Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on Ownership and Innovation

Autor:

Fernando Jorge Alves dos Santos

Directores:

Andrea Fosfuri

Neus Palomeras Vilches

Tutor:

Neus Palomeras Vilches

DEPARTAMENTO DE ECONOMÍA DE LA EMPRESA

Getafe, Noviembre de 2017

Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on Ownership and Innovation

Autor: *Fernando Jorge Alves dos Santos*

Directores: **Andrea Fosfuri**

Neus Palomeras Vilches

Firma del Tribunal Calificador:

Firma

Presidente:

Vocal:

Secretario:

Calificación:

Getafe, de de

To my supervisors for guiding and pushing, my parents for encouraging, my wife for being so patient and my son for making me more persevering.

Summary

We investigate three distinct 'bridges' between companies' ownership characteristics and innovation.

In the first chapter, we take an exploratory empirical approach to investigate the effect of going public on the risk-reward characteristics of the innovation portfolio. We argue that an IPO is a life-changing event that encompasses many changes that concur to impact the risk-reward characteristics of the innovation portfolio. However, we expect that an IPO is a more severe event for companies that were originated as new and independent entities, and remain independent till the IPO event – emerging growth companies – than it is for other companies – non-emerging growth companies. We find that going public does not affect the risk-reward characteristics of non-emerging companies. For emerging growth companies we find that the risk-reward characteristics of the innovation portfolio are: positively impacted by going public; and, negatively impacted by the percentage of shares offered at the IPO.

In the second chapter, we investigate pharmaceuticals' decision between acquiring and allying with a biotechnological, after recombinant DNA. The discovery of R-DNA represented an R&D competence-destroying event for incumbent pharmaceutical companies. As a response to that, organizational deals – namely, strategic alliances and mergers and acquisitions – have been extensively used as external sources of knowledge. We examine a potential determinant of pharmaceuticals' decision between these two alternatives: the width of applicability of the internal knowledge of the biotechnological company. We find compelling evidence that it has a positive effect on the likelihood that the pharmaceutical company acquires (or merges with) that biotechnological company.

In the third chapter, we examine the impact of SOX Section 404 on long-term investment of small innovative companies. We hypothesise that R&D intensity increases the impact of SOX 404 on long-term investment. Making use of a quasi-natural experiment, our results suggest that the impact of SOX Section 404 on companies' long-term investment is uneven, favouring R&D intensive companies. This may call for a re-centring of policy discussion around the distribution of the net benefit of coercive financial disclosure programs versus the overall economic impact of those programs.

Contents

Initial Public Offer and the Risk-Reward Characteristics of the Innovation Portfolio

Fernando Santos
Universidad Carlos III de Madrid
Department of Business Administration
fjasantos@gmail.com | fasantos@porto.ucp.pt

PhD BF | Working Paper
Supervisors: Andrea Fosfuri and Neus Palomeras

ABSTRACT

We take an exploratory empirical approach to investigate the effect of going public on the risk-reward characteristics of the innovation portfolio. We argue that an IPO is a life-changing event that encompasses many changes that concur to impact the risk-reward characteristics of the innovation portfolio. However, we expect that an IPO is a more severe event for companies that were originated as new and independent entities, and remain independent till the IPO event – emerging growth companies – than it is for other companies – non-emerging growth companies. We find that going public does not affect the risk-reward characteristics of non-emerging companies. For emerging growth companies we find that the risk-reward characteristics of the innovation portfolio are: positively impacted by going public; and, negatively impacted by the percentage of shares offered at the IPO.

INTRODUCTION

Does going public affect the risk-reward characteristics of the innovation portfolio? A few relevant statements suggest that it might. Regarding the IPO possibility, Facebook's founder stated: *"I tend to think that being private is better for us right now because of some of the big risks we want to take in developing new products. […] The experience of managing the company through launching controversial services is tricky, but I can only imagine it would be even more difficult if we had a public stock price bouncing around. There are a lot more new things left to build […] and I'd rather focus on building them than on going public right now"*[1]. On another example, Google's 2004 Founders' IPO Letter[2] addresses concerns regarding the changes on the innovation strategy driven by the emergence of conflict of interests: *"We will not shy away from high-risk, high-reward projects because of short-term earnings pressure. Some of our past bets have gone extraordinarily well, and others have not. Because we recognize the pursuit of such projects as the key to our long-term success, we will continue to seek them out. […] We will support selected high-risk, high-reward projects and manage our portfolio of projects."* Regarding Google's prospectus, Larry Page[3] argued, *"many companies are under pressure to keep their earnings in line with analysts' forecast. Therefore, they often accept smaller, but predictable, earnings rather than larger and more unpredictable returns. Sergey and I feel this is harmful, and we intend to steer in the opposite direction."*

Changes in innovation investment decisions seem to be a major concern. Yet, the analysis of how an IPO impacts innovation investment decisions is challenging for several reasons. First, an IPO is a major financing decision that impacts companies in multiple ways [Pagano *et al.* (1998)], many of which concurring to impact innovation, possibly in opposing directions [Wu (2012)]. Second, an IPO impacts companies distinctively, depending on whether or not the company is originated as new and independent entity, and remained independent till the IPO event [Pagano *et al.* (1998)][4]. Third, going public can either affect the level of innovation efforts and/or it

[1] www.businessinsider.com.
[2] From the S-1 Registration Statement. The content of the S-1 form is critical for any potential shareholder to decide whether or not to take part in a particular IPO. Insiders share information and manifest their intentions and interests, which is highly valuable in a context of so highly asymmetric information.
[3] Co-founder of Google, alongside with Sergey Brin.
[4] Hereafter we designate these companies as emerging growth companies (EGC). Non-emerging growth companies (NEGC) engaged in an IPO are typically divisions or subsidiaries of large diversified public companies. Divesture does not necessarily involve an IPO, although it does in many cases [Draho (2004)].

can impact the type of innovation efforts. This is because the fundamental differences between innovation and other more conventional investments – namely, innovative activities are harder to manage and communicate to outsiders, less tangible, have more uncertain and long-term returns – may also exist between innovative activities, and are the base for the development of empirical predictions. To illustrate, consider the following empirical prediction. By lowering owners' tolerance for failure and increasing their focus on short-termism, going public is expected to decrease investment in projects that are typically long-term. One can argue that, since innovation is long-term (as compared with conventional projects), the level of innovation efforts would decrease. But we also could argue that the fundamental differences between innovation and conventional investment – namely, innovative activities are harder to manage and communicate to outsiders, less tangible, have more uncertain and long-term returns – can be observed within the innovation portfolio. As such, instead of affecting the level of innovation efforts, it could be that the type of innovation efforts would be impacted, through a higher investment in less conservative innovative projects relative to more conservative innovative projects.

Accordingly, we assess both the impact of going public on the risk-reward characteristics of the innovation portfolio – which is associated with the type of innovation efforts – and on the overall reward of the innovation portfolio – which is associated with the level of innovation efforts. Yet, our main contribution is on assessing the former.

Conceptually, we consider an innovation portfolio with higher risk as one with higher level of less conservative projects within the innovation portfolio – namely, an innovation portfolio with relatively more projects that are harder to manage and communicate to outsiders, less tangible and have more uncertain and long-term returns. We assume the reward characteristics of the innovation portfolio (and not the overall reward of innovation), to be associated with the risk of the innovation portfolio[5]. We are obviously close to the existing literature on the relation between IPO and innovation [Wu (2012)[6]; Aggarwal and Hsu (2014)[7]; and, Bernstein (2015)[8]].

[5] To clarify the distinction between overall reward of innovation and reward characteristics of the innovation portfolio, consider the following scenario. Suppose a company decides to double its level of innovation efforts, without changing the type of innovation efforts. In this case, the overall reward of the innovation portfolio is expected to increase (perhaps, double), while the reward characteristics of the innovation portfolio is not impacted.

[6] It is a single-industry study (medical device). Finds negative impact on innovation quality, negative impact on exploration of new and recently developed knowledge and positive impact on exploration based in scientific knowledge.

[7] It is a single-industry study (biotech). Finds negative impact on innovation quality.

[8] It is a multi-industry study. Finds negative impact on innovation quality, negative impact on originality and no impact on generality.

Our research is closer to Wu (2012), and Bernstein (2015) because, as claimed before, we also inspect the characteristics of the innovation portfolio around the IPO. Bernstein (2015) investigates the impact of going public on generality and originality (in his designation, the fundamental nature of research). Wu (2012), on the other hand, investigates the impact on exploitation *versus* exploration. Our proposal is distinct from both and is closer to Wu (2012). Wu (2012) hypothesises on the proportion of exploratory search in the innovation portfolio based on the fundamental differences between exploitation and exploration – namely, relative to exploitation, returns from exploration are uncertain, distant and often negative [March (1991)]. We, on the other hand, are interested on the risk-reward characteristics of the innovation portfolio, which is operationalized by inspecting the project and time distribution of returns of the innovation portfolio – and these distributions, in turn, bases the distinctive features of exploitation/exploration. As such, besides complementing these proximate contributions – those that inspect the impact of going public on innovation – with novel evidence from a distinct dimension of innovation portfolio characterization, we believe this alternative approach to inspect innovation portfolio characteristics can be valuable for research streams that build upon the fundamental differences between innovation and other more conventional investments and/or the fundamental differences between exploitation and exploration.

We argue that the risk-reward of the innovation portfolio reflects companies' investment decisions, and that these are impacted by the IPO due to changes in companies' financial constraints and investment preferences of their owners and possibly their managers (in the case of an imperfect agency relation). We also argue that going public may impact each of these dimensions in distinct ways (what we designate by drivers). Another distinguishing feature of the paper is that, while we expect that an IPO is likely to be a more severe event for emerging growth companies (EGC) than it is for non-emerging growth companies (NEGC), we decide to investigate both.

We combine data from SDC, USPTO's TAF and Kenney and Patton (2013) databases and take an exploratory empirical approach that encompasses comparing companies that applied for an IPO and completed with those that withdrew[9], to assess the impact of going public on EGC and NEGC. Because we find it likely that the percentage of

[9] Following Bernstein (2015).

shares offered at the IPO captures a different set of drivers, we also assess the impact of percentage of shares offered at the IPO on EGC.

We find that, contrary to EGC, going public does not affect the risk-reward characteristics of NEGC. Moreover, for EGC: going public affects positively the risk-reward characteristics of the innovation portfolio; and, the percentage of shares offered at the IPO impacts negatively the risk-reward characteristics of the innovation portfolio.

LITERATURE

The impact of going public on the risk-reward characteristics of the innovation portfolio is fairly hard to predict in general. We claim that the risk-reward of the innovation portfolio reflects companies' investment decisions, and that these depend on: companies' financial constraints; and, investment preferences of their owners and possibly their managers (in the case of an imperfect agency relation). As described below and summarized in Table 1, going public may impact these dimensions in distinct ways.

As argued above, because the fundamental differences between innovation and other more conventional investments are the base for the development of empirical predictions, we assume that these fundamental differences also exist between innovative activities to develop predictions for both the level of innovation efforts and the type of innovation efforts.

Changes in financial constraints

Going public is an equity financing decision that relaxes companies' financial constraints. Besides the obvious immediate financial slack through IPO net proceeds[10], there is also a long-term effect[11]: it improves assess to equity financing [Bernstein (2015)][12]. It has been suggested that the access to stock market financing is particularly relevant for innovative activities (as opposed to more conventional investments) because equity (as opposed to debt) is a more suitable source of funds [Carpenter and Petersen (2002); Hall (2002); Brown *et al.* (2009); Hall (2010); Hall and Lerner (2010); and, Brown *et al.* (2013)]. The fundamental differences between innovative and more conventional investments justify this prediction. Namely, comparatively to other investments, R&D and innovative activities: are less tangible and thus lack of collateral (often needed in debt contracts); are harder to communicate to outsiders; have more skewed and uncertain returns, which does not fit a contract in which only downside returns are shared (often the case in debt contracts); and, have longer-term returns. As such, by easing access to equity financing, going public is

[10] There are non-negligible direct costs of going public such as: underwriting commissions; legal, printing and auditing expenses; etc. Using a sample of U.S. IPO between 1972 and 1982, Ritter (1987) estimates direct costs of $250 thousand plus 7% of gross proceeds.

[11] We focus on long-term effect.

[12] Moreover, and even though it is not a central argument in our analysis, it is worth noting the following. Since going public also requires higher information disclosure to outsiders, arm's-length debt (for example, bonds) may become relatively more attractive, which may in turn ease access to bank debt [Rajan (1992)].

expected to foster overall innovation efforts (which is expected to impact the overall reward of innovation) and/or a high-risk innovation portfolio. Hence, going public may ease access to equity financing, which in turn may increase the overall reward of innovation and/or the risk-reward embedded in the innovation portfolio.

There is an opposing argument, though. First, while relative to debt financing, equity financing is preferable to finance R&D and innovative activities, internal financing[13] is also quite appropriate [Carpenter and Petersen (2002); Hall (2002); Brown *et al.* (2009); Hall (2010); Hall and Lerner (2010); and, Brown *et al.* (2013)]. This is because R&D and innovative activities are associated with higher levels of information asymmetry between insiders (management) and outsiders (owners and creditors). Second, it has been found that, comparatively to private companies, public companies pay higher dividends and adopt a smoother dividend policy [Michaely and Roberts (2012)]. Accordingly, by favouring dividend policy in the use of internal finance, going public may decrease the overall reward of innovation and/or the risk-reward embedded in the innovation portfolio.

[Table 1 around here]

Changes in investment preferences

The risk-reward characteristics embedded in the innovation portfolio are also sensitive to changes in investment preferences. Going public is likely associated with changes of this nature because it implies significant changes in the ownership-control characteristics of the firm[14]. In particular, going public is simultaneously associated with ownership dilution, enhancement of diversification possibilities, rise of short-term earnings pressure and rise of the agency tensions. These are expected to impact owners' preferences and managers' preferences as follows.

