only 8.7% of projects are in the matured stages. The low number of total projects is consistent with high costs in developing a new drug as we discuss in the previous section. This may preclude the possible explanation based on luck.

Table 2 also shows that firm overall R&D competence in our sample is -0.064 and -0.063 if we only consider main segments. The negative R&D competence indicates that the number of projects suspended is larger than the number of projects advanced to the next phase on average. This is also consistent with the difficulties in developing a new drug. We also note that firms are more likely to stay in the same phase without frequent events of suspension or advance. The initial competence measure (the average of the first three initial observations) for main segments is 0.011 on average with a standard deviation of 0.235. Firm age is on average 7 years, thereby indicating our sample well captures startup firms in the drug industry sector. Approximately, 9% of projects are in the matured stage. Also, about 40% of projects have partner firms.

All our industry variables are based on firms' main segments. The average number of competing firms in a firm's main industry that include both privately- and publicly-held companies is approximately 50. The average industry-wide failure rate is 21.7% with a standard deviation of 10.4%. In an untabulated result, Orthopedics, Gastroenterology, and Ophthalmology industries' suspension rates are lowest as approximately 15 %, while suspension rates in Psychiatry, Cardiovascular, and Respiratory industries are highest as 25%. The average fraction of matured projects at the industry level is 23.3% with a standard deviation of 18.0%.

Next, in Table 3, we compare our variables between firms that have high versus low initial main competence in Panel A, and between firms that are diversified versus focused in their operations in Panel B. In Panel A, the sample consists of 3,851 firm year observations of 799 firms during the sample period of 1985-2014, while in Panel B the sample consists of 3,091 firm year observations of 646 firms that have low initial main competence. Firms with low initial main competence are the firms that have nonpositive initial main segment competence. A firm is diversified if the firm develops new drugs over multiple disease groups, and focused otherwise.

[Insert Table 3 Here]

In Table 3, Panel A shows that nearly all our variables are significantly different between the two groups. First, firms with high initial main competence are more likely to go public or get VC funding during our sample period. 20.7 % of firms with high main competence go public through IPO, whereas only 12.6 % of firms with low main competence go public. VC funding amounts are also significantly greater for the high competence group. 59.2% of firm years within high main competence group get VC funding, whereas only 50.5% of firm years within low main competence group get VC funding. The log amount of VC funding is 0.56 (1.75 million dollars) for the high group and 0.42 (1.52 million dollars) for the low group. High competence firms have the greater number of total projects but less matured-stage projects. High main segment competence firms are less likely to have projects jointly with partners.

In Panel B of Table 3, we compare between diversified and focused firms within the subset of low main competence firms. Firms in the drug industry sector are in general focused, because roughly 77% of firm-year observations in our sample period are in the focused status. The table shows that the diversified firms are also more likely to go public and get VC funding during our sample period. Approximately 4.4% more firms in the diversified group go public than firms in the focused group. VC funding amounts are also significantly larger for the diversified group. Diversified firms are about 8.2% more likely to get VC funding. It is also worth noting that diversified firms in low main competence group have both greater numbers of total projects and more matured-stage projects compared to the focused firms. Overall results indicate that there is heterogeneity in innovation strategies and subsequent growth patterns even within the group of firms with low initial main competence.

[Insert Figure 4 Here]

Figure 4 graphically confirms our findings in the table. The figure illustrates the time trends of the percentage firms of going public and the amount of VC funding over firm age. In Panel (a) and (b), the sample is split by the initial competence of main disease group, and in Panel (c) and (d) by diversification. Panel (a) shows that firms with high initial main competence are more likely to go public in the first 10 years and earlier than the firms with low initial competence in general. Firms with high initial main competence are the most likely to go public at the age of 5 and 6, but the likelihood stays relatively stable for the firms with low initial competence. Panel (b) shows that firms with high initial main competence continue to get the greater amount of VC funding at least for the first 10 years. VC funding amount is especially greater for the high competence group in the early years of the firm life. The VC funding amounts for the low competence group are smaller compared to the high competence group and do not fluctuate as much. The gap in the VC funding amount between the two groups becomes smaller as firms get older. It is partly because high competence firms tend to go public earlier and drop out from the sample.

Panels (c) and (d) present analogous results for focused and diversified firms within the subset of firms with low initial main competence. The figures show that, for the subset of firms with low initial main competence, diversified firms are more likely to go public and get the greater amount of VC funding on average. In Panel (c), focused firms are more likely to go public in the first two years. It is potentially because the relatively better firms within the group of low initial main competence are likely to stay focused than other firms. Also, in Panel (d), the converging pattern of VC funding at the later stage is likely associated with the fact that relatively better diversified firms with higher competence tend to go public earlier and thus drop out from the sample.

4 R&D Competence

4.1 Persistence

This section examines whether the initial R&D competence is persistent over time and the R&D performance at the beginning of a firm's life significantly determines the firm's subsequent on-going R&D performance. Our predictions and following tests are to show whether we can consider R&D competence as one of the firm initial conditions that a firm already had or could attain at the beginning of its life.

[Insert Figure 5 Here]

Figure 5 presents the evolution of firms' overall R&D competence over firm age. We split our sample by initial main R&D competence. Main competence is the R&D competence of the firm's main segment and we take the average of the first three non-missing values of main competence for the initial main R&D competence. High main competence is the group of the firms with positive initial main competence and low main competence is for the group of the firms with non-positive initial main competence. We find that the firm overall R&D competence does not fluctuate much in both high and low main competence groups. The firms in the high initial main competence group maintain their R&D competence positive. Also the firms in the low initial main competence group stay low all the time during the period that spans more than 10 years of firm life.

[Insert Table 4 Here]

Table 4 confirms the persistency of R&D competence that we find in Figure 5 using various regression models. Specifically, we estimate the following specification:

$$\begin{aligned} \textit{Main R\&D competence}_{i,t} &= \alpha + \beta_1 \textit{Initial main R\&D competence}_i + \beta_2 \textit{R\&D characteristics}_{i,t} \\ &+ \beta_3 \textit{Other firm characteristics}_{i,t} + \beta_4 \textit{Industry characteristics}_{i,t} + \alpha_i + \alpha_t + \epsilon_{i,t}, \end{aligned}$$

where α_i and α_t capture the industry fixed effects and year fixed effects, respectively. Before estimating the model, we first begin by showing the effect of firm and year fixed effects in predicting current R&D competence. Column one of the table shows that firm fixed effect alone explains 30% of current main R&D competence. This supports that in R&D competence there exists a permanent component and the permanent component is firm-specific nature. Column two compares the explanatory power of year fixed effects to that of firm fixed effects on current main R&D competence. We find the explanatory power of firm fixed effects is significantly greater than that of year fixed effects. Column three shows that combination of firm and year fixed effects explains 37.5% of current main R&D competence.

Then, as in the above model, we add variables that potentially affect R&D competence except the initial main R&D competence measure (our main variable of interest) as independent variables in column four. These variables include the number of phase 0 projects, firm age, percentage of matured projects, percentage of projects that have outside partners, number of competitors that also develop new drugs in the same industry, industry failure rate, and industry percentage of matured projects. These control variables are categorized into the following three groups: The R&D characteristics include the number of phase 0 projects, percentage of matured projects, and percentage of projects with partner. Other firm characteristics include firm age. Lastly, industry R&D characteristics include the number of firms in the same industry as a measure of competition, industry failure rate, and industry percentage of matured projects. The result in column four shows that these newly added independent variables also explain the variation of current main competence, but the increase in the adjusted R-square is just 1% relative to that in column three.

Next, we regress the current main R&D competence on the initial main R&D competence. Because the initial main R&D competence is the firm level variable, we include the industry fixed effects instead of the firm fixed effects. In column five and six, we find that initial main competence significantly explains the current main competence even after controlling for other variables. Overall, we find from the table that initial R&D competence persists over time and shows statistically significant and positive relation to the firm's future competence in the long periods of time.

4.2 Skill vs. Luck

Next, we take a deeper consideration of the notion of firm competence, especially whether firm competence is indeed firm skills compared to luck. One might argue that firm initial competence can be random, and the randomly achieved good luck for the first project might result in an ultimate firm success. We discuss this argument based on the mechanism motivated by a bandit problem and present evidence that initial competence is more likely to be firm skills.

In probability theory, a bandit problem is a setting in which a gambler sequentially pulls arms of a slot machine to maximize his/her expected total reward. The reward distribution of each arm is generally assumed to be independent and identically distributed ($i.i.d$) from a known distribution with unknown parameters. A basic formulation is Bernoulli k-armed bandit model, where the reward of each arm is one with probability p_k and zero otherwise. One of the important results of the model is the "stay-on-a winner" principle in Berry (1972). This principle states that if an arm proves to be successful at a stage, selecting the arm again is optimal at the following stage and the ultimate outcome is therefore path-dependent.

Bernoulli k-armed bandit problem can be translated into our context where a startup firm in our sample randomly selects one initial project (arm) among available k projects and the distribution of each project's success rate determines the firm's switching behavior among projects and ultimate success. The standard solution of Bernoulli k-armed bandit problem, stay-on-a-winner principle also applies to our problem. If a firm selects a new drug project with higher chance of a good outcome, the firm will stay with the project and persistently show high competence. Furthermore, the firm is also more likely to exit successfully with an IPO or receive greater amounts of venture capital funding. Hence, the ultimate success of the firm is path-dependent and significantly relies on the initial random selection of a project (*i.e.,* luck) not firm skills.

However, in the classical bandit problem, the assumption of independent arms is critical. When the rewards of arms are correlated at a different degree for each player, the final total rewards also vary across players. In Bernoulli k-armed bandit problem with *dependent* arms, the total reward is always greater when the correlation among arms is higher (Pandey, Chakrabarti, and Agarwal (2007)). The interpretation of this result in our setting is that if firms are heterogeneous in the correlation among the available projects, a firm with higher correlation (*i.e.,* skills) will always perform better in any stage than a firm with low correlation. We empirically test for this prediction using within-firm project performance correlations for a given time. Table 5 present the results.

[Insert Table 5 Here]

The dependent variable is performance of one project within a firm for a given year, an indicator variable that equals one if a firm's project advances to the next phase, minus one if the project is suspended, and zero otherwise. Peer project performance is a variable of interest, that is the average of all other projects' performance for the same firm in the given year. We examine whether the correlation between project performance within a firm is significant and positive, and whether the correlation is greater for high competence firms than low competence firms.

In column one of the table, we find that performance of a project is significantly associated with other projects' performance within the same firm in the same year. The average correlation between the projects within a firm is estimated as 13.9% and significant at the 1% level. In column two, we additionally include initial main competence and its interaction term with peer project performance, to examine whether the within-firm project correlation increases or decreases with firm initial competence. We find that firms with higher initial competence show greater correlation for within-firm projects performance, as the interaction term is significantly positive at the 1% level. We also consider a discrete version of initial main competence in column three and find the similar result that higher competence is significantly related to higher correlation between the project performance. Overall, the results in this table suggest that firms with high initial competence are more likely to perform better almost always due to the reason related to performance correlation within their own projects (*i.e.,* firm skills). Hence, our results are not merely driven by a luck-based explanation.

5 Empirical Methodology and Results

The key question is whether firm initial conditions, especially in R&D competence, explain the differential patterns of innovation strategies through diversification, and firm growth through an IPO exit or VC funding. We employ multiple different empirical methodologies to explore the direct and indirect relations between R&D competence and diversification, and firm growth.

First, we examine the effect of the initial R&D competence on diversification. We predict that higher opportunity costs prevent firms with high R&D competence in their main segment from diversifying into other industries. This prediction is associated with exploitative (explorative) innovation strategies by firms with higher (lower) competence. Second, we simultaneously estimate the effects of initial R&D competence on diversification and firm growth, and also the effect of diversification on firm growth. This empirical design that uses a simultaneous mediation model with diversification as a mediator enables us to disentangle the direct effect of R&D competence and the indirect effect of R&D competence through the diversification channel on firm growth. Lastly, we address potential endogeneity concerns between diversification decisions and firm growth, using the Medicare Part D legislation in 2003 as a shock to diversification incentives in firms with low R&D competence.

5.1 Initial R&D Competence and Diversification

In this section, we examine whether a firm's R&D competence in its main segment affects the firm's decision to diversify into other industries. We assess how initial main R&D competence or current main R&D competence affects diversification. We specifically estimate the following specification:

$$\begin{aligned} Diversification_{i,t} &= \alpha + \beta_1 Main\ R\&D\ competence_{i,t} + \beta_2 R\&D\ characteristics_{i,t} \\ &+ \beta_3 Other\ firm\ characteristics_{i,t} + \beta_4 Industry\ characteristics_{i,t} + \alpha_i + \alpha_t + \epsilon_{i,t}, \end{aligned}$$

where *Main R&D competence* is either the initial main R&D competence or the current main R&D competence based on specifications. *Diversification* is either the number of all different disease groups where a firm has new drug projects or one minus the sum of squared project share of each disease group in a given year. α_i and α_t capture the industry fixed effects and year fixed effects, respectively. We regress our time-varying firm diversification measure on the variables that potentially affect the diversification decision including the main R&D competence (our main variable of interest), firm age, number of phase 0 projects, percentage of matured projects, percentage of projects that have outside partners, indicator of VC financing, number of competitors in the same industry, industry failure rate, and industry percentage of matured projects. All control variables are one year lagged.

Table 6 presents the results from the above regression specification. In columns one and two, we use the diversification measure that is the total number of all disease groups where the firm has projects as a dependent variable. In columns three and four, we use the other diversification measure calculated by one minus the sum of squared project share of each disease group in a given year as a dependent variable. Columns one and three examine the effect of the initial main R&D competence (*Initial main competence*), and columns two and four examine the effect of the current main competence (*Main competence*).

[Insert Table 6 Here]

From columns one and three, we find that firms that have high initial competence in their main segment are less likely to diversify into other industries. The results are statistically and economically significant. A one standard deviation decrease in the initial R&D competence in a firm's main segment results in an increase in diversification by five percentage point (0.235*-0.206) as in column one, and by one percentage point (0.235*-0.0399) as in column three. This is consistent with our prediction that firms with high initial competence in their initial main segment are more likely to stay focused due to the high opportunity costs of entering into other industries. In columns two and four, we also find the similar results using the current main R&D competence instead, but the magnitudes of the coefficients are smaller than those of the initial main competence in columns one and three.

Table 6 also shows that other firm R&D characteristics and industry characteristics affect diversification decisions in several ways. First, competition in a firm's main industry increases diversification, as shown by the positive coefficients of the industry competition variable as measured by the total number of competing firms in the same industry. Second, firm R&D progresses as shown by the number of phase 0 projects and the percentage of matured projects are positively associate with diversification decisions. Firm age is also positively associated with diversification, indicating that firms start with a single segment in their early years and tend to diversify into multiple other industries as they grow older. The coefficients of the percentage of projects with partners and VC backing indicator are also positive, implying that firms with more operational and financial resources are more likely to diversify.

5.2 Diversification Patterns

Next, we further examine which industries firms with relatively low initial competence in their original main industry expand their business into. Table 7 presents summary statistics on characteristics of the destination industries where the diversifying firms newly enter. We find that 165 firms among 799 firms in our sample expand to other disease groups, with 220 unique diversification events. We examine the following relevant industry characteristics to better understand their diversification patterns. This is important to shed light on the fate of the unsuccessful startup firms in innovative industries, which has not been developed in the literature.

[**Insert Table 7 Here**]

We define the following indicator variables that equal to one if the industry average is above the median value for all industries. Those variables include industry R&D competence (High competence), project suspension (High suspend rate) and advance rates (High advance rate), percentage of matured projects (More matured product), percentage of Phase 0 product (More phase 0 product), competition (High Competition), number of industry competitors (More incumbents), and number of products (More products).

We find from the table that firms tend to enter into the industries with higher suspension rates, lower advance rates, more total products, more phase 0 products, higher competition, but less matured. All these patterns are significant at the 1% level. These stylized facts are consistent with unsuccessful startup firms in their initial main industries newly starting their businesses in other emerging industries. The high suspension rates, more products, and more products in the early stages are the characteristics of the first phase (fluid phase) industries in the industry life cycles models such as the models by Abernathy and Utterback (1978) and Klepper (1996).

5.3 Initial R&D Competence, Diversification and Firm Growth: Simultaneous Analysis

In this section, we simultaneously analyze the effects of initial R&D competence and diversification on firm growth. We use two different measures of firms growth including going public and VC funding. We use a system of linear equations that consider two dependent variables that are *Diversification* and *Going public* or *VC funding.* The dependent variable of the first equation, *Diversification* is continuous and the second dependent variable, *Going public* is a dummy variable (*VC funding* is continuous). To precisely incorporate the potential correlation in the residuals between the two equations, we estimate the system of equations using seemingly unrelated regressions (SUR). Specifically, we estimate the following specifications:

$$\begin{aligned} \mathit{Diversification}_{i,t} &= v + \beta_1 \mathit{Initial\ main\ competence}_{i,t} + \beta_2 \mathit{R\&D\ characteristics}_{i,t} \\ &+ \beta_3 \mathit{Other\ firm\ characteristics}_{i,t} + \beta_4 \mathit{Industry\ characteristics}_{i,t} + v_i + v_t + \epsilon_{i,t} \end{aligned}$$

$$\begin{aligned} \mathit{Going\ public\ or\ VC\ funding}_{i,t} = w &+ \gamma_1 \mathit{Initial\ main\ competence}_{i,t} + \gamma_2 \mathit{Diversification}_{i,t} \\ &+ \gamma_3 \mathit{R\&D\ characteristics}_{i,t} + \gamma_4 \mathit{Other\ Firm\ characteristics}_{i,t} \\ &+ \gamma_5 \mathit{Industry\ characteristics}_{i,t} + w_i + w_t + \eta_{i,t}, \end{aligned}$$

where *Diversification* is the firm-year level diversification measure calculated by the total number of different disease groups where the firm has projects in a given year, *Going public* is one if the firm went public via an IPO in a given year, and *VC funding* is Log (1+VC funding amount) in a given year. The same set of control variables as in Table 6 is included, except the VC backing indicator for the VC funding regression. As before, all control variables are one year lagged.

Table 8 reports the effects of initial main segment competence on either going public or getting VC funding directly and indirectly through firm diversification using a simultaneous mediation model. The results in columns one and three show that the diversification intensity declines with the initial main competence. The coefficients of *Initial main competence* is negative and statistically significant at the 1%. This is consistent with the result in Table 6, implying that low initial competence in a firm's main segment drives the firm to diversify into other industries.

[Insert Table 8 Here]

In column two, we find that both initial main competence and diversification increase the likelihood of going public. The initial R&D competence in the main segment alone significantly increases the probability of going public. This direct effect of the initial R&D competence on going public is also economically significant. The economic direct effect of the initial competence is comparable to those of the VC backing indicator and the percentage of joint projects with partners.[12] The results show that firm inherent R&D competence, which is highly persistent over time, is a fundamental component of comparative advantages in highly innovative industries that drive exits through IPOs at the end.

The result shows not only the direct effect of the initial R&D competence, but also the indirect effect of the initial R&D competence through diversification as a potential mediator. We find that a firm's low R&D competence in its main segment increases the firm's incentive to diversify its portfolio of new drug projects into multiple disease groups, and then the diversification increases the likelihood of going public. This indirect effect of the initial R&D competence on going public is statistically significant at the 5% level, but economically much weaker than the direct effect. The estimated coefficients in columns one and two imply that, regarding the likelihood of going public, the direct effect (0.0328) is roughly fourteen times greater than the indirect effect (-0.206*0.0115).

Regarding VC funding, we find in column four that only the indirect effect of the initial R&D competence through the diversification channel is statistically significant at the 5% level. The direct effect shown in the coefficient of initial R&D competence is positive but not statistically significant. The result in this specification implies that the initial R&D

[12] In unreported results, we find that the current main competence, not initial competence also has a significant and positive effect to increase the likelihood of going public.

competence in a firm's main segment affects the likelihood of getting more VC funding mainly through the diversification channel.

Overall, the results from our simultaneous analysis are consistent with the predictions that firms with higher R&D competence initially in their main segment are more likely to go public, whereas firms with lower R&D competence initially in their main segment increase the likelihood of going public or getting more VC funding by diversifying into multiple different industries.

We further stress-test our results by focusing only on between-effects in the cross-sections of our sample. We consider a refined sample that consists of only the final year observations for each firm. For a firm that eventually goes public, the IPO event takes place in the final year. For other firms, we select observations in the last available year. Then, we analogously run the simultaneous mediation model as in Table 8. The results are available in Appendix Table A.1. We confirm that our results are robust in the refined cross-sectional sample.