Owners' preferences

Going public impacts the ownership structure of companies. The fact that companies get listed in an organized exchange improves the liquidity of its equity and fosters its visibility [Pagano *et al.* (1998)]. It is likely for post-IPO ownership structure to be relatively more diluted. At the same time, it is also likely that, comparatively to pre-

[13] Internally generated cash flows.

[14] The literature that addresses the relationship between ownership-control characteristics and innovation is extensive, mainly empirical based, and traditionally uses theoretical insights from agency theory [Jensen and Meckling (1976)] and upper-echelons perspective [Hambrick and Mason, 1984].

IPO owners, post-IPO owners have a smaller fraction of their total wealth invested in the company[15]. Accordingly, pre-IPO owners are expected to avoid more the idiosyncratic risk because they are not able to diversify their investment portfolio as much as post-IPO owners. Moreover, after the lock-up period[16] the pre-IPO owners (in particular, managers and founders) can sell their equity stake, which will further enhance the diversification possibilities because: it enhances diversification possibilities of pre-IPO owners [Pagano *et al.* (1998)]; and, it is likely that post-lock-up period new owners have higher diversification possibilities that the pre-IPO owners. As such, by allowing higher levels of risk diversification for owners, going public is expected to increase the overall reward of innovation and/or the risk-reward embedded in the innovation portfolio.

Again, existing literature enable predictions in the opposite direction: private (public) ownership favours innovative (conventional) projects. The reasoning is as follows. On the one hand, owners of private companies can profit by adopting an exit strategy after receiving bad news, making them relatively tolerant to early failure; and, on the other hand, because market prices of public stocks quickly incorporate new information, owners of public companies have relatively more incentives to favour investments with higher short-term earnings (conventional projects) [Ferreira *et al.* (2014)][17]. As such, by lowering owners' tolerance for failure and increasing their focus on short-termism, going public is expected to decrease the overall reward of innovation and/or the risk-reward embedded in the innovation portfolio.

Managers' preferences

There are two reasons for taking into account managers' preferences. First, an IPO boosts the conflicts of interest between those who take corporate investment decisions (managers) and those who bear the risk embedded in those decisions (owners) [Jensen (1989)]. It also implies an immediate decrease in managers' equity claims, which lowers its incentives to focus on innovative and risky activities because they are harder to manage [Jensen and Meckling (1976)][18]. Second, since after the lock-up

[15] The portfolio theory argues the optimality of diversification [Fama (1980)]. Others link IPO to diversification [Zingales (1995); Black and Gilson (1998); Pagano *et al.* (1998); Certo *et al.* (2001); and, Certo *et al.* (2009)].
[16] The lock-up period is usually a 4 to 6 month window starting on IPO date in which the pre-IPO owners cannot sell their shares.
[17] Private companies differ from public companies in many ways. The essential difference for this argument is private companies are not required to publicly disclose information.
[18] 'on-the-job' utility argument.

period the manager can sell its equity stake, its incentives to focus on innovative and risky activities further decrease[19].

As such, by lowering managers' incentives for engaging on innovative and risky activities, going public is expected to decrease the overall reward of innovation and/or the risk-reward embedded in the innovation portfolio.

EGC *versus* NEGC

NEGC engaged in an IPO are typically divisions or subsidiaries of large diversified public companies [Draho (2004)]. Since the subsidiaries might get some benefit – for example, indirectly accessing public equity market – and also bear some of the costs of being public – for example, share the costs associated with disclosure requirements of the parent company – [Pagano *et al.* (1998)], we expect that an IPO will impact each of the drivers in a more significant way in the case of EGC, relative to NEGC.

IPO *versus* percentage of shares

Going public imply changes on financial constraints and on investment preferences. Yet, some of these changes are associated with the change in companies' type of ownership – public *versus* private – while others are associated with the extent to which companies are public (or private) – the percentage of shares offered at the IPO. For example, if a company completes and IPO, it is expected to comply with a set of disclosure requirements, independently on the percentage of shares offered at the IPO. Because "access to equity" and "tolerance-for-failure" are mostly impacted by changes in information asymmetry, we expect these drivers to be mostly insensitive to the percentage of shares offered at the IPO.

[19] 'on-the-job' utility argument, again.

EMPIRICS

We argue that investment decisions depend on financial constraints and on (owners and managers) investment preferences, and that an IPO can affect (directly or indirectly) all. This results in a set of empirical predictions that point in opposing directions, making it hard to predict the impact of going public on the risk-reward characteristics of the innovation portfolio. We take an exploratory empirical approach.

Empirical strategy

Our exploratory empirical approach focuses on U.S. companies and encompasses 2 main blocks of exercises enabled by combining data from SDC, USPTO's TAF and Kenney and Patton (2013) databases. We provide evidence on the impact of going public for EGC and NEGC, and on the impact of percentage of shares offered at the IPO for EGC.

IPO

To capture the effect of going public, we compare companies that applied and completed an IPO with those that applied but withdrew the IPO application[20]. We start by analysing the within-company changes in a set of innovation variables around the IPO event, for companies that completed an IPO and for companies that applied but withdrew the IPO, separately. For each subset of companies, the following form is estimated through OLS, for each dependent variable *Y*.

$$Y = \alpha + \sum_{k=0}^{7} \beta_k . Year_k + \theta . X + \varepsilon$$

$Year_k$ is a dummy variable indicating the distance in years relative to the IPO year (k=0 indicates the IPO year and is omitted). X is a vector of controls that include individual effects and time effects.

In an alternative approach, we try the following specification.

[20] Companies apply for an IPO by submitting an initial registration statement to the Security Exchange Commission, demonstrating an intention to go public. Yet, the decision to actually go public is differed in time. Companies have an option to withdraw (cancel or postpone) the IPO filling during the book-building phase that follows the filling.

$$Y_{after} = \alpha + \gamma . Y_{before} + \beta . IPO + \theta . X + \varepsilon$$

Y_{before} and Y_{after} are innovation variables measured before and after the IPO, respectively, as defined on data section. *IPO* is a dummy variable that equals to 1 when the firm completed the IPO and 0 otherwise. *X* is a vector of controls that include industry effects and time effects.

Further insights on the impacts of going public on the risk-reward characteristics of the innovation portfolio can be obtained by dividing the sample into EGC and NEGC, for which we follow Kenney and Patton (2013) definition by considering as EGC those that were originated as new and independent, and remain independent till the IPO date[21]. We think this is an informative exercise because, as argued by Draho (2004), corporate restructuring of larger firms frequently involves an IPO. As argued above, we expect that the drivers identified on literature section to be especially relevant in the case of EGC.

This constitutes an appealing setup because it enables comparison of companies that completed an IPO (either EGC or NEGC) with companies that applied and withdrew the IPO (either EGC or NEGC). That is, it enables comparison between a treatment group (those that completed) and a control group (those that withdrew), where both groups are expected to be similar enough because they are in the same stage of the life-cycle (they both apply for an IPO)[22].

Percentage of shares

To capture the effect of percentage of shares offered at the IPO, we try the following specification, for each innovation variable *Y*.

$$Y_{after} = \alpha + \gamma . Y_{before} + \beta . Shares + \theta . X + \varepsilon$$

Y_{before} and Y_{after} are innovation variables measured before and after the IPO, respectively, as defined on data section. *Shares* is the percentage of shares that go public at the IPO date, as defined on data section. *X* is a vector of controls that include industry effects and time effects.

[21] Ideally, we would need an identifier for EGC. However, this data is not available. We try to overcome this using a standard imputation procedure. We explain this in more detail on data section.

[22] The evidence on the validity of the control group is provided below.

As argued above, we provide evidence on the impact of percentage of shares offered at the IPO because we find it likely that the percentage of shares offered at the IPO captures a different set of impact drivers. Namely, the percentage of shares offered at the IPO is likely related (or considerably more related) with the following drivers: dividend policy (change in financial constraints); diversification (change in owners' investment preferences; and, on-job-utility (change managers' investment preferences). Unfortunately, due to data constraints, we are only able to provide evidence for EGC.

Data and variables

We use the SDC database to identify U.S. companies that applied and completed the IPO and U.S. companies that applied and cancelled or postponed (withdrew) the IPO, in the U.S. market. We use the USPTO's TAF database to define and characterize the patent portfolios of these firms[23]. Then, we combine these with Kenney and Patton (2013) database on emerging growth IPO to split the data into EGC and NEGC[24] and to collect data on the percentage of shares offered at the IPO, for EGC.

We start by selecting all IPO with issue dates between January 01 1985 and December 31 1993 – 5284 observations, of which 4462 completed while 882 withdrew the filing. As there are multiple observations per company[25], we keep the oldest and find a total of 4800 observations. From these, 4038 filings were actually completed, while the remaining 762 withdrew the filing. We start in 1985 because this is the year from which the SDC systematically covers withdraws. We end in 1993 to avoid truncation on the patent data.

We then use the fuzzy lookup algorithm to match the 'issuer' variable from SDC with the 'standard_name' variable from the NBER. We drive the algorithm twice and keep those companies that have the same match with 0.95 or higher on the similarity index[26]. We get a total of 1042 matched companies – 915 completed and 127 withdrew[27]. For these companies we identify the 65003 utility patents in the USPTO's TAF database, granted from 1976 to 2006. From these, 17234[28] were granted between (and including) four years before the IPO and 7 years after the IPO. We define these

[23] Hall *et al.* (2001).
[24] Through a standard imputation procedure, explained in detail below.
[25] Variable 'standard_name' of SDC.
[26] We also exclude those cases for which different issuers have the same match at NBER.
[27] The list of companies is available upon request.
[28] 1 out of the 17234 patents is owned by 2 eligible companies, so, in fact there are 17233 different patents in the dataset.

as being the eligible patents for the analysis. Restricting the analysis to companies that had at least one eligible patent implies loosing 162 firms. Hence, out of the 1042 firms, we keep 880 (777 completed and 103 withdrew).

For each of the 17234 eligible patents, we identify all citations received – 338168 citations. In order to avoid truncation, we use the following citation inclusion criteria. We only consider the citations that happened on the first 6 years of each patent life (-1 years < *cited-citing lag*[29] < 7 years)[30]. Out of the 17234 eligible patents, 15433[31] patents have at least one eligible citation, while others do not[32]. We end up with 153947 eligible citations[33]. Table 2 briefly describes what results from this process.

[Table 2 around here]

Year-by-year innovation variables

We define the patent portfolio of company *i* at time *k* as all patents granted to company *i* between (and including) *k-4* and *k*. We set the IPO year as *k=0*, and define the patent portfolio of each company *i* for *k=0, 1, 2, 3, 4, 5, 6 and 7*. That is, for every company *i* and every time *k,* we define the set of eligible patents to be included in the patent portfolio of company *i* at time *k* as all patents granted to company *i* between (and including) *k-4* and *k*.

For each time *k* and company *i*, we analyse the following innovation variables. (1) *Patents* – number of eligible patents; (2) *Cites* – number of eligible citations of eligible patents; (3) *Average Cites* – number of eligible citations divided by number of eligible patent; (4) *Dispersion Cites* – standard deviation of the eligible citations per eligible patent; (5) *Mean Lag Forward Cites* – average *cited-citing lag* of eligible patents. For eligible patents with eligible citations, this average is computed. Eligible patents with no eligible citations are assumed to have a mean lag of forward cites of 7 years – which is larger than the maximum value that this variable can take for eligible patents with eligible citations (6 years). The outcome is computed at the portfolio level, as a simple average over eligible patents; (6) *Dispersion Lag Forward Cites* – standard deviation of the *cited-citing lag* of eligible patents; and, (7) *Old* – average *k-*

[29] *cited-citing lag* is defined as the difference between the grant year of the citing patent and the grant year of the cited patent.
[30] This is why we confine the IPO dates to 1993. For these IPOs we need eligible patents till 2000 for which we need 7 years of observations of citations (till 2006). And citation data is only available till 2006.
[31] More precisely, 15432, as the patent that is shared by two companies has at least one eligible citation.
[32] We considered these 1801 patents to have no eligible citation. 703 were not cited at all, till 2006; 1098 were cited but don't fulfil the citation inclusion criteria.
[33] More precisely, 153944 as the patent shared by the two companies has 3 eligible citations.

cited lag[34] of eligible citations of eligible patents. Table 3 presents the descriptive statistics of the dependent variables for all 880 companies in our sample.

[Table 3 around here]

We are aware that using patent information to measure innovation is sensitive to changes in patenting strategy. Yet, (1) *Patents* is a widely used measure of overall innovation return. However, it is also widely recognized that patents may have distinct importance/quality. (2) *Cites* is typically used as a measure of innovation return that overcomes this difficulty[35]. (3) *Average Cites* and (4) *Dispersion Cites* enable a deeper insight of the project distribution of returns in the patent portfolio. (3) *Average Cites* captures the average return of the innovation portfolio and (4) *Dispersion Cites* capture the extent to which returns embedded on the patent portfolio are concentrated around the mean. (5) *Mean Lag Forward Cites* and (6) *Dispersion Lag Forward Cites* enable a deeper insight of the time distribution of returns in the patent portfolio. (7) *Old* investigate whether the returns embedded on the patent portfolio come from recently granted patents or not.

Key innovation variables

We focus on the risk-reward characteristics of the innovation portfolio, which depend on the type of innovation efforts. As argued before we consider an innovation portfolio with higher risk as one with higher level of innovation efforts on less conservative projects within the innovation portfolio – namely, an innovation portfolio with relatively more projects that are harder to manage and communicate to outsiders, less tangible and have more uncertain and long-term returns. So, riskier projects are expected to be associated with more uncertain and longer-term returns. In the former, we should empirically observe a higher dispersion of returns – captured by (4) *Dispersion Cites*. In the later, we should empirically observe a higher (5) *Mean Lag Forward Cites*. As argued before, we expect the reward characteristics of the innovation portfolio, to be associated with the risk of the innovation portfolio. We

[34] *k-cited lag* is defined as the difference between the grant year of the cited patent and *k*. As such, this will take values between (and including) 0 and -4.