5.4 Simultaneous Analysis with the Medicare Part D Legislation

We recognize that the two main variables of interest in our analysis - diversification and firm growth (going public or getting more VC funding) are potentially endogenous. For example, managerial risk aversion may lead to diversification decision and the successful exit through an IPO as well. Therefore, we consider an exogenous shock that is likely to increases only diversification incentives to a certain group of firms to address these potential endogeneity concerns. Particularly, we use the passage of the Medicare Part D legislation (also called the Medicare prescription drug benefit) in 2003 as an exogenous shock to the diversification incentives.

Medicare Part D is a government program to subsidize the costs of prescription drugs and prescription drug insurance premiums for Medicare beneficiaries. Before Medicare Part D, Medicare program covered the medicine only associated with physician services. Part D significantly expanded drug coverage covered by Medicare. In particular, under Part D the drug coverage must include all drugs within the following indication classes: antipsychotics, antidepressants, anticonvulsants, immunosuppressants, cancer, and HIV/AIDS. Therefore, only specific firms with lack of the project pipelines in such indication classes should be materially affected by the shock in terms of their increased incentives to expend into the relevant disease groups. The prior literature on Medicare Part D shows that the passage and implementation of the program is associated with significant increases in pharmaceutical R&D for therapeutic classes that have higher Medicare market shares (Duggan and Morton (2011) and Blume-Kohout and Sood (2013)). Importantly in our context, the effect of Medicare Part D on firm growth comes only through the diversification channel by which firms without any project in the affected Part D drug classes can be exposed to the shock. Thus, the Medicare Part D shock is relevant and likely satisfies the exclusion restriction.

Based on the differential effects of Medicare Part D on diversification incentives, we identify the treated and control groups as follows: Firms without any projects in the required Part D drug classes before the legislation are included in the treated group, while firms with active projects in Part D drug classes before the legislation are included in the control group. The Medicare Part D shock should significantly affect the diversification incentives of the firms in our treated group, but not (or less) affect those of the firms in our control group. Furthermore, we expect that within the treated group the increased diversification incentives following the shock will be more pronounced for firms that show low competence in their original disease groups. Specifically, we estimate the following specification:

$$\begin{aligned} \mathit{Diversification}_{i,t} = v + \beta_1 \mathit{Treated}_i \times \mathit{Post}_t \times \mathit{Low\ competence}_i \\ \beta_2 \mathit{Treated}_i \times \mathit{Post}_t + \beta_3 \mathit{Treated}_i \times \mathit{Low\ competence}_i + \beta_4 \mathit{Post}_t \times \mathit{Low\ competence}_i \\ + \beta_5 \mathit{Treated}_i + \beta_6 \mathit{Post}_t + \beta_7 \mathit{Low\ competence}_i \\ + \beta_8 \mathit{R\&D\ characteristics}_{i,t} + \beta_9 \mathit{Other\ firm\ characteristics}_{i,t} \\ + \beta_{10} \mathit{Industry\ characteristics}_{i,t} + v_i + v_t + \epsilon_{i,t}, \end{aligned}$$

$$\begin{aligned} \mathit{Going\ public\ or\ VC\ funding}_{i,t} = w + \gamma_1 \mathit{Low\ competence}_i + \gamma_2 \mathit{Diversification}_{i,t} \\ + \gamma_3 \mathit{R\&D\ characteristics}_{i,t} + \gamma_4 \mathit{Other\ Firm\ characteristics}_{i,t} \\ + \gamma_5 \mathit{Industry\ characteristics}_{i,t} + w_i + w_t + \eta_{i,t}, \end{aligned}$$

where *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation, *Post* is one after the passage of the Medicare Part D legislation in 2003, and *Low competence* is an indicator variable for a firm with low initial main competence.

Post alone is subsumed by year fixed effects.

[Insert Table 9 Here]

Table 9 reports the estimation results of the system of equations with SUR as previously used. We find that firms with lower competence in the treated group indeed increase diversification after the passage of the Medicare Part D legislation, consistent with our prediction. The coefficients of the triple interaction terms in the second row are significantly positive at the 1% level for both columns one and three. The economic impact of the estimated triple interaction terms (1.132 and 1.125) implies that following the Medicare Part D shock the treated firms that have low initial competence increased their number of disease groups by roughly 1.3. The implications of other firm and industry characteristics on diversification are comparable to those in Table 8.

In column two, regarding IPO exits, we find that both the initial R&D competence (*Low competence* in this specification) and the instrumented diversification significantly increase the likelihood of going public. As before, the economic significance of the direct effect of the initial R&D competence on the likelihood of going public is estimated as comparable to those of the VC backing indicator and the percentage of joint projects with partners. Also, the result shows that the effect of diversification, even after our control of potential endogeneity concerns using the Medicare Part D shock, is significant and positive at the 5% level. We find that a firm with low R&D competence in its main segment that is not covered by the Medicare Part D program has higher incentives to diversify into the new industries supported by the Medicare Part D program, thereby being able to increase the likelihood of going public.[13]

Regarding VC funding, we also find consistent results. In column four, the result shows that both coefficients for the initial R&D competence (*Low competence*) and instrumented diversification are positive and significant. This indicates that the diversification channel as well as the initial R&D competence increases the likelihood of getting more VC funding. For the former channel, we especially address potential endogeneity concerns between diversification and VC funding by using the Medicare Part D shock as an instrument for diversification.

[13]In Appendix Table A.2, we consider both IPOs and acquisitions as successful exits of startup firms and run the analogous IPO regressions as in columns one and two of Table 8 and Table 9. We find that our results are robust of considering both exit choices.

As robustness, we use two alternative measures of R&D competence. First, we drop all suspended projects when counting the total number of projects in the subsequent years after they are initially identified as suspended. This is based on the assumption that all suspended projects are not resumed. Second, we consider a conditional success probability weighted R&D competence to reflect phase-specific difficulties of transition based on success rates estimated in DiMasi, Hansen, and Grabowski (2003). Our findings are robust to these alternative measures of R&D competence and the regression results are available in Appendix Table A.3 and A.4, respectively. Also, Povel, Sertsios, Kosova, and Kumar (2016) find that there is a significant and persistent entry year effect, more specifically whether it is during industry booms or glooms, on the performance of hotel investment. We thus consider the entry year fixed effects instead of the year fixed effects and run the analogous test. The results are available in Appendix Table A.5.[14]

Overall, our results are consistent with our predictions that firms with higher R&D competence initially in their main segment are more likely to go public or get more VC funding, whereas firms with lower R&D competence increase the likelihood of such successful growth by exploring other industries, especially where positive demand shocks are expected, for example in the Medicare Part D shock.

6 Conclusion

We examine private firms' diversification and growth in instances in which initial conditions in R&D competence exert strong effects on both simultaneously. Our study focuses on heterogeneous levels of R&D competence inherent at a firm's inception or acquired in the earliest stages of its life. Our detailed project-level drug development data that spans 30 years and 21 disease groups in the drug industry sector enables us to measure innate R&D competence overall and in a firm's main segment by calculating the suspension and success rates of new drug development projects.

We find that R&D competence within its first three years persists throughout a firm's lifetime. We further examine how this persistent R&D competence affects innovation strategies,

[14]Our results are also robust to dropping firm years where all projects are identified as suspended in a given year and thereafter. These firms are considered as firms that no longer operate.

as measured by diversification into multiple disease groups and firm growth through IPO exit and venture capital funding. We find that firms with high initial R&D competence in their main segment are more likely to focus on their best segment and grow faster and more successfully than firms with low initial competence. Specifically, they exit through IPOs earlier or receive greater amounts of venture capital funding. By contrast, firms with low initial R&D performance in their main segment tend to diversify into other disease groups, thereby also increasing the likelihood of going public at the end or securing more venture capital funding.

Using a simultaneous mediation model, we show that both the direct effect of initial R&D competence on firm growth and the indirect effect, whereby the mediation effect of diversification exists, are significant, but the direct effect is much greater than the indirect effect. We confirm the mediation effect of diversification on firm growth to likely be causal, using the Medicare Part D legislation as an exogenous shock to diversification incentives. Our results add new evidence to the literature that initial conditions in the earliest stage of firm life largely explain variations in corporate diversification and growth patterns.

Appendix. Variable Definitions

- *Going public* is the firm year dummy variable that is one if the firm goes public via an IPO in a given year and zero otherwise.
- *Exit* is the firm year dummy variable that is one if the firm either goes public via an IPO or is acquired in a given year, and zero otherwise.
- *VC funding* is the log of one plus the amount of VC funding (in million dollars) in a given year.
- *Diversification w/ # of industries* is the total number of different disease groups where the firm has projects.
- *Diversification w/ project shares* is one minus the sum of squared project share of each disease group in a given year. The project share is the number of projects in each disease group divided by the number of total projects over all disease groups.
- *R&D competence* is the number of advancing events to the next phase minus the number of suspension events divided by the total number of projects in the firm's new drug development pipeline in a given year.
- *Main competence* is R&D competence of the firm's main segment. A firm's main segment is the disease group where the firm has the largest number of projects.
- *Initial main competence* is the average of the first three non-missing values of R&D competence of the firm's main segment.
- *Phase 0*: The drug or treatment is in pre-clinical stage, or Investigational New Drug (IND) application is submitted to the FDA.
- *Phase 1*: Researchers test the drug or treatment in a small group of healthy volunteers for the first time to evaluate its safety, determine a safe dosage range, and identify side effects.
- *Phase 2*: The drug or treatment is given to a small group of people who have a certain disease or condition to see if it is effective and to further evaluate its safety.
- *Phase 3*: The drug or treatment is given to large groups of people from several hundreds to 3,000 to confirm its effectiveness, monitor side effects, compare it to commonly used treatments, and collect information that will allow the drug or treatment to be used safely.
- *Phase 4*: Clinical studies are done. New Drug Application (NDA) or Biologic License Application (BLA) is submitted for the FDA review process.
- *Phase 5*: The FDA has approved the drug or treatment for marketing in the United States.
- *Matured phase projects* is the drug projects in the pipeline in phase 4 and 5.

- *# of projects* is the firm's total number of projects in the pipeline in all phases in a given year.
- *% of matured projects* is the percentage of matured projects in the firm's pipeline in a given year.
- *% of projects with partner* is the percentage of the projects in the firm's pipeline that have partners in a given year.
- *Log(1+firm age)* is the log of one plus the firm's founding years. Founding years for IPO firms are from Jay Ritter's IPO data web site. We thank Jay Ritter for kindly providing us founding year data. Other private firms' founding years are hand-collected.
- *VC backed* is firm-level dummy variable of venture capital funding that equals one if a firm has ever received funding from venture capital investors.
- *Log(# of projects)* is the log of one plus the firm's total number of projects in the pipeline in all phases in a given year.
- *Log(# of phase 0 projects)* is the log of one plus the firm's phase 0 in the pipeline in a given year.
- *Log(# of competitors)* is the log of the total number of firms with new drug development in each industry in a given year.
- *Ind failure rate* is the industry average of firm failure rate in a given year.
- *Ind % matured projects* is the industry average of firm % matured projects in a given year.