[35] We are aware that alternative measures for patent quality exist – namely, the number of claims could be used instead of forward cites.

think that (3) *Average Cites* captures this particularly well because it is insensitive to changes in the level of innovation efforts[36].

Before-and-after innovation variables

Y_{before} and Y_{after} are innovation variables measured before and after the IPO, respectively. We define those as the average value of the innovation variables when *k=0,1,2 and 3* and *k=4,5,6 and 7*, respectively. We deliberately consider post-IPO years for defining Y_{before} to account for the typical lag between investment decisions and patenting decisions. In general, results are robust to changes in the before-after threshold.

Table 4 compares the means of each innovation variable, measured at the pre-IPO (Y_{before}). It is interesting to note that companies that complete an IPO present higher overall return of the innovation portfolio – measured by (2) – and an innovation portfolio with higher risk – measured by (4). Because our Pre-IPO variables (Y_{before}) include citations that happen after the IPO event, we may attribute these differences to an increase in patent visibility attributed to the IPO itself. As such, while some differences exist among the observables related with innovation outcomes, we interpret these results as evidence of validity of the control group.

[Table 4 around here]

Imputation procedure

As argued before, ideally, we would need an identifier for EGC because the motives to and impacts of going public of an independent company likely differ from those of a subsidiary of a public company. Unfortunately this data is not available. We try to overcome this with the following imputation procedure, which results are described on Table 5.

[Table 5 around here]

We match our previous SDC-USPTO's TAF combination (880 companies) with Kenney and Patton (2013) database. This database encompasses all (and only) emerging growth IPO on American stock exchanges that were filed with the SEC

[36] As opposed to (2) *Cites*, which captures particularly well the overall reward of innovation.

from January 1990 through December 2010. As such, we are only able to match IPO fillings that happened between (and including) 1990 and 1993, that were EGC and that completed an IPO. We matched names using the fuzzy lookup algorithm and keep the cases with 0.95 or higher on the similarity index and we find a total of 277 matches[37]. We rely on the precision of this procedure to distinct EGC and NEGC (for companies that complete an IPO between 1990 and 1993). The imputation procedure further uses all seven pre-IPO innovation variables of interest (Y_{before}) to predict an EGC status (for companies that applied for an IPO between 1985 and 1993)[38]. This procedure leads to 362 EGC and 315 NEGC[39].

Percentage of shares

Kenney and Patton (2013) database provides data on the number of shares offered and the number of shares outstanding at the IPO date. We then define our independent variable of interest *Shares* as the ratio between both – the percentage of shares offered at the IPO. The histogram is presented on Figure 1. It reveals that the percentage of shares involved at an EGC IPO is variable, though typically below 50%.

[Figure 1 around here]

Results

IPO

For firms that completed the IPO, results reported on Panel A of Table 6 capture a positive trend on (4) *Dispersion Cites*, and no trend on remaining 2 key innovation variables – (3) *Average Cites* and (5) *Mean Lag Forward Cites*. The results for companies that withdrew the IPO, reported on Panel B of Table 6, are of lower global significance. Moreover, taken together with results provided on Table 4, we find no evidence of a mean reversion mechanism.

[Table 6 around here]

[37] In this case, this turned out to be quite effective because they also rely on SDC.
[38] We are aware of the limitations of such procedure and find that this would be, perhaps, the first effort to be taken by future research.
[39] This implies that the procedure does not predict an ECG status for 203 companies.

Table 7 presents the effect of completing an IPO on each innovation variable. Results suggest that the risk-reward characteristics of the innovation portfolio are not significantly impacted[40,41].

[Table 7 around here]

Table 8 presents the effect of going public for the 677 companies (Panel A), for EGC (Panel B) and NEGC (Panel C). Results of Panel A are relatively similar to those of Table 7. Yet, the separate analysis between EGC and NEGC reveal quite distinct results. The risk-reward characteristics of NEGC are not impacted, suggesting no change on the type of innovation efforts. Yet, The risk-reward characteristics of EGC are significantly impacted. Results for EGC are statistically significant only for the project distribution of returns – (3) *Average Cites* and (4) *Dispersion Cites* – suggesting that the change on the type of innovation efforts is towards projects with more uncertain returns. The difference in results sustain our idea that the drivers identified on literature section to be especially relevant in the case of EGC. Moreover, taken together with results provided on Table 5 (Panels D and E), we find no evidence of a mean reversion mechanism[42].

[Table 8 around here]

Percentage of Shares

Table 9 presents the effect of shares offered on each innovation variable. The percentage of shares offered at the IPO impacts negatively the risk-reward characteristics of the innovation portfolio. Again, only the project distribution of returns is affected[43]. If the percentage of shares captures only the drivers "dividend policy", "diversification" and "on-the-job utility"[44], these results suggest that the

[40] Except a slightly significant lower (5) *Mean Lag Forward Cites*.
[41] As a robustness check, Table 11 on Appendix B is similar to Table 7, except for the definition Y_{before} and Y_{after}. In this case, Y_{before} and Y_{after} are defined as the average value of the innovation variables when *k=0,1 and 2* and *k=3,4,5,6 and 7*, respectively. In general, the direction of impacts is the same and statistical significance varies slightly.
[42] As a robustness check, Table 12 on Appendix B is analogous to panels B and C of Table 8, except that it is a Difference-in-Differences approach. In general, the direction of impacts is the same and statistical significance drops significantly for key dependent variables.
[43] As a robustness check: we discretised *Shares* and obtained analogous results; we also tried a different definition for Y_{before} and Y_{after} – defined as the average value of the innovation variables when *k=0,1 and 2* and *k=3,4,5,6 and 7*, respectively, and obtained similar results. These are available upon request.
[44] Our claim here is that we expect that the remaining drivers ("access to equity" and tolerance-for-failure and short-term pressure") to be particularly relevant when the company becomes public, but likely insensitive to *Shares*.

driver “diversification” is dominated by the drivers “dividend policy” and/or “on-the-job utility”.

Comparing these results with those of Panel B of Table 8 is further informative. If we assume that the average impact of going public that is jointly attributable to the drivers captured by the percentage of shares (“dividend policy”, “diversification” and “on-the-job utility”) is also negative, it follows that the drivers “access to equity” dominates the impact of “tolerance-for-failure and short-term pressure”.

[Table 9 around here]

DISCUSSION

We argue that going public may impact the risk-reward characteristics of the innovation portfolio because it potentially impacts companies' financial constraints, and investment preferences of their owners and possibly their managers, each of which driving (possibly opposing) impacts on the type of innovation efforts. Combining data from SDC, USPTO's TAF and Kenney and Patton (2013) databases and focusing on U.S. companies, take an exploratory approach that encompasses assessing the impact of going public by EGC and NEGC and assessing the impact of shares offered by EGC at the IPO. Our main findings are the following.

First, results on time distribution of results – in which we include one dimension of innovation portfolio risk, (5) *Mean Lag Forward Cites* – are generally fuzzy and mostly non-significant. This suggests that the impact of going public is not clearly towards a more explorative portfolio nor it is to a more exploitative portfolio. Or alternative, we don't see a significant impact on the time dimension variable that typical is incorporated in the definition of explorative versus exploitative research – (5) *Mean Lag Forward Cites*. This is a particularly interesting result when set in parallel with evidence from Wu (2012).

Second, contrary to NEGC, results in project distribution of results – in which we include another dimension of innovation portfolio risk, (4) *Dispersion Cites*, and the innovation portfolio reward, (3) *Average Cites* – are strongly significant for EGC, in both *IPO* specification and *Shares* specification. Suggesting that this dimension of risk, in specific, is highly correlated with reward.

However, results for EGC are mostly contrary, depending on whether we consider the *IPO* specification (positive impact on reward and risk embedded in the innovation portfolio) or the *Shares* specification (negative impact on reward and risk embedded in the innovation portfolio). We argue the alternative specifications capture a different set of empirical predictions (drivers). If, as we advanced, the percentage of shares captures only the drivers "dividend policy", "diversification" and "on-the-job utility", these results suggest that the positive driver "diversification" is dominated by the negative drivers "dividend policy" and/or "on-the-job utility". If we further assume that the average impact of going public that is jointly attributable to these 3 drivers is also negative, it follows that the drivers "access to equity" dominates the impact of "tolerance-for-failure and short-term pressure".

Yet, think that the advanced predictions are not exhaustive though. There may well exist alternative drivers in play. For example, going public may also be associated with inventors' mobility [Wu (2012); and, Bernstein (2015)], impacting companies' capabilities. This can be a consequence of the drivers just exposed (a shift in inventors' job-market demand), but it also can be a parallel direct consequence of IPO event, if inventors' have a relative preference for a particular stage of companies' life-cycle (a shift in inventors' job-market supply). We think that further research is needed to investigate the impacts on inventors' job-market.

We are aware that results are limited, in the sense that we are not able to provide a full explanation for the evidence on the risk-return of the innovation portfolio for ECG, of both: positive average impact of IPO; and, negative of shares offered at the IPO. However, we think that this reflects the complexity of the IPO event, and of its' impact on innovation.

There are several limitations worth pointing out. On the one hand, we are concerned with measurement-related issues because we do not observe possible changes in patenting strategy, and we do not observe patents' visibility changes (which could impact the *Average Cites* through *Cites*). We think this is more concerning in our *IPO* specification than it is on our *Shares* specification because we find it likely that the percentage of shares offered at the IPO is unrelated with visibility – potentially being more effective on isolating, as a whole, the arguments presented in the theoretical framework section. I any case, we don't think that the measuring issue is driving the difference in results found between *IPO* and *Shares* specifications. On the other hand, one could question the exogeneity of *IPO* and *Shares* variables. Finally, we rely on the assumption that the decision to withdraw/complete is unrelated to companies' innovation policies or opportunities. This is questionable, because some factors that justify the decision to withdraw may also impact the outcome variables. Numerous factors justify the decision to withdraw, all of which occurring in the book-building phase: *"the IPO process itself (for example, the inability to comply with SEC disclosure requirements or resolve SEC staff comments, a poor road show, or resignation of the enterprise's underwriters or auditors) […] market conditions (for example, reduced market liquidity or demand for IPOs, changes in interest rates and costs of capital, or changes in market sector valuations) […] adverse business developments (for example, loss of a customer or prospective customer, loss of key personnel, or inability to obtain financing) […] unexpected change in the outlook or*

profile of the industry […] in other cases, an IPO might be withdrawn because a financial or strategic buyer acquires the enterprise." [AICPA (2016)]. Hence, selection issues may arise, as for example, an increase in the cost of capital in the book-building phase could explain both the IPO withdraw and a lower innovation outcome.

Future research could also investigate the role of corporate governance – in line with Hill and Snell (1988), Baysinger *et al.* (1991) and Francis and Smith (1995) – and managers' characteristics – in line with Barker and Mueller (2002) and Kor (2006) in this particular setting. A possible extension would be to assess the impact of corporate governance effectiveness, around the IPO event. We would expect that more effective corporate governance would favour owners' characteristics (as opposed to managers') – which, in the case of Shares specification, and according to our predictions, would positively foster the risk-reward embedded in the innovation portfolio.

REFERENCES

Aggarwal, V.; and, Hsu, D. 2014. Entrepreneurial Exits and Innovation. Management Science 60(4): 867-887.

American Institute of Certified Public Accountants (AICPA). 2016. Accounting and Valuation Guide: valuation of privately-held-company equity securities issued as compensation. Wiley. ISBN: 978-1-937352-22-6.

Barker, V.; and, Mueller, G. 2002. CEO Characteristics and Firm R&D Spending. Management Science. 48(6): 782-801.

Baysinger, B.; Kosnik, R.; and, Turk, T. 1991. Effects of Board and Ownership Structure on Corporate R&D Strategy. Academy of Management Journal. 34(1): 205-214.

Bernstein, S. 2015. Does Going Public Affect Innovation? The Journal of Finance. 70(4): 1365-1403.

Black, B.; and, Gilson, R. 1998. Venture Capital and the Structure of Capital Markets: banks versus stock markets. Journal of Financial Economics. 47: 243-277.

Brown, J.; Fazzari, S.; and, Petersen, B. 2009. Financing Innovation and Growth: cash flow, external equity, and the 1990s R&D boom. The Journal of Finance. 64(1): 151-185.

Brown, J.; Martinsson, G.; and, Petersen, B. 2013. Law, Stock Markets, and Innovation. The Journal of Finance. 68(4): 1517-1549.

Carpenter, R.; and, Petersen, B. 2002. Capital Market Imperfections, High-Tech Investment, and New Equity Financing. The Economic Journal. 112(477): 54-72.

Certo, S.; Covin, J.; Daily, C.; and, Dalton, D. 2001. Wealth and the Effects of Founder Management Among IPO-stage New Ventures. Strategic Management Journal. 22(6-7): 641-658.

Certo, S.; Holcomb, T.; and, Holmes, R. 2009. IPO Research in Management and Entrepreneurship: Moving the Agenda Forward. Journal of Management. 35(6): 1340-1378.

Draho, J. 2004. The IPO Decision: why and how companies go public. Edward Elgar Publishing.

Fama, E. 1980. Agency Problems and the Theory of the Firm. Journal of Political Economy. 88(2): 288–307.

Ferreira, D.; Manso, G.; and, Silva, A. 2014. Incentives to Innovate and the Decision to Go Public or Private. The Review of Financial Studies. 27(1): 256-300.

Francis, J.; and, Smith, A. 1995. Agency Costs and Innovation: some empirical evidence. Journal of Accounting and Economics. 19: 383-409.

Hambrick, D.; and, Manson, P. 1984. Upper Echelons: the organization as a reflection of its top managers. The Academy of Management Review. 9(2): 193-206.