References

Abernathy, William J, and James M Utterback, 1978, Patterns of industrial innovation, *Technology review* 64, 254–228.

Ayyagari, Meghana, Asli Demirgüç-Kunt, and Vojislav Maksimovic, 2015, What determines entrepreneurial outcomes in emerging markets? the role of initial conditions, *World Bank Policy Research Working Paper*.

Benner, Mary J, and Michael Tushman, 2002, Process management and technological innovation: A longitudinal study of the photography and paint industries, *Administrative Science Quarterly* 47, 676–707.

Berry, Donald A, 1972, A bernoulli two-armed bandit, *The Annals of Mathematical Statistics* pp. 871–897.

Blume-Kohout, Margaret E, and Neeraj Sood, 2013, Market size and innovation: Effects of medicare part d on pharmaceutical research and development, *Journal of public economics* 97, 327–336.

Chemmanur, Thomas J, Jie He, Shan He, and Debarshi K Nandy, 2015, The exit choices of entrepreneurial firms, in *Working Paper*.

Chemmanur, Thomas J, Shan He, and Debarshi K Nandy, 2009, The going-public decision and the product market, *Review of Financial Studies* p. hhp098.

DiMasi, Joseph A, Ronald W Hansen, and Henry G Grabowski, 2003, The price of innovation: new estimates of drug development costs, *Journal of health economics* 22, 151–185.

Drucker, Peter Ferdinand, 2007, *Management challenges for the 21st century* (Routledge).

Duggan, Mark G, and Fiona Scott Morton, 2011, The medium-term impact of medicare part d on pharmaceutical prices, *The American Economic Review* 101, 387–392.

Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist, 2016, The bright side of patents, *NBER Working paper*.

Ferreira, Daniel, Gustavo Manso, and André C Silva, 2014, Incentives to innovate and the decision to go public or private, *Review of Financial Studies* 27, 256–300.

Gao, Huasheng, Po-Hsuan Hsu, and Kai Li, 2015, Public market listing and corporate innovation strategies: Evidence from private firms, Discussion paper, Working paper.

Griliches, Zvi, Ariel Pakes, and Bronwyn H Hall, 1988, The value of patents as indicators of inventive activity, *NBER Working paper*.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, 2001, The nber patent citation data file: Lessons, insights and methodological tools, Discussion paper, National Bureau of Economic Research.

Hallen, Benjamin L, 2008, The causes and consequences of the initial network positions of new organizations: From whom do entrepreneurs receive investments?, *Administrative Science Quarterly* 53, 685–718.

Hamel, Gary, and Coimbatore K Prahalad, 1990, The core competence of the corporation, *Harvard business review* 68, 79–91.

Henderson, Rebecca, and Iain Cockburn, 1994, Measuring competence? exploring firm effects in pharmaceutical research, *Strategic management journal* 15, 63–84.

Holmstrom, Bengt, 1989, Agency costs and innovation, *Journal of Economic Behavior & Organization* 12, 305–327.

King, Robert G, and Ross Levine, 1993, Finance, entrepreneurship and growth, *Journal of Monetary economics* 32, 513–542.

Klepper, Steven, 1996, Entry, exit, growth, and innovation over the product life cycle, *American economic review* pp. 562–583.

Kortum, Samuel, and Josh Lerner, 2000, Assessing the contribution of venture capital to innovation, *RAND journal of Economics* pp. 674–692.

Lemmon, Michael L, Michael R Roberts, and Jaime F Zender, 2008, Back to the beginning: persistence and the cross-section of corporate capital structure, *Journal of Finance* 63, 1575–1608.

Lerner, Josh, 1995, Venture capitalists and the oversight of private firms, *Journal of Finance* 50, 301–318.

Levinthal, Daniel A, and James G March, 1993, The myopia of learning, *Strategic management journal* 14, 95–112.

Maksimovic, Vojislav, and Gordon Phillips, 2002, Do conglomerate firms allocate resources inefficiently across industries? theory and evidence, *Journal of Finance* pp. 721–767.

——— , and Liu Yang, 2013, Private and public merger waves, *Journal of Finance* 68, 2177–2217.

Maksimovic, Vojislav, and Gordon M Phillips, 2013, Conglomerate firms, internal capital markets and the theory of the firm, *Robert H. Smith School Research Paper.*

Manso, Gustavo, 2011, Motivating innovation, *Journal of Finance* 66, 1823–1860.

March, James G, 1991, Exploration and exploitation in organizational learning, *Organization science* 2, 71–87.

Marquis, Christopher, and András Tilcsik, 2013, Imprinting: Toward a multilevel theory, *The Academy of Management Annals* 7, 195–245.

McGrath, Rita Gunther, 2001, Exploratory learning, innovative capacity, and managerial oversight, *Academy of management journal* 44, 118–131.

Pagano, Marco, Fabio Panetta, and Luigi Zingales, 1998, Why do companies go public? an empirical analysis, *Journal of Finance* 53, 27–64.

Pandey, Sandeep, Deepayan Chakrabarti, and Deepak Agarwal, 2007, Multi-armed bandit problems with dependent arms, in *Proceedings of the 24th international conference on Machine learning* pp. 721–728. ACM.

Pástor, L'uboš, Lucian A Taylor, and Pietro Veronesi, 2009, Entrepreneurial learning, the ipo decision, and the post-ipo drop in firm profitability, *Review of Financial Studies* 22, 3005–3046.

Povel, Paul, Giorgo Sertsios, Renata Kosova, and Praveen Kumar, 2016, Boom and gloom, *The Journal of Finance*.

Powell, Walter W, and Kurt W Sandholtz, 2012, Amphibious entrepreneurs and the emergence of organizational forms, *Strategic Entrepreneurship Journal* 6, 94–115.

Thakor, Anjan, 2013, Innovation and growth: What do we know?, *World Scientific*.

Figure 1: Mediation Model

The figures illustrate our model with one mediator, diversification. Figure (a) displays the direct effect of R&D competence on going public decisions (β_0) without the mediator. Figure (b) displays the effects of R&D competence both directly (β_1) and indirectly through the mediator, diversification ($\beta_2 \times \beta_3$).

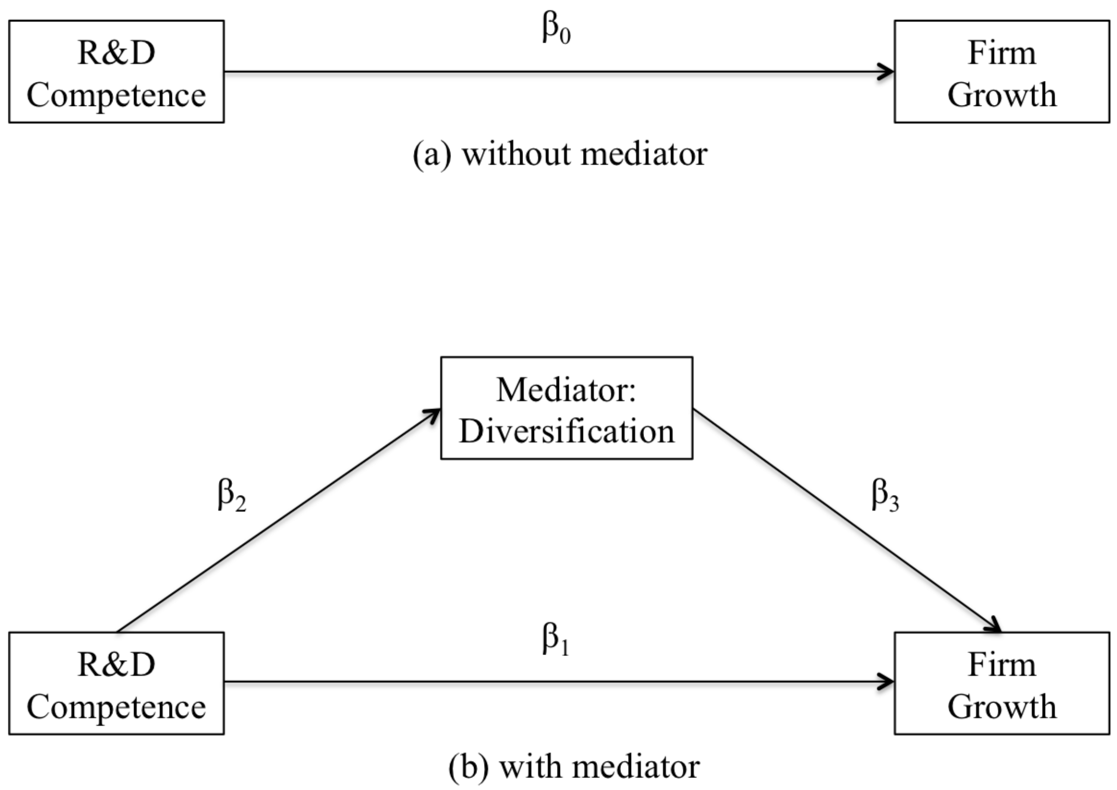


Figure 2: Histogram of Initial Main R&D Competence

This figure illustrates the distribution of initial R&D competence of a firm's main segment. We define a firm's main segment based on the disease group where the firm has the largest number of projects. *Initial main competence* is the average of the first three non-missing values of R&D competence for the firm's main segment. 206 firms (26%) out of 799 sample firms have non-zero initial main competence.