Hall, B. ; Jaffe, A.; and, Trajtenberg, M. 2001. The NBER patent citation data file: lessons, insights and methodological tools. NBER Working Paper 8498.

Hall, B. 2002. The Financing of Research and Development. Oxford Review of Economic Policy. 18(1): 35-51.

Hall, B. 2010. The Financing of Innovative Firms. Review of Economics and Institutions. 1(1): 1-30.

Hall, B.; and, Lerner J. 2010. The Financing of R&D and Innovation. In Handbook of the Economics of Innovation, ed. Bronwyn H. Hall and Nathan Rosenberg. 609-639. Elsevier.

Hill, C.; and, Snell, S. 1988. External Control, Corporate Strategy, and Firm Performance in Research-Intensive Industries. Strategic Management Journal. 9(6): 577-590.

Jensen, M. 1989 (rev. 1997). Eclipse of the Public Corporation. Harvard Business Review.

Jensen, M.; and, Meckling, W. 1976. Theory of the Firm: managerial behavior, agency costs and ownership structure. Journal of Financial Economics. 3(4): 305-360.

Kenney, M.; and, Patton, D. 2013. Firm Database of Emerging Growth Initial Public Offerings (IPOs) from 1990 through 2010. Inter-university Consortium for Political and Social Research.

Kor, Y. 2006. Direct and Interaction Effects of Top Management Team and Board Compositions on R&D Investment Strategy. Strategic Management Journal. 27(11): 1081-1099.

March, J. 1991. Exploration and Exploitation in Organizational Learning. Organization Science. 2(1): 71–87.

Michaely, R.; and, Roberts, M. 2012. Corporate Dividend Policies: lessons from private firms. The Review of Financial Studies. 25(3): 711-746.

Pagano, M.; Panetta, F.; and, Zingales, L. 1998. Why do Companies Go Public? An empirical analysis. The Journal of Finance. 53(1): 27–64.

Rajan, R. 1992. Insiders and Outsiders: the choice between informed and arm's-length debt. The Journal of Finance. 47(4): 1367-1400.

Ritter, J. 1987. The Costs of Going Public. Journal of Financial Economics. 19: 269-281.

Wu, G. 2012. The Effect of Going Public on Innovative Productivity and Exploratory Search. Organization Science. 23 (4): 928–950.

Zingales, L. 1995. Insider Ownership and the Decision To Go Public. Review of Economic Studies. 62(3): 425-448.

APPENDIX A – main figures and tables

Table 1 – Synthesis of predictions

The table synthesises the empirical predictions on the overall reward of innovation and/or on the risk-reward of the innovation portfolio.

Scope of impact		Driver	Main references	Predicted effect on overall reward of innovation and/or on the risk-reward of the innovation portfolio
Changes in financial constraints		Access to equity	Brown, Martinsson and Petersen (2013)	+
		Dividend policy	Michaely and Roberts (2012) | Brown, Fazzari and Petersen (2009)	-
Changes in investment preferences	Owners' preferences	Diversification	Pagano, Panetta and Zingales (1998)	+
		Tolerance-for-failure and short-term pressure	Ferreira, Manso and Silva (2014)	-
	Managers' preferences	On-the-job utility	Jensen and Meckling (1976)	-

Table 2 – IPO and patent data

The table presents the sample distribution of IPO, patents and citations over time.

Year	IPO		Patents		Citations by Grant Year of Cited		Citations by Grant Year of Citing	
	Completed	Withdrew	Completed	Withdrew	Completed	Withdrew	Completed	Withdrew
1981			2	4	0	12	0	0
1982			124	4	280	62	2	3
1983			285	21	782	71	36	4
1984			367	33	1094	92	120	12
1985	48	3	352	31	1081	122	266	32
1986	115	15	313	23	1389	78	428	55
1987	97	14	518	42	2801	121	685	71
1988	38	7	644	51	3240	276	990	78
1989	43	2	860	95	4842	496	1677	124
1990	45	13	936	98	6160	541	1969	158
1991	106	2	1078	82	7342	471	2559	222
1992	116	30	1280	102	10409	724	3404	306
1993	169	17	1419	86	13335	485	4844	356
1994			1319	78	14731	508	6099	483
1995			1081	45	14257	259	7597	422
1996			1173	57	15998	304	9391	447
1997			1171	96	14782	598	10192	553
1998			1548	99	17204	541	15239	557
1999			1059	63	11539	288	15088	431
2000			579	16	6546	86	16128	441
2001							15204	343
2002							12585	350
2003							10581	407
2004							7424	179
2005							3733	86
2006							1571	15
Total	777	103	16108	1126	147812	6135	147812	6135

Table 3 – Descriptive statistics

The table presents the descriptive statistics of the innovation variables for all 880 companies in the sample. Each company has 8 observations (*k=0, 1, 2, 3, 4, 5, 6* and *7*). For (1) and (2), there are no missing values. For (3) to (7) there are missing values because some of the portfolios have zero eligible patents. (7) is further restricted to observations with at least 1 eligible citation.

Innovation Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
(1) – Patents	7040	7.692898	28.44974	0	708
(2) – Cites	7040	67.24361	295.759	0	10812
(3) – Average Cites	5129	8.274504	9.196147	0	133
(4) – Dispersion Cites	5129	4.127362	5.973097	0	70.74876
(5) – Mean Lag Forward Cites	5129	4.310513	0.8595701	1	7
(6) – Dispersion Lag Forward Cites	5129	0.6637803	0.6123937	0	3
(7) – Old	4992	-1.656017	1.115773	-4	0

Table 4 – Pre-IPO differences between completed and withdrew companies

The table compares the means of each of the pre-IPO values of innovation variables (Y_{before}) between companies that completed an IPO and companies that withdrew an IPO. The innovation variables are: (1) *Patents*; (2) *Cites*; (3) *Average Cites*; (4) *Dispersion Cites*; (5) *Mean Lag Forward Cites*; (6) *Dispersion Lag Forward Cites*; and, (7) *Old*. It is an independent sample t-test assuming unequal variances. We restrict the analysis to those companies that have no missing values on both Y_{before} and Y_{after} – the average value of the innovation variables when *k=0,1,2 and 3* and *k=4,5,6 and 7*, respectively. Regarding Y_{before} and Y_{after} constructs: in (3) and (5) the average is conditional on those observations that have more than zero eligible patents; for the dispersion measures (4) and (6) we decided to restrict the average conditional on those observations that have more than one eligible patent (to avoid cases in which standard deviation is hard to interpret); in (7) we confined the analysis to those observations that have at least one eligible citation. The number of companies is reported in parenthesis. *P<10%*; **P<5%; ***P<1% (two-tailed p-values).

Innovation Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Withdrew	4.01699 (103)	21.06553 (103)	6.738466 (76)	3.406136 (48)	4.334577 (76)	1.004438 (48)	-1.267203 (73)
Completed	5.759331 (777)	42.75676 (777)	8.26522 (586)	5.070864 (435)	4.277963 (586)	0.8537027 (435)	-1.104182 (575)
Difference	-1.74234	-21.69122***	-1.526754	-1.664729***	0.0566142	0.150735*	-0.1630213

Table 5 – Pre-IPO differences between groups – imputation

The table compares the means of each of the pre-IPO innovation variables (Y_{before}) between selected groups. The innovation variables are: (1) *Patents*; (2) *Cites*; (3) *Average Cites*; (4) *Dispersion Cites*; (5) *Mean Lag Forward Cites*; (6) *Dispersion Lag Forward Cites*; and, (7) *Old*. It is an independent sample t-test assuming unequal variances. We restrict the analysis to those companies that have no missing values on both Y_{before} and Y_{after} – the average value of the innovation variables when *k=0,1,2 and 3* and *k=4,5,6 and 7*, respectively. Regarding Y_{before} and Y_{after} constructs: in (3) and (5) the average is conditional on those observations that have more than zero eligible patents; for the dispersion measures (4) and (6) we decided to restrict the average conditional on those observations that have more than one eligible patent (to avoid cases in which standard deviation is hard to interpret); in (7) we confined the analysis to those observations that have at least one eligible citation. The number of companies is reported in parenthesis. *P<10%; **P<5%; ***P<1% (two-tailed p-values).

Panel A – emerging growth companies and non-emerging growth companies – Kenney and Patton (2013) criteria – that completed an IPO between (and including) 1990 and 1993.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Non-emerging growth	6.580189 (159)	47.16352 (159)	7.322459 (119)	4.713641 (82)	4.282581 (119)	0.9828356 (82)	-1.2773539 (118)
Emerging growth	5.011733 (277)	61.27256 (277)	11.03087 (216)	6.755832 (171)	4.296502 (216)	0.6869671 (171)	-1.113085 (214)
Difference	1.568456	-14.10904	-3.708406***	-2.042191***	-0.0139211	0.2958685***	0.160444*

Panel B – emerging growth companies and non-emerging growth companies – imputation – that applied for an IPO between (and including) 1985 and 1993.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Non-emerging growth	8.75 (315)	50.95714 (315)	6.017356 (262)	3.809364 (218)	4.324198 (262)	1.057081 (218)	-1.336408 (259)
Emerging growth	5.735497 (362)	52.48204 (362)	10.24295 (308)	5.829088 (264)	4.260335 (308)	0.7163934 (264)	-1.065732 (306)
Difference	3.014503	-1.524901	-4.225594***	-2.019724***	0.0638631	0.3406975***	-0.2706761***

Panel C – companies considered and not considered by the imputation procedure.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Not considered	0.2770936 (203)	1.683498 (203)	6.78442 (92)	0 (1)	4.252072 (92)	0 (1)	-0.6646586 (83)
Considered	7.138109 (677)	51.77253 (677)	8.300659 (570)	4.915603 (482)	4.28969 (570)	0.8704848 (482)	-1.189812 (565)
Difference	-6.861016***	-50.08903***	-1.516239	-	-0.0376176	-	0.5251534***

Panel D – companies that completed and companies that withdrew between (and including) 1985 and 1993 and considered emerging growth by the imputation.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Withdrew	6.409091 (22)	32.17045 (22)	6.116851 (22)	3.477117 (21)	4.337194 (22)	0.9633122 (21)	-1.332505 (22)
Completed	5.691912 (340)	53.79632 (340)	10.56034 (286)	6.032345 (243)	4.254423 (286)	0.6950547 (243)	-1.045067 (284)
Difference	0.7171791	-21.62587	-4.443491*	-2.555228*	0.082771	0.2682575**	-0.2874382*

Panel E – companies that completed and companies that withdrew between (and including) 1985 and 1993 and considered non-emerging growth by the imputation.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Withdrew	8.844828 (29)	46.11207 (29)	5.415533 (27)	3.350928 (27)	4.20365 (27)	1.036424 (27)	-1.723114 (27)
Completed	8.740385 (286)	51.44843 (286)	6.086501 (235)	3.874169 (191)	4.338049 (235)	1.060012 (191)	-1.291404 (232)
Difference	0.104443	-5.336358	-0.670968	-0.5232405	-0.1343985	-0.235883	-0.4317097***

Figure 1 – Percentage of shares offered by EGC IPO

The table presents the histogram for the variable *Shares* of emerging growth companies. *Shares* is the number of shares offered divided by the number of shares outstanding at the IPO date – hence, the percentage of shares that go public at the IPO date. IPOs between (and including) 1990 and 1993.

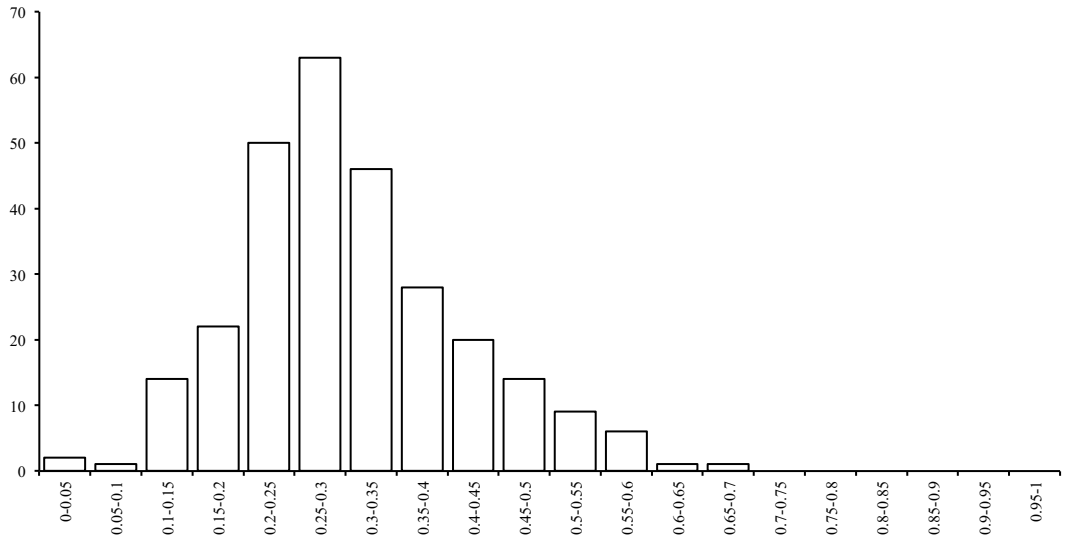


Table 6 – Within-company changes around IPO

The table presents within-company changes in innovation portfolio around the IPO event. The innovation variables are: (1) *Patents*; (2) *Cites*; (3) *Average Cites*; (4) *Dispersion Cites*; (5) *Mean Lag Forward Cites*; (6) *Dispersion Lag Forward Cites*; and, (7) *Old*. *Year* are dummy variables indicating the distance in years relative to the IPO year ($year_0$ is a dummy variable indicating the IPO year and is omitted). For the dispersion measures – (4) and (6) – we decided to restrict the analysis to those observations for which eligible patents are higher than 1. This is because it is hard to interpret the standard deviation for portfolios that have zero or 1 single patent. For (7), we confined the analysis to those observations that have at least one eligible citation. All models are estimated with individual effects and time effects. Robust standard errors, clustered at the company level, are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel A – companies that completed the IPO.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Year_1$	0.9419218***	11.36039***	0.30432	0.1869706	-0.0356332	0.0028217	-0.0720904
	(0.2110434)	(3.036001)	(0.2436527)	(0.1979076)	(0.0302327)	(0.0272987)	(0.0554371)
$Year_2$	1.839024***	20.30196***	0.2243728	0.4891198*	-0.0177603	0.0395291	-0.2677822***
	(0.3877727)	(5.623021)	(0.2899147)	(0.2561703)	(0.0447419)	(0.0379448)	(0.0772906)
$Year_3$	2.751789***	28.41094***	0.2823738	0.5270903*	0.0037788	0.0459603	-0.5182458***
	(0.5413942)	(7.706803)	(0.3784376)	(0.3118352)	(0.0571334)	(0.0478068)	(0.0957407)
$Year_4$	3.645189***	37.4708***	0.3886569	0.6697476**	-0.0048901	0.0522687	-0.8909923***
	(0.6643438)	(9.129791)	(0.414465)	(0.3341235)	(0.0696846)	(0.0582759)	(0.1086564)
$Year_5$	4.646364***	51.256***	0.3384813	0.9258513**	-0.0134343	0.0556057	-1.229551***
	(0.7400288)	(9.628333)	(0.4472251)	(0.3630331)	(0.082007)	(0.0670863)	(0.1189182)
$Year_6$	6.09482***	67.93346***	0.2563481	0.9714929**	-0.0001672	0.1033767	-1.523302***
	(0.8126755)	(10.11565)	(0.4719717)	(0.3925792)	(0.0927741)	(0.0769662)	(0.1325815)
$Year_7$	7.642803***	88.03155***	0.2377998	1.279353***	-0.0114008	0.1373309	-1.823641***
	(0.8871147)	(11.7862)	(0.4639487)	(0.3830887)	(0.1022097)	(0.0862495)	(0.1455712)
Overall R^2	0.0090	0.0181	0.0160	0.0343	0.0016	0.0045	0.0546
Prob>F	0.0000	0.0000	0.0250	0.0000	0.6688	0.0070	0.0000
Observations	6216	6216	4551	3367	4551	3367	4443
Clusters	777	777	777	612	777	612	763

Panel B – companies that withdrew the IPO.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Year_1$	0.6936815*	3.600747**	-0.1679462	-0.2433977	-0.1083188	0.1111133	0.0688153
	(0.3570655)	(1.542921)	(0.3842141)	(.3706215)	(0.1104805)	(0.086806)	(.2505061)
$Year_2$	1.188553**	4.247701	-0.8340949	-0.5095872	-0.039621	0.1260473	-0.2320224
	(0.5535449)	(2.556391)	(0.5541454)	(.4056793)	(0.1508337)	(0.0881683)	(.3147209)
$Year_3$	1.696358**	7.307061*	-0.630386	-0.665594	-0.0044839	0.0256581	-0.3035067
	(0.692577)	(3.795701)	(0.6351442)	(.5975086)	(0.1920293)	(0.1258628)	(.3956928)
$Year_4$	1.916244**	8.837203	-0.5249066	-1.105454*	0.1469317	0.0369521	-0.7588851*
	(0.9574119)	(5.660899)	(0.7098703)	(.6269939)	(0.1969379)	(0.1436581)	(.4516066)
$Year_5$	1.619121	7.658726	-0.1938081	-0.7189939	0.2179866	0.1134348	-0.9281189*
	(1.139365)	(6.829197)	(0.7422564)	(.7436413)	(0.1948683)	(0.1407505)	(.494045)
$Year_6$	1.154896	7.171304	0.0414306	-0.3819971	0.3131802*	0.1162212	-1.034922*
	(1.18015)	(6.874059)	(0.7808569)	(.7900586)	(0.1872403)	(0.1552514)	(.558678)
$Year_7$	1.53589	11.08175	-0.151978	-0.136813	0.4047379**	0.1096068	-1.242043**
	(1.182978)	(6.856124)	(0.7331384)	(.7875467)	(0.2029859)	(0.1760236)	(.6076176)
Overall R^2	0.0075	0.0145	0.0047	0.0530	0.0225	0.0089	0.0407
Prob>F	0.0320	0.0821	0.6814	0.4387	0.3123	0.0025	0.0000
Observations	824	824	578	364	578	364	549
Clusters	103	103	103	69	103	69	97

Table 7 – The effect of going public

The table presents the effect of completing an IPO on each innovation variable. The innovation variables are: (1) *Patents*; (2) *Cites*; (3) *Average Cites*; (4) *Dispersion Cites*; (5) *Mean Lag Forward Cites*; (6) *Dispersion Lag Forward Cites*; and, (7) *Old*. *IPO* is a dummy variable that equals to 1 when the company completed the IPO and 0 otherwise. We restrict the analysis to those companies that have no missing values on both Y_{before} and Y_{after} – the average value of the innovation variables when *k=0,1,2 and 3* and *k=4,5,6 and 7*, respectively. Regarding Y_{before} and Y_{after} constructs: in (3) and (5) the average is conditional on those observations that have more than zero eligible patents; for the dispersion measures (4) and (6) we decided to restrict the average conditional on those observations that have more than one eligible patent (to avoid cases in which standard deviation is hard to interpret); in (7) we confined the analysis to those observations that have at least one eligible citation. We included industry effects and time effects. Industry effects are 1 digit SIC dummies and time effects are IPO year dummies. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.
IPOs between (and including) 1985 and 1993.

Innovation Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Y_{before}	1.184741*** (0.1971084)	2.137991*** (0.2589002)	0.6769843*** (0.0662234)	0.763925*** (0.0591634)	0.543162*** (0.0460891)	0.5839384*** (0.0358264)	-0.0299006 (0.0431537)
IPO	3.134382*** (1.175529)	26.50044** (10.76371)	0.5354535 (0.5620708)	0.6894623 (0.4198007)	-0.1445835* (0.0782126)	0.0514913 (0.0616211)	0.1903776* (0.1092102)
Constant	-2.290808 (1.998253)	-6.051064 (22.14764)	0.6739 (0.8146907)	0.9046653 (1.361754)	2.097642*** (0.2753247)	0.3103412** (0.1492047)	-2.240268*** (0.3999673)
R^2	0.7652	0.8068	0.6224	0.6399	0.4228	0.4620	0.4228
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0046
Observations	880	880	662	483	662	483	648

Table 8 – The effect of going public – EGC vs. NEGC

The table presents the effect of completing an IPO on each innovation variable. The innovation variables are: (1) *Patents*; (2) *Cites*; (3) *Average Cites*; (4) *Dispersion Cites*; (5) *Mean Lag Forward Cites*; (6) *Dispersion Lag Forward Cites*; and, (7) *Old*. *IPO* is a dummy variable that equals to 1 when the company completed the IPO and 0 otherwise. We restrict the analysis to those companies that have no missing values on both Y_{before} and Y_{after} – the average value of the innovation variables when *k=0,1,2 and 3* and *k=4,5,6 and 7*, respectively. Regarding Y_{before} and Y_{after} constructs: in (3) and (5) the average is conditional on those observations that have more than zero eligible patents; for the dispersion measures (4) and (6) we decided to restrict the average conditional on those observations that have more than one eligible patent (to avoid cases in which standard deviation is hard to interpret); in (7) we confined the analysis to those observations that have at least one eligible citation. We included industry effects and time effects. Industry effects are 1 digit SIC dummies and time effects are IPO year dummies. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.
IPOs between (and including) 1985 and 1993.

Panel A – all firms.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Y_{before}	1.182224*** (0.1997103)	2.14235*** (0.2607179)	0.6349729*** (0.0682928)	0.7638656*** (0.0591865)	0.4561661*** (0.0517028)	0.5796115*** (0.0358647)	-0.0139195 (0.0435967)
IPO	4.331687* (2.34937)	44.46702** (18.90238)	0.9051967 (0.5984079)	0.690825 (0.4199111)	-0.026022 (0.0912823)	0.0529808 (0.0617902)	0.199428 (0.1266881)
Constant	7.65843 (10.85141)	-10.26343 (64.00168)	0.6790802 (0.9826475)	1.487397 (1.155935)	2.217275*** (0.2692032)	0.2112065* (0.1137845)	-2.25159*** (0.2599032)
R^2	0.7624	0.8077	0.6122	0.6391	0.3018	0.4611	0.0541
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0634
Observations	677	677	570	482	570	482	565

Panel B – emerging growth companies.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Y_{before}	1.282103*** (0.2660878)	1.972412*** (0.4354881)	0.5886456*** (0.0749539)	0.740224*** (0.0658016)	0.5324226*** (0.0622801)	0.6297698*** (0.0541939)	0.0713118 (0.0645914)
IPO	3.467196 (2.355683)	35.97809 (23.0467)	2.657417*** (0.8683433)	1.815739*** (0.5606962)	-0.069227 (0.1309316)	0.0203394 (0.0810901)	0.2901295 (0.1825057)
Constant	-6.705789 (4.768386)	-95.32949 (46.66141)	-1.089324 (2.591242)	-1.931741 (1.338119)	2.510019*** (0.5378878)	0.1369963 (0.1448667)	-2.116738*** (0.3391738)
R^2	0.7722	0.6903	0.5851	0.6659	0.4164	0.5076	0.0491
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.7883
Observations	362	362	308	264	308	264	306

Panel C – non-emerging growth companies.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Y_{before}	1.135197*** (0.2694649)	2.254367*** (0.2550358)	0.7597101*** (0.0595037)	0.8378859*** (0.1237133)	0.4033025*** (0.0814413)	0.5463501*** (0.0548714)	-0.1236435** (0.0592363)
IPO	5.066868 (3.688949)	60.73981** (29.8315)	-0.2788624 (0.0595037)	-0.151073 (0.6484997)	0.0321174 (0.1312559)	0.0935623 (0.0946229)	0.1338358 (0.1878276)
Constant	-12.32688* (6.32868)	-310.6078* (186.612)	21.67437*** (2.645204)	0.2999629 (0.9505093)	2.473077*** (0.3913057)	0.0269411 (0.1236539)	-2.195544*** (0.4066457)
R^2	0.7719	0.8996	0.7258	0.5451	0.2354	0.2354	0.1213
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	315	315	262	218	262	218	259

Table 9 – The effect of percentage of shares offered by EGC IPO

The table presents the effect of shares offered on each innovation variable. The innovation variables are: (1) *Patents*; (2) *Cites*; (3) *Average Cites*; (4) *Dispersion Cites*; (5) *Mean Lag Forward Cites*; (6) *Dispersion Lag Forward Cites*; and, (7) *Old*. *Shares* is the number of shares offered divided by the number of shares outstanding at the IPO date – hence, the percentage of shares that go public at the IPO date. We restrict the analysis to those companies that have no missing values on both Y_{before} and Y_{after} – the average value of the innovation variables when *k=0,1,2 and 3* and *k=4,5,6 and 7*, respectively. Regarding Y_{before} and Y_{after} constructs: in (3) and (5) the average is conditional on those observations that have more than zero eligible patents; for the dispersion measures (4) and (6) we decided to restrict the average conditional on those observations that have more than one eligible patent (to avoid cases in which standard deviation is hard to interpret); in (7) we confined the analysis to those observations that have at least one eligible citation. We included industry effects and time effects. Industry effects are 1 digit SIC dummies and time effects are IPO year dummies. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

IPOs between (and including) 1990 and 1993.

Innovation Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Y_{before}	2.030545*** (0.0826693)	2.30173*** (0.2897348)	0.6067885*** (0.0921178)	0.733647*** (0.0755794)	0.4828294*** (0.0802728)	0.6075597*** (0.0777935)	0.0111856 (0.0735631)
Shares	-14.65472** (6.399349)	-122.9997 (112.2886)	-8.787497** (3.439304)	-8.600904*** (2.7023)	0.2206267 (0.3416232)	0.1085451 (0.2331961)	-1.477064*** (0.4741043)
Constant	4.955831 (4.812494)	121.0066 (92.00639)	10.61993*** (3.397128)	3.559533*** (1.233806)	2.060714*** (0.3793858)	0.2015629 (0.1427283)	-1.271942** (0.4890888)
R^2	0.8962	0.8655	0.5887	0.6987	0.3649	0.4689	0.0597
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2481
Observations	277	277	216	171	216	171	214

APPENDIX B – secondary figures and tables

Table 10 – Within-company changes – robustness check

The Table is similar to Table 6, except that the analysis is confined to those patents that have at least one eligible citation – disregarding 1801 patents with no eligible citation and, consequently, 20 companies (14 completed and 6 withdrew). Provided that the analysis is confined to this subset of companies and patents, we change the definition of (5) as it turns out to be possible to compute a weighted average. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.
IPOs between (and including) 1985 and 1993.