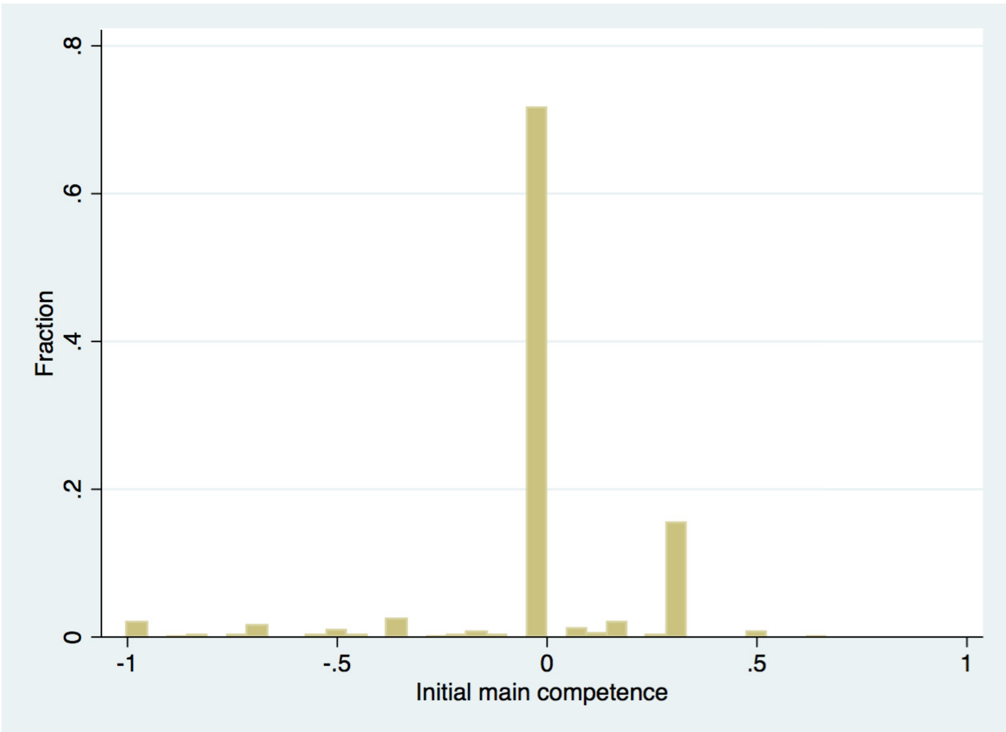
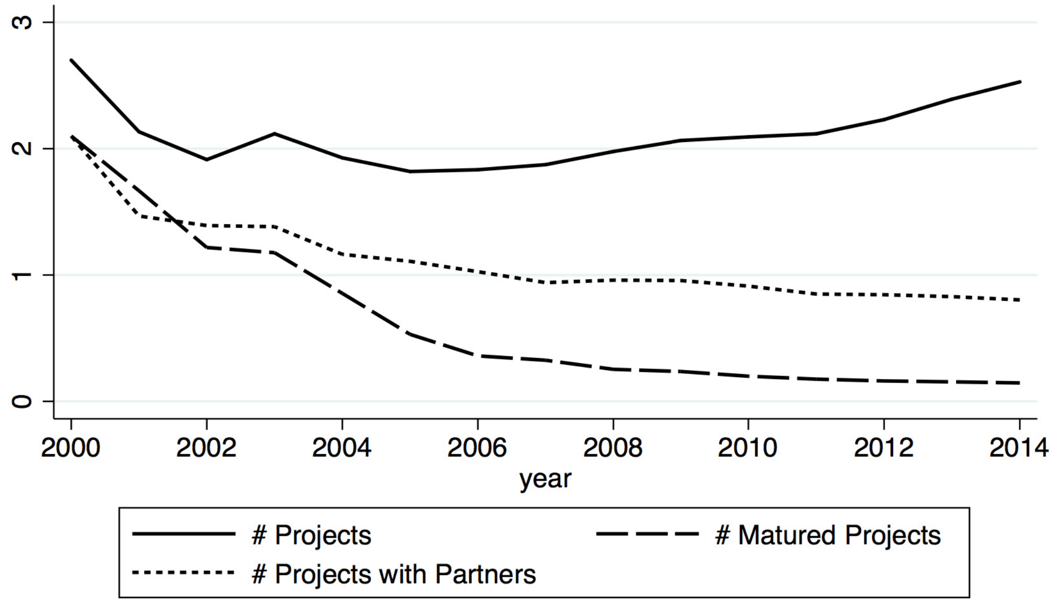
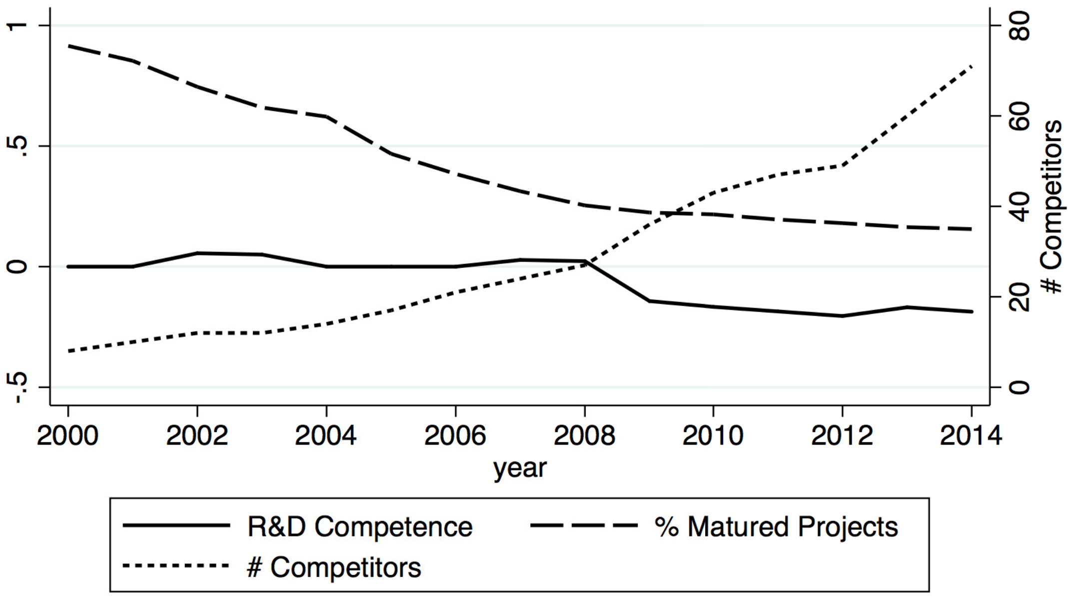


Figure 3: Time Trends of R&D Variables

The figures display the time trends of (a) firm-specific and (b) industry-specific R&D variables of our sample firms in the drug industry sector. The averages of the following variables are displayed. *# of projects* is the firm's total number of projects in the pipeline in all phases from phase 0 to phase 5 in a given year. *# of matured projects* is the firm's total number of projects in the pipeline in phase 4 and 5 in a given year. *# of projects with partners* is the firm's total number of projects in the pipeline that have partners in a given year. *# of competitors* is the number of all firms in the firm's main disease group including all publicly- and privately-held companies. *R&D competence* is the industry average of firm R&D competence in a given year. *% matured projects* is the industry percentage of matured projects (phase 4 and 5) in a given year.



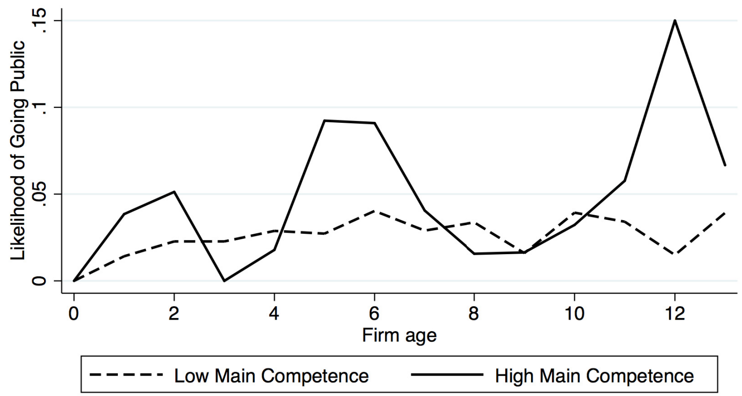
(a) Firm R&D Variables



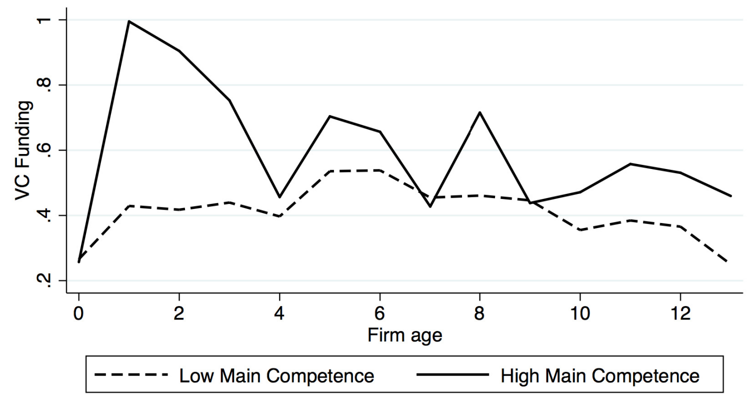
(b) Industry Variables

Figure 4: Time Trends of Competence, Diversification, and Firm Growth

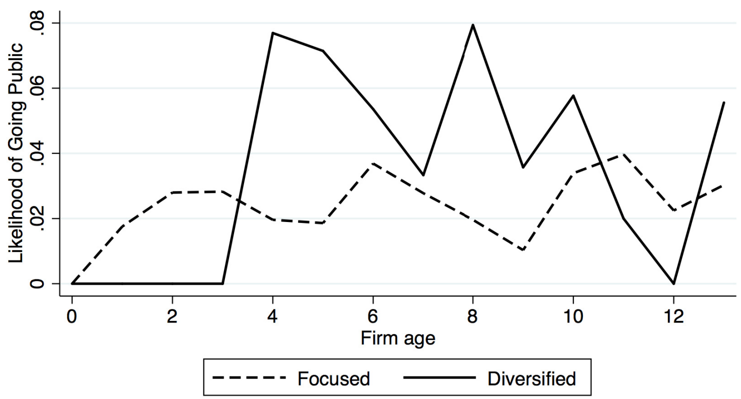
The figures display the time trends of percentage firms of going public and VC funding over firm age. In Panels (a) and (b), the sample is split by initial competence of firm's main segment, and in Panels (c) and (d) by diversification. *Initial main competence* is the average of the first three non-missing values of R&D Competence of the firm's main segment. *Initial main competence* is high (solid) if it is positive, and low (dash) otherwise. Panels (c) and (d) are for focused (dash) versus diversified (solid) firms within the subset of firms with low initial main competence. A firms is diversified if the firm develops new drugs over multiple disease groups, and focused otherwise.