Panel A – companies that completed the IPO.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Year_1$	0.938552*** (0.1981605)	11.49696*** (3.095407)	0.2784644 (0.2492361)	0.2250239 (0.2065297)	-0.00138 (0.0234833)	.0177648 (.0160319)	-0.0720904 (0.0554371)
$Year_2$	1.804434*** (0.374871)	20.60858*** (5.733569)	0.1504631 (0.2999371)	0.546148** (0.2646895)	0.0131249 (0.0364792)	.0237472 (.0217804)	-0.2677822*** (0.0772906)
$Year_3$	2.699948*** (0.5279115)	28.82107*** (7.859052)	0.1180485 (0.395606)	0.6442583** (0.3186645)	0.0349589 (0.0488135)	.0358275 (.0264707)	-0.5182458*** (0.0957407)
$Year_4$	3.580379*** (0.6508512)	38.09371*** (9.301747)	0.1984181 (0.4378275)	0.825828** (0.3340766)	0.039132 (0.0627126)	.0492754 (.0325241)	-0.8909923*** (0.1086564)
$Year_5$	4.563875*** (0.7253003)	52.14761*** (9.806216)	0.1539459 (0.4776766)	1.127021*** (0.352666)	0.041874 (0.0749968)	.0501331 (.0388674)	-1.229551*** (0.1189182)
$Year_6$	5.924525*** (0.8008607)	69.16544*** (10.2834)	0.1270953 (0.5126924)	1.198043*** (0.372529)	0.0410303 (0.0855944)	.0712831 (.0446968)	-1.523302*** (0.1325815)
$Year_7$	7.376082*** (0.8682029)	89.61902*** (11.96419)	0.1370789 (0.5150055)	1.537401*** (0.3494555)	0.02237 (0.0954303)	.0791138 (.0499461)	-1.823641*** (0.1455712)
Overall R^2	0.0112	0.0183	0.0177	0.0261	0.0221	0.0049	0.0546
Prob>F	0.0000	0.0000	0.0381	0.0000	0.0034	0.0147	0.0000
Observations	6104	6104	4443	3227	4443	4294	4443
Clusters (IPO firms)	763	763	763	595	763	742	763

Panel B – companies that withdrew the IPO.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Year_1$	0.7180217** (0.3103503)	3.733216** (1.643228)	-0.1461439 (0.3912316)	-0.3226089 (.4406156)	-0.0651446 (0.0695401)	0.0769239* (0.0433973)	0.0688153 (.2505061)
$Year_2$	1.212892** (0.485751)	4.246834 (2.680655)	-0.7592538 (0.5607449)	-0.6053809 (0.4486819)	-0.036819 (0.112198)	0.0881527 (0.0584824)	-0.2320224 (.3147209)
$Year_3$	1.657313** (0.6550879)	7.510677* (3.943507)	-0.6290596 (0.6382699)	-0.7510603 (0.6472263)	-0.0064148 (0.1660578)	0.020583 (0.0901879)	-0.3035067 (.3956928)
$Year_4$	1.78253* (0.8981263)	9.339202 (5.874014)	-0.3382725 (0.7184288)	-1.244977* (0.6600977)	0.0086387 (0.193341)	0.040063 (0.0972581)	-0.7588851* (.4516066)
$Year_5$	1.493395 (1.07767)	8.25676 (7.117061)	-0.0887942 (0.727355)	-0.742032 (0.7812898)	0.0776922 (0.2024902)	0.0236101 (0.1028969)	-0.9281189* (.494045)
$Year_6$	1.064755 (1.131306)	7.904082 (7.246043)	0.0993446 (0.7354271)	-0.342682 (0.8377789)	0.2038126 (0.1999465)	0.0370363 (0.1151528)	-1.034922* (.558678)
$Year_7$	1.290064 (1.119697)	12.136* (7.20832)	0.0611765 (0.7006742)	-0.0270154 (0.8327477)	0.2387452 (0.2163389)	0.0168088 (0.1243079)	-1.242043** (.6076176)
Overall R^2	0.0073	0.0141	0.0086	0.0544	0.0348	0.0135	0.0407
Prob>F	0.0270	0.0786	0.3794	0.1790	0.0815	0.2171	0.0000
Observations	776	776	549	324	549	515	549
Clusters (withdrew firms)	97	97	97	62	97	91	97

Table 11 – The effect of going public – robustness check

The Table is similar to Table 7, except for the definition Y_{before} and Y_{after}. In this case, Y_{before} and Y_{after} are defined as the average value of the innovation variables when *k=0,1 and 2* and *k=3,4,5,6 and 7*, respectively. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.
IPOs between (and including) 1985 and 1993.

Innovation Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Y_{before}	1.112791*** (0.1810505)	2.131289*** (0.3050299)	0.607898*** (0.0660658)	0.7450908*** (0.0699744)	0.4674154*** (0.0469813)	0.5486075*** (0.0405839)	-0.0558851*** (0.0346566)
IPO	3.153935*** (1.11493)	28.71303*** (10.44681)	0.9272478* (0.5125851)	0.6692544 (0.4397908)	-0.0609697 (0.0791938)	0.074994 (0.065633)	0.1651411* (0.0914544)
Constant	-0,1464931 (2.058992)	10.71256 (21.22547)	3.181605*** (1.114661)	1.350579 (1.153915)	2.490249*** (0.3525174)	0.4627669 (0.3669862)	-2.247325*** (0.4119121)
R^2	0.536	0.7786	0.6053	0.6208	0.3768	0.4503	0.0680
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0018
Observations	880	880	598	424	598	424	584

Table 12 – The effect of going public – EGC vs. NEGC – Difference-in-Differences

The table is analogous to panels B and C of Table 8, except that it is a Difference-in-Differences approach. We included industry effects and time effects. Robust standard errors, clustered at the company level, are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel B – emerging growth companies.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After	5.159091**	37.63636**	-0.5465446	0.1135149	0.0818059	0.0426662	-0.9185876***
	(2.143259)	(15.8818)	(0.5378977)	(0.4216692)	(0.1279866)	(0.0902691)	(0.2562287)
IPO	3.874404	46.37486**	3.777747***	2.274692***	-0.1037456	-0.2772129*	0.2516864
	(3.775849)	(22.16538)	(1.089525)	(0.7872055)	(0.1204907)	(0.1501919)	(0.174158)
IPO.After	**2.172527**	**52.64084****	**0.3927083**	**0.8907788***	**-0.0726405**	**0.1153387**	**-0.0499488**
	(2.332186)	**(21.23351)**	**(0.7134487)**	**(0.5075366)**	**(0.1326725)**	**(0.0931454)**	**(0.2637913)**
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.0843	0.0621	0.1011	0.0850	0.1280	0.0809	0.3124
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	724	724	616	528	616	528	612

Panel C – non-emerging growth companies.

Innovation Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After	-2.017241	-13.96552	0.4332913	0.5019675	0.1033094	-0.0950448	-0.7236505***
	(3.001727)	(11.84075)	(0.7283845)	(0.5222906)	(0.1526871)	(0.0833559)	(0.2377939)
IPO	-0.9815811	-2.344077	-0.1327319	0.0826412	0.1510857	0.0439891	0.3971898***
	(3.673184)	(23.72668)	(0.8918031)	(0.6068839)	(0.159517)	(0.1062545)	(0.1406497)
IPO.After	**5.333675**	**61.78195*****	**0.0801267**	**0.111175**	**-0.0753986**	**0.0466631**	**-0.2472508**
	(3.233856)	**(21.62935)**	**(0.7607207)**	**(0.5597379)**	**(0.1611904)**	**(0.0904055)**	**(0.2503222)**
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.0278	0.0279	0.2816	0.2286	0.0473	0.1181	0.3041
Prob>F	0.0001	0.0030	0.0000	0.0000	0.4705	0.0000	0.0000
Observations	630	630	524	436	524	436	518

Pharmaceuticals' R&D After Recombinant DNA: the decision between acquiring and allying with a biotechnological

Fernando Santos
Universidad Carlos III de Madrid
Department of Business Administration
fjasantos@gmail.com | fasantos@porto.ucp.pt

PhD BF | Working Paper
Supervisors: Andrea Fosfuri and Neus Palomeras

ABSTRACT

The discovery of R-DNA represented an R&D competence-destroying event for incumbent pharmaceutical companies. As a response to that, organizational deals – namely, strategic alliances and mergers and acquisitions – have been extensively used as external sources of knowledge. We examine a potential determinant of pharmaceuticals' decision between these two alternatives: the width of applicability of the internal knowledge of the biotechnological company. We find compelling evidence that it has a positive effect on the likelihood that the pharmaceutical company acquires (or merges with) that biotechnological company.

INTRODUCTION

Till the 1980s vertical integrated companies characterized the pharmaceutical industry. Yet, the discovery of recombinant DNA (R-DNA) implied a radical transformation of the knowledge base of the industry that, in turn, indubitably triggered a vertical disintegration and a fragmentation of the R&D layer – with many specialized biotechnological firms operating on drug discovery and fuelling the pipelines of incumbent pharmaceutical companies operating the downstream layers of the industry value chain[1]. In a way, drug delivery became much more dependent on market transactions (as compared to hierarchies). Yet, mergers and acquisitions (MA) and strategic alliances (SA) proliferated within the industry [van Beuzekom and Arundel (2006); Sytch and Bubenzer (2008); Shibayama *et al*. (2008)] – both between biotechnological companies (horizontal deals) and between these and pharmaceutical companies (vertical deals) – being widely used as means to get access to valuable biotechnological knowledge, technologies and innovative capabilities [Schweizer (2005)][2]. Incumbent pharmaceutical companies, in particular, frequently use SA and MA with biotechnological companies as alternative external sources of technological knowledge [Powell *et al*. (1996); Venhaverbeke *et al*. (2002); Hagedoorn and Duysters (2002); Hill and Rothaermel (2003); Rothaermel and Deeds (2004); Schweizer (2005); Higgins and Rodriguez (2006); Rothaermel and Hess (2007); and, Grigoriou and Rothaermel (2017)]. But, what determines their choice between these alternatives?

This context has received a lot of attention from scholars, leading to a rich and rather dispersed literature. A few take the perspective of biotechnological and have suggested that biotechnological companies have some discretion when choosing pharmaceutical allies [Katila *et al.* (2008); and, Diestre and Rajagopalan (2012)][3]. Others, as we, take the perspective of the pharmaceutical company and investigate

[1] The value chain of the pharmaceutical industry encompasses the following ordered steps (which are needed to deliver a drug to the end-user): R&D; manufacturing; supply and logistics; and, marketing and sales. The R&D step encompasses the following ordered sub-steps: drug discovery (gene or genome sequencing; target discovery; target validation; and, lead discovery); pre-clinical tests (animal studies); and, clinical tests (phase I; phase II; and, phase III).

[2] While we recognize the existence of vertical deals where the object is related with downstream activities (like sales and marketing), arguably motivated by biotechnological companies' lack of capabilities on these areas, we intentionally avoid this discussion because it is out of the scope of this paper.

[3] Katila *et al.* (2008) suggested that biotechnological companies face a critical tension when selecting an incumbent pharmaceutical company as an alliance partner, characterized by a trade-off between value creation and value appropriation. Building upon this, Diestre and Rajagopalan (2012) find evidence suggesting that new biotechnological companies are less likely to select for R&D allies incumbent pharmaceutical companies that have the ability to absorb and exploit their knowledge when they perceive that the pharmaceutical company will also have incentives to misappropriate the biotechnological company's knowledge.

how they respond to this competence-destroying event. Results suggest that solutions involve: attracting human capital [Zucker and Darby (1997)]; establishing strategic alliances with biotechnological companies [Rothaermel (2001)]; acquiring biotechnological companies [Schweizer (2005); and, Higgins and Rodriguez (2006)]; or, a combination of these [Arora and Gambardella (1990); Rothaermel and Hess (2007); Hess and Rothaermel (2011); and, Grigoriou and Rothaermel (2017)][4]. While enriching, this particular literature stream does not address pharmaceuticals' choices between acquiring and allying with biotechnological companies. While SA and MA may be seen as alternative mechanisms to access external knowledge, they are, in fact, substantially different.

It is on a broader stream of literature that we find further insights. Previous work suggests that MA would be preferred when external sources of innovative capabilities relate with the core business, due to the dangers of uncontrolled technological transfer (relative more present in SA) [Hagedoorn and Duysters (2002)]. While we agree with this argument, in our view, the decision between MA and SA depends, not only on commitment requirements, but also on flexibility requirements [Ghemawat and Sol (1998); Smit and Trigeorgis (2004); and, Yin and Shanley (2008)]. In particular, we expect pharmaceutical companies not to be indifferent to the usage-flexibility of the resources of the biotechnological company. We claim that usage-flexibility of resources of the biotechnological company should decrease the flexibility requirements of organizational form, and thus be associated with MA. Since knowledge base is the critical resource in this setting [Arora and Gambardella (1990)], we hypothesise that the width of applicability of the internal knowledge of the biotechnological company will have a positive effect on the likelihood that the pharmaceutical company will acquire (or merge with) that biotechnological company.

[4] Rothaermel (2001) suggests that incumbents that focus their network strategy on exploiting complementary assets (downstream alliances) outperform incumbents that focus their network strategy on exploring the new technology (upstream alliances). Schweizer (2005) suggests that biotechnological effective integration by pharmaceutical companies (upstream mergers and acquisitions) demand to use a versatile post-acquisition integration approach. Higgins and Rodriguez (2006) find that the performance of mergers and acquisitions is correlated with acquirer's prior access to information of the targets' R&D activities (measured by previous established alliances) and with a superior negotiation position. Our main distinctive feature is that we focus on upstream alliances and compare that to upstream mergers and acquisitions.
Arora and Gambardella (1990) suggest that alternatives are complementary because they target distinct and complementary sets of resources. Our main distinctive feature is that we disregard agreements with other entities (such as universities) and include strategic alliances with biotechnological companies. We also focus on a slightly more mature stage of the event.
Rothaermel and Hess (2007) explore the interdependencies between human, firm and network levels of analysis. Hess and Rothaermel (2011) assess the degree of complementary between incumbent's star scientists and strategic alliances (upstream and downstream). Grigoriou and Rothaermel (2017) suggest that the effectiveness of incumbent's use of external knowledge is contingent on its internal knowledge characteristics. Our main distinctive feature is that we focus uniquely on incumbent's use of external knowledge.

Combining data from SDC, WRDS (North American Compustat data) and USPTO, we find results that support the hypothesis.

We think this is particularly enriching for several reasons. First, it provides additional insights on how pharmaceutical companies vertically integrate R&D (vertical backward integration strategy) after discovery of R-DNA. Second, it evidences the role of flexibility, and in this sense, it adds to a broader literature stream that has been addressing the preferences that companies have for external sources of innovative competencies [Hagedoorn and Duysters (2002)]. Finally, this enhances our understanding of how the biotechnological and pharmaceutical sector is affected by the dynamics of innovation [Dosi and Mazzucato (2009)], as well as other industries facing disruption [Gottinger *et al*. (2010)].