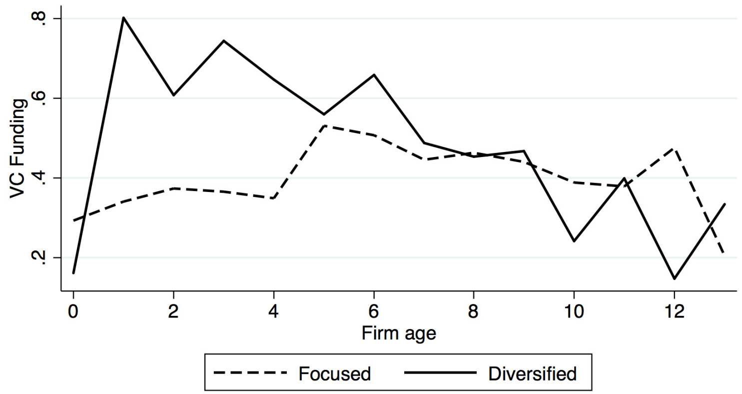
(a) Competence and IPO



(b) Competence and VC Funding



(c) Diversification and IPO



(d) Diversification and VC Funding

Figure 5: Persistency of R&D Competence

The figure illustrates the persistence of overall R&D competence. *Main competence* is R&D competence of the firm's main segment. *Initial main competence* is the average of the first three non-missing values of R&D competence of the firm's main segment. *Initial main competence* is high (solid) if it is positive, and low (dash) otherwise.

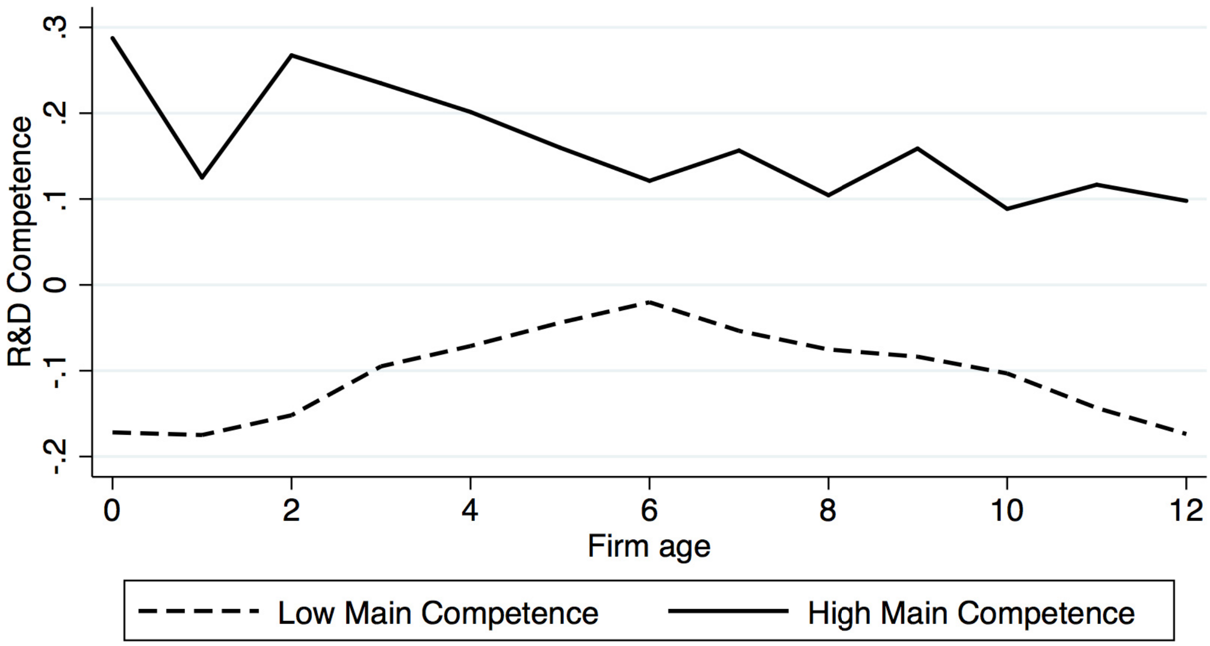


Table 1: Industry Classification

The table presents our industry classification based on disease codes from the BioMedTracker database. Our sample comprises of all private firms in the drug industries that are required to report to the U.S. Food and Drug Administration (FDA). The sample consists of 3,851 firm year observations of 799 firms during the sample period of 1985-2014.

Disease code	Disease group	Observations	Percent
1	Allergy	9	0.23
2	Autoimmune/Immunology	232	6.02
3	Cardiovascular	315	8.18
4	Dermatology	19	0.49
5	Ear, Nose, Throat/Dental	25	0.65
6	Endocrine	289	7.50
7	Gastroenterology	59	1.53
8	Hematology	113	2.93
9	Infectious	381	9.89
10	Metabolic	120	3.12
11	Neurology	563	14.62
12	Obstetrics/Gynecology	18	0.47
13	Oncology	1,172	30.43
14	Ophthalmology	136	3.53
15	Orthopedics	3	0.08
16	Psychiatry	119	3.09
17	Renal	30	0.78
18	Respiratory	151	3.92
19	Rheumatology	19	0.49
20	Urology	60	1.56
21	Not Specified	18	0.47
Total		3,851	100

Table 2: Summary Statistics

The table presents summary statistics for private firms in our sample from the drug industries that are required to report to the U.S. Food and Drug Administration (FDA). The sample consists of 3,851 firm year observations of 799 firms during the sample period of 1985-2014. All variable definitions are in Appendix.

	Mean	Std. Dev	Min	Median	Max	Obs.
Diversification Decisions and Firm Growth						
Diversification w/ # of industries	1.380	0.874	1.000	1.000	9.000	3851
Diversification w/ project shares	0.123	0.229	0.000	0.000	0.821	3851
Going public (firm)	0.142	0.349	0.000	0.000	1.000	3851
Going public (firm year)	0.032	0.177	0.000	0.000	1.000	3851
VC backed	0.522	0.500	0.000	1.000	1.000	3851
VC fund	3.812	15.413	0.000	0.000	471.070	3851
Log(1+VC fund)	0.447	1.076	0.000	0.000	6.157	3851
Firm Characteristics						
R&D competence	-0.064	0.381	-1.000	0.000	1.000	3851
Main competence	-0.063	0.390	-1.000	0.000	1.000	3851
Initial Main competence	0.011	0.235	-1.000	0.000	0.667	3851
Firm age	7.318	4.447	0.000	7.000	36.000	3851
Log(1+firm age)	1.946	0.650	0.000	2.079	3.611	3851
# of projects	2.218	2.144	1.000	1.000	23.000	3851
# of phase 0 projects	0.234	0.684	0.000	0.000	7.000	3851
# of phase 1 projects	0.483	0.837	0.000	0.000	10.000	3851
# of phase 2 projects	0.781	1.139	0.000	0.000	13.000	3851
# of phase 3 projects	0.145	0.414	0.000	0.000	4.000	3851
# of phase 4 projects	0.018	0.155	0.000	0.000	3.000	3851
# of phase 5 projects	0.239	1.047	0.000	0.000	12.000	3851
% matured projects	0.087	0.270	0.000	0.000	1.000	3851
% projects with partner	0.396	0.466	0.000	0.000	1.000	3851
Industry Characteristics						
# of competitors	47.294	17.253	2.000	47.000	71.000	3851
Log(# of competitors)	5.277	0.856	1.792	5.493	6.366	3851
Ind failure rate	0.217	0.104	0.000	0.241	0.414	3851
Ind % martured projects	0.233	0.180	0.056	0.200	1.000	3851

Table 3: Comparison in Main Segment Competence and Diverification

The table presents summary statistics for the firms that have high vs low initial main competence (Panel A) and that are diversified vs focused (Panel B) among our private firm sample in the drug industries. In Panel A, the sample consists of 3,851 firm year observations of 799 firms during the sample period of 1985-2014. In Panel B, the sample consists of 3,091 firm year observations of 646 firms from the subset of firms with low initial main competence. Firm initial main competence is low when it is nonpositive. A firm is diversified if the firm develops new drugs over multiple disease groups, and focused otherwise. All variable definitions are in Appendix. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: High vs. Low Main Segment Competence

	High Main Competence		Low Main Competence		
	Mean	Median	Mean	Median	Mean Difference
VC backed	0.592	1.000	0.505	1.000	0.087***
Log(1+VC fund)	0.560	0.000	0.420	0.000	0.140***
Going public (firm)	0.207	0.000	0.126	0.000	0.080***
Going public (firm year)	0.049	0.000	0.028	0.000	0.021***
R&D competence	0.111	0.000	-0.107	0.000	0.217***
Main competence	0.121	0.000	-0.108	0.000	0.229***
Initial main competence	0.306	0.333	-0.061	0.000	0.368***
Diversification	0.128	0.000	0.122	0.000	0.006
Log(1+firm age)	1.994	2.079	1.934	2.079	0.059**
# of projects	2.462	2.000	2.158	1.000	0.304***
% matured projects	0.064	0.000	0.093	0.000	-0.029***
% projects with partner	0.327	0.000	0.414	0.000	-0.087***
Log(# of competitors)	5.317	5.501	5.267	5.493	0.050
Ind failure rate	0.220	0.236	0.217	0.242	0.004
Ind % matured projects	0.214	0.200	0.238	0.202	-0.024***
Observations	760		3091		

Panel B: Diversified vs. Focused

	Diversified		Focused		
	Mean	Median	Mean	Median	Mean Difference
VC backed	0.568	1.000	0.486	0.000	0.082***
Log(1+VC fund)	0.471	0.000	0.405	0.000	0.066
Going public (firm)	0.160	0.000	0.116	0.000	0.044***
Going public (firm year)	0.043	0.000	0.024	0.000	0.020***
R&D competence	-0.122	0.000	-0.102	0.000	-0.020
Main competence	-0.130	0.000	-0.102	0.000	-0.028*
Initial main competence	-0.069	0.000	-0.059	0.000	-0.009
Log(1+firm age)	2.053	2.197	1.899	2.079	0.154***
# of projects	4.143	3.000	1.563	1.000	2.580***
% matured projects	0.152	0.000	0.075	0.000	0.078***
% projects with partner	0.422	0.333	0.411	0.000	0.010
Log(# of competitors)	5.072	5.226	5.326	5.509	-0.254***
Ind failure rate	0.230	0.254	0.212	0.238	0.018***
Ind % martured projects	0.250	0.196	0.234	0.208	0.016**
Observations	713		2378		