HYPOTHESIS DEVELOPMENT

The importance of external sources of technological knowledge

The switch from chemical based to biological based research implied that solid understanding of molecular actions and pathologies became necessary for drug research [Gambardella (1995)]. This represented an R&D competence-destroying event for incumbent pharmaceutical companies [Tushman and Anderson (1986); Powell *et al*. (1996); and, Grigoriou and Rothaermel (2017)]. At the same time, it opened a vast domain of research venues, leading to a higher division of innovative labour [Dosi and Mazzucato (2009)]. The result was a vertical disintegration and a fragmentation of the R&D layer [Grabowski and Kyle (2008); and, Dosi and Mazzucato (2009)], operated by many specialized biotechnological firms.

It has been claimed that, in the context of an emerging technological paradigm, incumbent companies frequently use SA and MA as alternative external sources of technological knowledge [Powell *et al*. (1996); Venhaverbeke *et al*. (2002); Hagedoorn and Duysters (2002); Hill and Rothaermel (2003); Rothaermel and Deeds (2004); Schweizer (2005); Higgins and Rodriguez (2006); Rothaermel and Hess (2007); and, Grigoriou and Rothaermel (2017)]. As such, we find it reasonable to conjecture that incumbent pharmaceutical companies respond to the competence-destroying event by establishing SA and MA with biotechnological companies. We study the decision between these two alternatives, in the post-R-DNA period.

The decision between MA and SA

MA and SA are fundamentally different corporate arrangements. Comparatively to MA, a SA is a more flexible – namely, it is more easily reverted – and provides less control over joint assets [Hoffmann and Schaper-Rinkel (2001); and, Yin and Shanley (2008)]. Accordingly, it has been claimed that companies base their decision between these two alternatives on the balance of flexibility and commitment requirements [Ghemawat and Sol (1998); Smit and Trigeorgis (2004); and, Yin and Shanley (2008)], where the flexibility leg states that, in uncertain environments, the necessity to adapt favours the adoption of waiting and staging strategies; and the commitment leg states that, in competitive markets, engaging in irreversible strategic investments may be beneficial for a company because it can influence competitors' decisions in its

favour. Accordingly, it is reasonable to conjecture that pharmaceuticals' decision between MA and SA will depend on the flexibility requirements associated to biotechnological companies. Regarding this, it has been claimed that usage-flexible resources ease the problem of commitment under conditions of uncertainty, by enabling companies that control them to try a wider domain of applications as key uncertainties are resolved [Ghemawat and Sol (1998)][5]. This suggests that, for the pharmaceutical company, the decision to acquire or ally with a biotechnological company depends on the level of usage-flexibility of resources of the biotechnological company[6]. In particular, usage-flexibility of resources of the biotechnological company should decrease the flexibility requirements of organizational form (that is, decrease the needs to revert an organizational deal), and thus be associated with MA (as opposed to SA). Since knowledge base is the critical resource in this setting [Arora and Gambardella (1990)], our main proposal is that the level of usage-flexibility of the knowledge base of the biotechnological company will have a positive effect on the likelihood that the pharmaceutical company will acquire (or merge with) that biotechnological company.

Usage-flexible knowledge base

Providing a precise definition of usage-flexibility of a knowledge base requires a definition of knowledge base. We define knowledge base of the biotechnological company as its internal knowledge – the knowledge within the boundaries of the biotechnological company – as opposed to a broader definition that would also include external knowledge – the knowledge within the boundaries of their horizontal partners[7]. Building upon this, and in line with Ghemawat and Sol (1998), we define usage-flexibility of a knowledge base as the extent to which internal knowledge enable trying a wider domain of applications. As such, we hypothesise the following.

Hypothesis – the width of applicability of the internal knowledge of the biotechnological company will have a positive effect on the likelihood that the pharmaceutical company will acquire (or merge with) that biotechnological company.

[5] Usage-flexible resources should be understood as resources with larger domains of application, when compared to usage-specific resources, in line with Ghemawat and Sol (1998).

[6] This is consistent with the dynamic capabilities' proposal that, in high uncertainty environments, companies are likely to pursue strategies that favour building and monitoring organizational capabilities that support flexibility [Teece *et al.* (1997)].

[7] Horizontal partners are biotechnological companies that recently engaged in an MA or SA with that biotechnological company. We provide more precise/empirical definitions below.

EMPIRICS

Data and sample

We combine data from SDC, WRDS (North American Compustat data) and USPTO. We start by using SDC data to identify MA and SA[8] established between pharmaceutical and biotechnological companies[9] (what we consider vertical deals), between 1989[10] and 1998. In order to do so, we first follow Cortright and Mayer (2002), to define the pharmaceutical industry by those companies with one of the following primary SIC codes: 2833 (medicinal chemicals and botanic products), 2834 (pharmaceutical preparations), 2835 (in vitro and in vivo diagnostic substances), 2836 (biological products, except diagnostic substances), 3826 (laboratory analytical instruments), 8731 (services-commercial physical and biological research) or 8733 (non commercial biological research)[11]. Second, we follow Rothaermel (2001), and partition the pharmaceutical industry in two sets: pharma companies as those companies having 2834 as primary SIC code; and, biotech companies as those that don't. We identified 1003 vertical deals between (and including) 1989 and 1998 – 288 of which are MA, while 715 are SA[12]. Regarding the 288 MA, 18 are mergers, 160 are acquisitions where the target is the biotech company, and 110 are acquisitions where the target is the pharma company. Since the goal is to model the pharma's decision to acquire a biotech, we keep the 18 mergers and the 160 acquisitions where the target is the biotech. Contrary to mergers and acquisitions, identifying the focus of strategic alliances is not as straightforward. We adopt the following procedure[13]. We classify SA by the nature and existence of joint activities, and drop 203 SA that are exclusively focusing on downstream activities (those that have joint manufacturing operations and/or joint marketing arrangements but do not have joint R&D), and keep the remaining 512 SA[14]. We end up with 690 vertical deals – 178 MA and 512 SA –

[8] We rely on SDC classification of strategic alliances. These include, among other: equity purchase; funding agreement; joint manufacturing operations; joint marketing arrangements; joint research and development; joint venture; licensing agreement; royalties; and supply agreements.
[9] Hereafter we indistinctively use: 'pharmaceutical' and 'pharma'; and, 'biotechnological' and 'biotech'.
[10] Data on SA is only available for time window 1985-1998. Moreover, the need to construct independent variables forces disregarding vertical deals established before 1989. Yet, the time window is comprehended in the knowledge adaptation process period (roughly between 1974 and 1998) and avoids the influence of alternative external shocks [Grigoriou and Rothaermel (2017)].
[11] The only exception is 3826, which was not on theirs.
[12] SDC provides up to 8 classifications for each strategic alliance. For these 715 SA we find: 96 equity purchase; 73 funding agreement; 142 joint manufacturing operations; 372 joint marketing arrangements; 425 joint research and development; 76 joint venture; 316 licensing agreement; 164 royalties; 25 supply agreements; and, 38 other classification.
[13] Similar to the one used by Hess and Rothaermel (2011) to distinguish between upstream and downstream alliances.
[14] 208 cases have either type of joint activities (R&D and manufacturing and/or marketing), while 87 cases have neither joint activity.

established between (and including) 1989 and 1998. This establishes our baseline sample (sample A). Table 1 describes the sample generating process from this point on, which encompasses the following adjustments to the sample.

[Table 1 around here]

First, we drop 205 vertical deals in which both companies are not identified as parent companies. This is to exclude cases where the identified company (say, a biotech) is a subsidiary of a parent company operating mainly at a distinct domain (say, a pharma company). Second, we drop 51 vertical deals in which pharma companies are identified as biotech companies in another vertical deal, and vice-versa. This is to focus on cases of more extreme (or less fuzzy) vertical boundaries[15]. After these restrictions we end up with 434 vertical deals – 99 MA and 335 SA – established between (and including) 1989 and 1998. Moreover, the probit model forces further restrictions on sample size[16]. First, we drop 224 vertical deals for which the biotech is not identified at WRDS (North American Compustat data)[17, 18]. Finally, we drop 24 vertical deals for which the biotech has missing values in core controls. We end up with 186 vertical deals (31 MA and 155 SA), 188 companies (89 biotech and 99 pharma), with considerably more U.S. representativeness at the biotech side – 85% vs. 52%. This establishes our sample E, which is our focus sample.

Variables and descriptive statistics

We define 1 independent variable and 3 sets of independent variables: biotech knowledge variables; biotech deal variables; and, other biotech and pharma variables. We use USPTO data for biotech knowledge variables[19]. Part of which are our key independent variables, while others are biotech knowledge controls. We use SDC data

[15] While we rely on primary SIC code variables on both SDC database for mergers and acquisitions and for strategic alliances, some of the dropped cases are situations in which these databases deliver different primary SIC codes. Yet, for some cases we identify that the same database delivers distinct primary SIC codes, which we expect to be associated with a change in companies' main activities. We decided to drop these cases to make sure we focus on cases of more extreme (or less fuzzy) vertical boundaries.

[16] As explained below the dependent variable is binary, and due to that we use a probit model to estimate the likelihood of a pharma to acquire a biotech.

[17] Note that we restrict biotech and not pharmaceutical companies because this would severely shrink the sample size. If we depart from baseline sample and only constraint to WRDS (North American Compustat data) data availability on core controls for both biotech and pharma, we would reach a sample size of 100 deals. As explained below in more detail, controls for pharmaceutical companies use only SDC data.

[18] Because Compustat North America covers publicly traded companies in the U.S. and Canada, kept biotech companies are publicly-traded, and most of which are North American.

[19] Note the following. To construct knowledge variables we depart from security identifiers provided by SDC, and used NBER files that match USPTO to WRDS (North American Compustat data). These files provide a unique company identifier and take into account company reorganizations.

for biotech deal variables, capturing: upstream horizontal deal activity of the biotech; and, vertical deal activity of the biotech. We use SDC and WRDS data for a more complete set of core controls: biotech and pharma variables. Table 2 describes each of these variables in more detail.

[Table 2 around here]

Dependent variable

We define the dependent variable – (1) *Vertical MA* – as a dummy variable, equal to 1 if the established vertical deal between the pharmaceutical company and the biotechnological company is a MA, and 0 if it is a SA.

Key independent variables

It has been claimed that in industries in which know-how is critical, both internal and external knowledge are of major importance [Powell *et al*. (1996)]. Biopharmaceutical industry is definitely not an exception here. Biotechnological companies do rely on organizational deals with their peers to complement their (internal) knowledge base, and the extent to which they do so most likely depends on achievable complementarities. Accordingly, our empirical proposal is to capture the width of applicability of the internal knowledge of the biotech as the extent to which external knowledge is: technologically distant from internal knowledge; and technologically dispersed[20,21]. That is, an internal knowledge base with wider applicability is one that is applicable to technological distant and technological dispersed domains.

We define internal knowledge as the knowledge within the boundaries of the biotech company in the year before the (vertical) event year, and external knowledge as the knowledge within the boundaries of their horizontal partners[22] in the (vertical) event year[23]. We operationalize this using patent data. The amount/size of internal

[20] We intentionally use organizational deals (SA and MA) to grasp knowledge 'relation'. We are assuming that two biotechnological companies enter into an organizational deal if and only if their knowledge bases are significantly related. We consider this a weak assumption because: on the one hand, biotechnology is knowledge-intensive, and as such it is likely that organizational deals are knowledge-driven; and, on the other hand, biotechnological companies are highly specialized.

[21] Note that we consider width of applicability to be conceptually related to the specificity of the internal knowledge, and not to the level of specialization of the biotech – which we consider to be conceptually related with the diversity of the internal knowledge. As explained below, we also control for level of specialization/diversity.

[22] Horizontal partners are biotech companies engaged with the biotech leg of the vertical deal (through an MA or an SA), within 4 years prior to the vertical link event (excluding the event year).

[23] In order to isolate biotech knowledge from possible impacts of the vertical deal (with the pharmaceutical company), we disregard assignee identifiers that are newly attributable to the biotech company in the (vertical) event year – as such we use the year before the (vertical) event year.

knowledge corresponds to the variable *Biotech Patents* and the amount/size of external knowledge corresponds to the variable *Partners Patents*[24,25,26].

To capture technological distance, we use two measures of technological relatedness between external and internal knowledge – variables (6) *Dissimilarity*[27] and (7) *Complementarity*[28]. We expect the impact of *Dissimilarity* to be positive and the impact of *Complementarity* to be negative. To capture technological dispersion, we use an inverted measure of class-concentration within the external knowledge – variable (8) *Partners Knowledge Dispersion*[29]. Accordingly, we expect the impact of *Partners Knowledge Dispersion* to be positive. Note that this is missing for cases where external knowledge is empty (no granted patents). We overcome this difficulty by adopting a dummy variable adjustment[30]. This entails plugging zero values to missing data cases and including a dummy equal to 1 if the horizontal partners of the biotech leg of the vertical deal have no granted patens (thus no classes), within 4 years prior to the vertical link event (excluding the event year) – variable (9) *Partners no Patent*.

Biotech knowledge controls

It is reasonable to admit that pharmaceutical companies may value both internal and external knowledge, and that they do so differently. Because both external and internal knowledge are controllable[31], and given the higher organizational flexibility associated to external knowledge, it can be argued that, keeping everything else constant, high external knowledge incidence would offer high customization

[24] As explained below, these are set as control variables.