Table 4: Persistency of R&D Competence

The table examines the persistence of R&D competence in firm's main disease group. *Main competence* is the R&D competence of the firm's main segment. Initial main competence is the average of the first three non-missing values of R&D competence for the firm's main segment. The table presents coefficient estimates and adjusted R-squares for mutiple different model specifications. *t-statistics* (in parenthesis) are robust and adjusted for firm clustering. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Main competence					
	(1)	(2)	(3)	(4)	(5)	(6)
Initial main competence					0.752***	0.736***
					(13.74)	(14.51)
Log(# of phase 0 projects)				0.0169		0.0461***
				(0.37)		(2.70)
Log(1+firm age)				0.303*		-0.172***
				(1.74)		(-5.55)
% matured projects				0.362**		0.147***
				(2.52)		(4.73)
% projects with partner				-0.0883		-0.0600**
				(-0.85)		(-2.53)
Log(# of competitors)				-0.0967*		-0.108
				(-1.92)		(-1.05)
Ind failure rate				-0.564		-0.456*
				(-1.23)		(-1.65)
Ind % matured projects				-0.837***		-0.363*
				(-2.84)		(-1.93)
Observations	3071	3071	3071	3071	3071	3071
R^2	0.457	0.014	0.522	0.530	0.179	0.215
Adjusted R^2	0.300	0.005	0.375	0.385	0.165	0.200
Fixed Effects	Firm	Year	Firm, Year	Firm, Year	Ind, Year	Ind, Year

Table 5: Within-firm Project Performance

The table reports performance correlation between within-firm projects. The dependent variable, *Project performance* equals one if a firm's project advances to the next phase, minus one if the project is suspended, and zero otherwise for a given year. *Peer project performance* is the average of all other projects' performance for the firm in the given year. *Initial main competence* is the average of the first three non-missing values of R&D competence of the firm's main segment. Initial main competence is high if it is positive, and low, otherwise. We define a firm's main segment based on the disease group where the firm has the largest number of projects. Industry and year fixed effects are included. Standard errors (in parenthesis) are robust and adjusted for clustering within firm. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Project performance		
	(1)	(2)	(3)
Peer project performance	0.139***	0.142***	0.116***
	(0.02)	(0.02)	(0.02)
Peer project performance * Initial main competence		0.233***	
		(0.04)	
Initial main competence		0.096***	
		(0.02)	
Peer performance * High initial main competence			0.071**
			(0.03)
High initial main competence			0.044***
			(0.01)
Log(# of phase 0 projects)	-0.013**	-0.012**	-0.010**
	(0.01)	(0.01)	(0.01)
Log(1+firm age)	-0.011**	-0.006	-0.010**
	(0.01)	(0.01)	(0.01)
% matured projects	0.029	0.031*	0.036*
	(0.02)	(0.02)	(0.02)
% projects with partner	-0.002	0.001	0.002
	(0.01)	(0.01)	(0.01)
Log(# of competitors)	-0.098**	-0.085**	-0.096***
	(0.04)	(0.03)	(0.04)
Ind failure rate	-0.039	0.009	-0.009
	(0.13)	(0.13)	(0.13)
Ind % matured projects	0.013	0.009	-0.009
	(0.08)	(0.07)	(0.08)
VC backed	0.006	0.004	0.005
	(0.01)	(0.01)	(0.01)
Observations	8636	8583	8583
Adjusted R^2	0.032	0.038	0.036
Fixed Effects	Ind, Year	Ind, Year	Ind, Year

Table 6: Initial R&D Competence and Diversification

The table examines the relation between diversification and initial main competence as well as time-varying main competence. The sample comprises of all firm years in our sample period, 1985-2014. In the first two columns, the dependent variable, *Diversification with # of industries* is the total number of all unique disease groups where the firm has projects. In the last two columns, the dependent variable, *Diversification with project shares* is the firm-year level diversification measure calculated by one minus the sum of squared project share of each disease group at the given year. The project share is the number of projects in each disease group divided by the number of total projects over all disease groups. *Main competence* is the R&D competence of the firm's major segment. *Initial main competence* is the average of the first three non-missing values of R&D competence for the firm's main segment. We define a firm's main segment based on the disease group where the firm has the largest number of projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for industry and year clustering. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification with # of industries		Diversification with project shares	
	(1)	(2)	(3)	(4)
Initial Main competence	-0.206***		-0.0399**	
	(-3.28)		(-2.40)	
Main competence		-0.190***		-0.0347***
		(-5.27)		(-3.98)
Log(# of phase 0 projects)	0.430***	0.439***	0.116***	0.118***
	(6.64)	(6.72)	(9.24)	(9.34)
Log(1+firm age)	0.125***	0.108***	0.0361***	0.0329***
	(6.21)	(5.83)	(6.34)	(6.06)
% matured projects	0.432***	0.447***	0.0986***	0.101***
	(7.90)	(8.15)	(6.63)	(6.79)
% projects with partner	0.108***	0.101***	0.0256***	0.0245***
	(4.37)	(4.13)	(3.56)	(3.48)
Log(# of competitors)	0.604***	0.583***	0.223***	0.219***
	(3.74)	(3.62)	(4.83)	(4.78)
Ind failure rate	0.463	0.379	0.174	0.159
	(1.10)	(0.91)	(1.27)	(1.17)
Ind % matured projects	0.599**	0.561**	0.142*	0.136*
	(2.22)	(2.08)	(1.75)	(1.68)
VC backed	0.0374*	0.0346*	0.0205***	0.0200***
	(1.81)	(1.66)	(3.67)	(3.54)
Observations	3851	3851	3851	3851
Adjusted R^2	0.146	0.150	0.144	0.146
Fixed Effects	Ind, Year	Ind, Year	Ind, Year	Ind, Year

Table 7: Diversification Pattern

The table presents summary statistics for destination industry of diversifying firms in our sample during the sample period of 1985-2014. The sample consists of 220 firm year observations of 165 firms' diversification patterns. We define the following indicator variables that equal one if the industry average is above the median value for all industries. Variables include industry R&D competence (High competence), project suspension (High suspend rate) and advance rates (High advance rate), percentage of matured projects (More matured product), percentage of Phase 0 product (More phase 0 product), competition (High Competition), number of industry competitors (More incumbents), and number of products (More products). The table presents the percentages of the firms that enter into the specific destination among all diversified firms, *p-value* and *t-statistics*.

	Percentage	p-value	t-statistics
High competence	49.5%	0.55	-0.13
High suspend rate	96.8%	0.00	39.47
High advance rate	47.3%	0.79	-0.81
More products	81.4%	0.00	11.92
More matured products	1.4%	0.00	-62.06
More phase 0 products	93.2%	0.00	25.35
High competition	99.5%	0.00	109.00
More incumbents	99.5%	0.00	109.00
Observations		220	

Table 8: Initial Competence, Diversification and Firm Growth

The table reports the effects of initial main competence on (1) Going public and (2) VC funding directly and indirectly through diversification using a simultaneous mediation model. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO at a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. *Initial main competence* is the average of the first three non-missing values of R&D competence for the firm's main segment. We define a firm's main segment based on the disease group where the firm has the largest number of projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.0115**		0.0470**
		(2.30)		(2.19)
Initial main competence	-0.206***	0.0328***	-0.202***	0.0903
	(-3.30)	(2.69)	(-3.25)	(1.29)
Log(# of phase 0 projects)	0.430***	0.00488	0.429***	0.0338
	(6.69)	(0.31)	(6.64)	(0.54)
Log(1+firm age)	0.125***	0.000319	0.129***	-0.0210
	(6.26)	(0.10)	(6.38)	(-0.80)
% matured projects	0.432***	-0.0119	0.427***	-0.397***
	(7.96)	(-0.88)	(7.92)	(-6.11)
% projects with partner	0.108***	0.0131**	0.106***	-0.00947
	(4.41)	(2.50)	(4.36)	(-0.23)
Log(# of competitors)	0.604***	0.0208	0.605***	0.0757
	(3.77)	(0.49)	(3.77)	(0.36)
Ind failure rate	0.463	-0.387***	0.463	-0.450
	(1.11)	(-2.99)	(1.10)	(-0.59)
Ind % matured projects	0.599**	0.176*	0.590**	-0.294
	(2.24)	(1.84)	(2.19)	(-0.60)
VC backed	0.0374*	0.0475***		
	(1.82)	(5.16)		
Observations	3851		3851	
Berndt R^2	0.206		0.186	
Fixed Effects	Ind, Year		Ind, Year	

Table 9: Medicare Part D Shock on Diversification and Firm Growth

The table reports the effects of diversification on (1) Going public and (2) VC funding using an instrumental variable approach with the passage of Medicare Part D legislation. *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation. *After* is one after the passage of the Medicare Part D legislation in 2003. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO at a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.0108**		0.0439**
		(2.17)		(2.05)
Treated * Post * Low competence	1.132***		1.125***	
	(2.95)		(2.94)	
Treated * Post	-0.401		-0.390	
	(-1.48)		(-1.45)	
Treated * Low competence	-1.171***		-1.170***	
	(-3.11)		(-3.12)	
Post * Low competence	-0.388		-0.389	
	(-1.26)		(-1.27)	
Treated	0.784**		0.775**	
	(2.55)		(2.52)	
Low competence	0.347	-0.0192**	0.348	-0.130***
	(1.14)	(-2.40)	(1.15)	(-2.60)
Log(# of phase 0 projects)	0.434***	0.00613	0.434***	0.0411
	(6.63)	(0.39)	(6.59)	(0.66)
Log(1+firm age)	0.110***	0.00224	0.114***	-0.0166
	(5.90)	(0.71)	(5.99)	(-0.63)
% matured projects	0.423***	-0.00966	0.419***	-0.388***
	(7.06)	(-0.71)	(7.03)	(-6.00)
% projects with partner	0.130***	0.0124**	0.128***	-0.00589
	(5.11)	(2.42)	(5.08)	(-0.14)
Log(# of competitors)	0.663***	0.0186	0.661***	0.0770
	(3.49)	(0.44)	(3.48)	(0.37)
Ind failure rate	0.0700	-0.391***	0.0651	-0.456
	(0.17)	(-3.03)	(0.16)	(-0.60)
Ind % matured projects	1.010***	0.170*	1.003***	-0.309
	(3.34)	(1.78)	(3.31)	(-0.62)
VC backed	0.0318	0.0474***		
	(1.45)	(5.10)		
Observations	3851		3851	
Berndt R^2	0.213		0.195	
Fixed Effects	Ind, Year		Ind, Year	

Table A.1: Initial Competence, Diversification and Firm Growth: Cross-sectional Analysis

The table reports the effects of initial main competence on (1) Going public and (2) VC funding directly and indirectly through diversification using a simultaneous mediation model. We use a refined sample that consists of only the final year observations for each firm during our sample period of 1985-2014. For a firm that eventually goes public, the IPO event takes place in the final year. For other firms, we select observations in the last available year. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO during the period, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the total VC funding amount during the period. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects at the final observation. *Initial main competence* is the average of the first three non-missing values of R&D competence for the firm's main segment. We define a firm's main segment based on the disease group where the firm has the largest number of projects. Industry fixed effects are included. Standard errors (in parenthesis) are robust and adjusted for clustering within industry and entry-year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.021*		0.251***
		(0.01)		(0.07)
Initial main competence	-0.319*	0.158***	-0.307*	0.309
	(0.18)	(0.04)	(0.18)	(0.25)
Log(# of phase 0 projects)	0.601***	0.005	0.599***	-0.244*
	(0.10)	(0.02)	(0.11)	(0.15)
Log(1+firm age)	0.227***	-0.013	0.237***	0.300**
	(0.06)	(0.02)	(0.06)	(0.12)
% matured projects	0.653***	-0.051	0.636***	-1.005***
	(0.18)	(0.06)	(0.17)	(0.28)
% projects with partner	0.074	0.058**	0.070	0.046
	(0.06)	(0.02)	(0.06)	(0.13)
Log(# of competitors)	-0.554	-1.567***	-0.585	-4.352**
	(0.85)	(0.45)	(0.85)	(1.92)
Ind failure rate	0.995	1.667	0.937	5.071
	(2.54)	(1.41)	(2.54)	(5.84)
Ind % matured projects	-0.575	-3.022***	-0.621	-10.110***
	(1.45)	(0.91)	(1.46)	(3.64)
VC backed	0.074	0.166***		
	(0.07)	(0.02)		
Observations	796		796	
Fixed Effects	Ind, Entry Year		Ind, Entry Year	

Table A.2: Exits through an IPO or an Acquisition

The table reports the effects of initial main competence on the two exit choices including IPOs and acquisitions directly and indirectly through diversification using a simultaneous mediation model. The dependent variable, *Exit* is one if the firm either goes public via an IPO or is acquired in a given year, and zero otherwise. The first two columns report analogous regression results as in the first two columns of Table 8 that replaces the dependent variable *Going public* with *Exit*. The last two columns report analogous regression results as in the first two columns of Table 9 that also replaces the dependent variable *Going public* with *Exit*. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Exit	Diversification	Exit
	(1)	(2)	(3)	(4)
Diversification		0.012**		0.012**
		(0.01)		(0.01)
Initial main competence	-0.218***	0.020		
	(0.06)	(0.02)		
Treated * Post * Low competence			1.535***	
			(0.35)	
Treated * Post			-0.407	
			(0.29)	
Treated * Low competence			-1.615***	
			(0.34)	
Post * Low competence			-0.345	
			(0.31)	
Treated			0.852***	
			(0.32)	
Low competence			0.310	-0.017**
			(0.31)	(0.01)
Log(# of phase 0 projects)	0.417***	0.006	0.423***	0.007
	(0.07)	(0.01)	(0.07)	(0.01)
Log(1+firm age)	0.129***	-0.011**	0.122***	-0.010**
	(0.02)	(0.00)	(0.02)	(0.00)
% matured projects	0.407***	-0.008	0.403***	-0.007
	(0.05)	(0.02)	(0.06)	(0.02)
% projects with partner	0.122***	0.014**	0.143***	0.014**
	(0.02)	(0.01)	(0.03)	(0.01)
Log(# of competitors)	0.568***	0.035	0.583***	0.034
	(0.12)	(0.05)	(0.15)	(0.05)
Ind failure rate	0.645	-0.509***	0.043	-0.512***
	(0.44)	(0.19)	(0.39)	(0.19)
Ind % matured projects	0.331	0.217*	0.888***	0.213*
	(0.26)	(0.12)	(0.28)	(0.12)
VC backed	0.019	0.054***	0.009	0.054***
	(0.02)	(0.01)	(0.02)	(0.01)
Observations	3708		3708	
Fixed Effects	Ind, Year		Ind, Year	

Table A.3: Alternative Competence Measure: Dropping Suspended Projects

The table reports the effects of diversification on (1) Going public and (2) VC funding using an instrumental variable approach with the passage of Medicare Part D legislation as in Table 9 using an alternative measure of R&D competence. With this R&D competence measure, we drop all suspended projects when counting the total number of projects in the subsequent years after they are initially identified as suspended. *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation. *Post* is one after the passage of the Medicare Part D legislation in 2003. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO at a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.010**		0.041*
		(0.01)		(0.02)
Treated * Post * Low competence	1.211***		1.206***	
	(0.39)		(0.39)	
Treated * Post	-0.453*		-0.445	
	(0.27)		(0.27)	
Treated * Low competence	-1.247***		-1.247***	
	(0.38)		(0.38)	
Post * Low competence	-0.405		-0.407	
	(0.31)		(0.31)	
Treated	0.827***		0.819***	
	(0.31)		(0.31)	
Low competence	0.360	-0.017**	0.362	-0.117**
	(0.31)	(0.01)	(0.30)	(0.05)
Log(# of phase 0 projects)	0.439***	0.002	0.438***	0.022
	(0.06)	(0.02)	(0.07)	(0.06)
Log(1+firm age)	0.131***	0.003	0.134***	0.005
	(0.02)	(0.00)	(0.02)	(0.03)
% matured projects	0.459***	-0.010	0.455***	-0.397***
	(0.06)	(0.01)	(0.06)	(0.07)
% projects with partner	0.189***	0.007	0.188***	0.007
	(0.03)	(0.00)	(0.03)	(0.03)
Log(# of competitors)	0.739***	0.021	0.737***	-0.015
	(0.19)	(0.04)	(0.19)	(0.21)
Ind failure rate	0.151	-0.395***	0.146	-0.542
	(0.43)	(0.13)	(0.43)	(0.78)
Ind % matured projects	1.102***	0.178*	1.094***	-0.468
	(0.31)	(0.10)	(0.31)	(0.49)
VC backed	0.029	0.049***		
	(0.02)	(0.01)		
Observations	3658		3658	
Fixed Effects	Ind, Year		Ind, Year	

Table A.4: Alternative Competence Measure: Phase-specific Difficulties

The table reports the effects of diversification on (1) Going public and (2) VC funding using an instrumental variable approach with the passage of Medicare Part D legislation as in Table 9 using an alternative measure of R&D competence. This alternative is a conditional success probability weighted R&D competence that reflects phase-specific difficulties of transition based on success rates estimated in DiMasi, Hansen, and Grabowski (2003). *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation. *Post* is one after the passage of the Medicare Part D legislation in 2003. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO at a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.010**		0.040*
		(0.01)		(0.02)
Treated * Post * Low competence	1.231***		1.226***	
	(0.39)		(0.39)	
Treated * Post	-0.469*		-0.461*	
	(0.27)		(0.27)	
Treated * Low competence	-1.249***		-1.249***	
	(0.38)		(0.38)	
Post * Low competence	-0.423		-0.425	
	(0.31)		(0.31)	
Treated	0.828***		0.821***	
	(0.31)		(0.31)	
Low competence	0.361	-0.020**	0.363	-0.141***
	(0.31)	(0.01)	(0.30)	(0.05)
Log(# of phase 0 projects)	0.439***	0.002	0.439***	0.023
	(0.06)	(0.02)	(0.07)	(0.06)
Log(1+firm age)	0.131***	0.003	0.134***	0.006
	(0.02)	(0.00)	(0.02)	(0.03)
% matured projects	0.460***	-0.009	0.456***	-0.395***
	(0.06)	(0.01)	(0.06)	(0.07)
% projects with partner	0.190***	0.007*	0.189***	0.009
	(0.03)	(0.00)	(0.03)	(0.03)
Log(# of competitors)	0.740***	0.021	0.738***	-0.013
	(0.19)	(0.04)	(0.19)	(0.21)
Ind failure rate	0.146	-0.395***	0.141	-0.542
	(0.43)	(0.13)	(0.43)	(0.78)
Ind % matured projects	1.106***	0.178*	1.099***	-0.470
	(0.31)	(0.10)	(0.31)	(0.49)
VC backed	0.029	0.049***		
	(0.02)	(0.01)		
Observations	3658		3658	
Fixed Effects	Ind, Year		Ind, Year	

Table A.5: Medicare Part D Shock on Diversification and Firm Growth: Entry Year Effects

The table reports the effects of diversification on (1) Going public and (2) VC funding using an instrumental variable approach with the passage of Medicare Part D legislation as in Table 9 with the exception of entry year fixed effects instead of year fixed effects. *Treated* is one if a firm does not have any projects in the required Part D drug classes before the legislation. *Post* is one after the passage of the Medicare Part D legislation in 2003. In the first two columns, the dependent variable, *Going Public* is one if the firm went public via an IPO at a given year, and zero otherwise. In the last two columns, the dependent variable, *VC funding* is the log of one plus the VC funding amount in a given year. The mediator variable, *Diversification* is the total number of different disease groups where the firm has projects. Industry and entry year fixed effects are included. *t-statistics* (in parenthesis) are robust and adjusted for clustering within industry and year. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

	Diversification	Going public	Diversification	VC funding
	(1)	(2)	(3)	(4)
Diversification		0.012**		0.074***
		(0.01)		(0.02)
Treated * Post * Low competence	0.436**		0.415**	
	(0.20)		(0.20)	
Treated * Post	-0.704***		-0.671***	
	(0.13)		(0.12)	
Treated * Low competence	-0.407**		-0.399**	
	(0.19)		(0.20)	
Post * Low competence	0.238*		0.229	
	(0.14)		(0.14)	
Treated	1.282***		1.254***	
	(0.22)		(0.22)	
Low competence	-0.331**	-0.020***	-0.322**	-0.131***
	(0.14)	(0.01)	(0.14)	(0.05)
Log(# of phase 0 projects)	0.500***	0.010	0.500***	-0.009
	(0.06)	(0.01)	(0.06)	(0.06)
Log(1+firm age)	-0.053***	0.012***	-0.044**	0.053*
	(0.02)	(0.00)	(0.02)	(0.03)
% matured projects	0.274***	0.001	0.262***	-0.354***
	(0.07)	(0.02)	(0.07)	(0.07)
% projects with partner	0.023	0.016***	0.020	0.040
	(0.02)	(0.01)	(0.02)	(0.04)
Log(# of competitors)	0.529***	0.110***	0.518***	-0.048
	(0.08)	(0.03)	(0.08)	(0.15)
Ind failure rate	0.212	-0.127*	0.199	-1.255***
	(0.23)	(0.07)	(0.23)	(0.38)
Ind % matured projects	0.436*	0.317***	0.427*	-0.008
	(0.25)	(0.08)	(0.25)	(0.40)
VC backed	0.076***	0.046***		
	(0.02)	(0.01)		
Observations	3851		3851	
Fixed Effects	Ind, Entry Year		Ind, Entry Year	