[25] The procedure to identify *Biotech Patents* and *Partners Patents* assumes that when a company is acquired the patents go to the new owner. Because biotech companies engaged in a vertical merger and acquisition are parent companies, if horizontal MA motivate a change in the ownership of patents, it is likely that these patents will be mostly captured in *Biotech Patents*, as opposed to *Partners Patents*. Note that this procedure also implies disregarding patents granted to partners which where, in the meanwhile – that is, between the year of the horizontal deal and the (vertical) event year – acquired by another company.

[26] We are careful enough to avoid double counting, mainly because two biotech companies may be engaged in more than one deal. Even so, in very few cases assignee identifiers are repeated. In these cases we attribute the patents to the biotech. Moreover, we keep possible co-assigned patents. We are aware that this may inflate the variable *Partners Patents* (if co-assignment between partners exists) and the variable *Upstream Patents* (for that reason and also if co-assignment exists between the biotech and the partners).

[27] Technological dissimilarity is defined as equal to one minus the ratio of overlap of patents in the same class (*Biotech Patents* vs. *Partners Patents*) weighted by the relative importance of common classes for the biotech leg of the vertical deal (*Biotech Patents*). In other words, it is defined as one minus a measure of technological similarity proposed by Makri *et al.* (2010). Table 2 provides a more precise definition.

[28] This is similar to the measure of technological complementarity proposed by Makri *et al.* (2010). It is defined as the ratio of overlap of patents in the same subcategory but in different class (*Biotech Patents* vs. *Partners Patents*) weighted by the relative importance of common subcategories for the biotech leg of the vertical deal (*Biotech Patents*). Table 2 provides a more precise definition.

[29] This is constructed in the logic of a Herfindahl Index, where the shares of classes are computed with the corresponding number of patents.

[30] We do so because this approach keeps observations that would otherwise be dropped and because values are not missing at random.

[31] Internal knowledge is controllable for the obvious reason that it is within the boundaries of the company. External knowledge is controllable by means of redefining the set of horizontal partners.

possibilities. We control for this organizational dimension of flexibility by including variable (5) *Biotech-to-Upstream Patents Ratio*.

Finally, we add 2 knowledge related controls: variable (2) *Upstream Patents* – defined as the sum of variables (3) *Biotech Patents* and (4) *Partners Patents* – to control for the size of internal and external knowledge; and; variable (10) *Biotech Classes* to control for the diversity of internal knowledge.

Biotech deal controls

Because it can be argued that the value generated by a SA or an MA depends on companies' capabilities to engage in those arrangements, which are developed through repeated experience [Villalonga and McGahan (2005)][32], we add prior biotech-biotech deal variables to control for possible relative preference for a certain type of deal (SA or MA). To control for the possibility that the strength of engagement of SA influences pharmaceuticals' decisions, we further differentiate SA deals according to 2 criteria: either or not it is a licensing agreement; either or not it is an agreement for joint efforts[33]. While the first captures the scope of engagement (narrower set of assets in the case of licensing agreement vs. broader set of assets), the later captures the degree of engagement (high in the case of joint efforts vs. low in the absence of joint efforts). Furthermore, in line with previous literature [Higgins and Rodriguez (2006)], we control for the number of prior SA established with the same pharmaceutical and with other pharmaceutical company.

Other biotech and pharma controls

For the pharmaceutical and for the biotechnological company we add a dummy indicating if the nationality is U.S. and whether or not they share the same nationality. We also control for size of the biotechnological company.

Descriptive statistics

Table 3 presents descriptive statistics for our sample E. It is interesting to note that the relative incidence of mergers and acquisitions in the case of upstream horizontal deals is similar to the case of our 186 vertical deals, and considerably below 50%. In line with this, the mean of *Upstream LC SA* is higher than the mean of *Upstream HC SA*. Variables *Upstream MA* and *Upstream SA* show large ranges, suggesting that biotech companies vary significantly on their deal activity with their horizontal partners. Moreover, in 58% of the 186 vertical deals, the biotech leg of the vertical deal has

[32] Building upon: Dyer and Singh (1998); Haleblian and Finkelstein (1999); Anand and Khanna (2000); Hayward (2002); and, Kale *et al.* (2002).

[33] Joint efforts can either be: joint R&D and/or joint manufacturing operations and/or joint marketing arrangements.

established a SA with another pharma company, and in 8% of the 186 vertical deals, the biotech leg of the vertical deal has established a SA with the same pharma company. The amount of upstream knowledge also varies significantly by source: biotech (internal) vs. partners (external). Two upstream vertical SA illustrate that. In one case, the biotech has 0 patents and their partners have 2965; while in the other, the biotech has 1919 patents and their partners have 299 patents. *Dissimilarity* varies between 0.18 and 1, while *Complementarity* varies between 0 and 0.52. *Partners Knowledge Dispersion* varies between 0 and 0.95. Note that results for *Partners Knowledge Dispersion* should be interpreted cautiously as they are affected by the imputation procedure (dummy variable adjustment) – for example, the correlation between *Dissimilarity* and *Partners Knowledge Dispersion*. Finally, *Biotech Classes* shows a range that is larger than expected, taking into consideration the arguments of high level of specialization of biotechnological companies. The histogram in Panel A of Figure 1 clarifies that this is due to few outliers, further justifying the inclusion of this control.

[Table 3 around here]

[Figure 1 around here]

Analysis

Since the dependent variable is binary, we estimate the likelihood of a pharma to acquire a biotech using a probit model, with errors clustered at the pharmaceutical company level.

Results

Table 4 displays the results of alternative probit models. Ideally we would like to include biotech industry dummies (SIC) and year dummies (event year) to control for industry characteristics and time effects. This forces dropping a significant number of observations due to perfect prediction. We propose to overcome this difficulty by replacing these dummies with the following dummy variables: a dummy equal to one if the event year is in year 1994 or later; and, a dummy variable equal to one if boarder SIC code is 2833, 2835 or 2836. Models A1 and A2 try these alternatives for a set of controls. Results don't seem significantly sensitive to this alternative

specification for industry and time control. Hence, we proceed by using these last proposed dummies on Models A3 to B2. Model A3 further adds biotech-biotech deal level controls to account for possible relative preference for deal type (MA or SA), and Model A4 includes biotech knowledge controls and drops observations for which the variable *Upstream Patents* is zero. Models B1 to B2 further include our key dependent variables.

[Table 4 around here]

We find support for the hypothesis. First, the coefficients associated to the variable *Dissimilarity* are positive and statistically significant in model B1. Second, in model B2, the coefficient associated to the variable *Complementarity* is negative and statistically significant. Third, the coefficients associated to the variable *Partners Knowledge Dispersion* are positive and statistically significant in models B1 and B2. Taken together, we interpret these results as compelling evidence that the width of applicability of the internal knowledge of the biotechnological company has a positive effect on the likelihood that the pharmaceutical company acquires (or merges with) that biotechnological company.

Our confidence is further enhanced by the circumstance that several additional results are as expected. First, the coefficients associated to the variable *Biotech-to-Upstream Patents Ratio* are negative in all models B, which is consistent with the arguments supporting the existence of a relative preference for organizational-flexible knowledge (external knowledge). Second, results for biotech-biotech deal level controls reveal that a relative preference for a certain type of deal (MA vs. SA) is in play. Finally, the coefficients associated to the variable *Own Vertical SA* are positive, which is consistent with the view that strategic alliances may serve as a prior step to mergers and acquisitions.

Robustness tests

All these results are robust to replacement of variable *Upstream Patents* by the variable *Biotech Patents*[34] and the replacement of variable *Upstream SA* by the variables *Upstream HC SA* and *Upstream LC SA*[35].

[34] These results are reported on Table 5.
[35] These results are reported on Table 6.

We also add variables *Pharma MA* – number of vertical MA established by the pharma leg of the vertical deal, within 4 years prior to the vertical link event (excluding the event year) – and *Pharma SA* – number of vertical SA established by the pharma leg of the vertical deal, within 4 years prior to the vertical link event (excluding the event year) – to control for pharma deal experience[36]. While results suggest that deal experience is also in play for pharma, our main results remain mostly unchanged.

One important limitation of our study is that the inclusion of further pharma characteristics would imply a severe drop in sample size. Yet, because it could also be argued that pharma characteristics – namely, pharma size and pharma financial constraints – could provide alternative explanations for the decision between acquiring and merging with a biotech company, we test that by focusing on variables *Pharma Assets* – total assets of the pharma leg of the vertical deal, reported 1 year prior to the event year – and *Pharma OCF* – operating cash flow of the pharma leg of the vertical deal, reported 1 year prior to the event year – for different sample restrictions[37]. Results do not suggest that pharma size and pharma financial constraints play a significant role.

Finally, results are also robust if the analysis is confined to those observations for which the variable *Biotech Classes* is below 30 or below 50[38].

[36] These results are reported on Table 7.
[37] These results are reported on Table 8.
[38] These results are available upon request.

DISCUSSION

The discovery of R-DNA represented an R&D competence-destroying event for incumbent pharmaceutical companies. As a response to that, pharmaceutical companies have been using SA and MA as external sources of technological knowledge. In this study we made an effort to examine pharmaceutical's decision between these two alternatives. Recognizing the importance of usage-flexible resources of the biotechnological company, we hypothesise that the width of applicability of the internal knowledge of the biotechnological company will have a positive effect on the likelihood that the pharmaceutical company will acquire (or merge with) that biotechnological company.

Combining data from SDC, WRDS (North American Compustat data) and USPTO, we analyse internal and external knowledge of the biotechnological companies that are engaged in an organizational deal (MA or SA) with a pharmaceutical company, and find that MA is positively associated with: the technological distance between external and internal knowledge of the biotechnological company; and, the technological dispersion of external knowledge of the biotechnological company. We interpret these results as supportive of our hypothesis.

The main implication of this paper is that it suggests a backward vertical integration strategy for pharmaceutical companies: allying with biotechnological companies with more specific knowledge, and acquiring biotechnological companies with less specific knowledge. This is consistent with the proposition that *"the more specific the biotech know-how within an acquired biotechnology company, the more autonomy it is granted by an acquiring pharmaceutical company, and the sooner it becomes an independent centre of excellence within the pharmaceutical company"* [Schweizer (2005)]. More generally, this enhances our understanding of how the biotechnological and pharmaceutical sector is affected by the dynamics of innovation [Dosi and Mazzucato (2009)], as well as other industries facing disruption [Gottinger *et al*. (2010)]. Moreover, by evidencing the role of flexibility, it adds to a broader literature stream that has been addressing the preferences that companies have for external sources of innovative competencies [Hagedoorn and Duysters (2002)].

Our study entails several limitations. First, due to data limitations, we were unable to include further pharmaceutical level controls (in particular, the knowledge base of the

pharmaceutical company) and further pharma-biotech controls (in particular, the complementarity of knowledge bases). This is particularly critical because, from the perspective of the pharmaceutical company, both internal and external knowledge are potentially relevant. Second, we were unable to control for possible variation in commitment requirements, and as such, inference on the role of flexibility is conditional on the assumption that requirements for commitment are invariant. Third, it is likely that the way we capture knowledge 'relation' calls for refinement. If we fail to capture knowledge relation with deal relation – that is, if our assumption, that biotechnological companies enter into an organizational deal if and only if their knowledge bases are significantly related, does not hold – our measures of width of applicability of internal knowledge could be capturing alternative elements. For example, it could be argued that the variation in the technological distance between internal and external knowledge and technological dispersion of external knowledge could be associated to variation in risk-taking behaviour instead of variation of width of applicability of internal knowledge. While, as claimed before, we consider this assumption to be relatively weak – because biotechnological companies are highly specialized and biotechnology is knowledge-intensive, and as such it is likely that organizational deals are knowledge-driven – alternative empirical strategies exist. For example, one could assume that there is a knowledge relation if and only if there is a citation and use knowledge flows (forward and backward citations) to capture knowledge relation. Fourth, we are aware that the way we distinguish between internal from external knowledge is not free of criticism. Alternative ways to distinguish are feasible, namely by considering all horizontal MA as part of external knowledge. Finally, we share the typical problems of using patent data to measure knowledge. However, we don't think this is too severe here, for a couple of reasons: on the one hand, patenting is a standard activity in this industry; and, on the other hand, our 'older' observations are reasonably after the discovery of R-DNA.

ACKNOWLEDGMENTS

Wharton Research Data Services (WRDS) was used in preparing firm-level and dyad-level controls. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

REFERENCES

Anand B.; and, Khanna T. 2000. Do Firms Learn to Create Value? the case of alliances. Strategic Management Journal. 21(3): 295–315.

Arora, A.; and, Gambardella, A. 1990. Complementarity and External Linkages: the strategies of the large firms in biotechnology. The Journal of Industrial Economics. 38(4): 361-379.

Cortright, J.; and, Mayer, H. 2002. Signs of Life: the growth of biotechnology centers in the U.S.. The Brookings Institution. Center on Urban and Metropolitan Policy.

Diestre, L.; and, Rajagopalan, N. 2012. Are All 'Sharks' Dangerous? new biotechnology ventures and partner selection in R&D alliances. Strategic Management Journal. 33: 1115-1134.

Dosi, G.; and, Mazzucato, Knowledge Accumulation and Industry Evolution: the case of pharma-biotech. Cambridge University Press.

Dyer J.; and, Singh H. 1998. The Relational View: cooperative strategy and sources of interorganizational competitive advantage. Academy of Management Review. 23: 660 – 679.

Gambardella, A. 1995. Science and Innovation: the US pharmaceutical industry during the 1980s. Cambridge University Press. Cambridge.

Ghemawat, P.; and, Sol, P. 1998. Commitment Versus Flexibility? California Management Review. 40(4): 26-42.

Grabowski, H.; and, Kyle, M. 2008. Mergers and Alliances in the Pharmaceuticals: effects on innovation and R&D productivity. In K. Gugler and B. Yurtoglu (eds.), The Economics of Corporate Governance and Mergers. 262-287. Edward Elgar.

Grigoriou, K.; and, Rothaermel, F. 2017. Organizing for Knowledge Generation: internal knowledge network and the contingent effect of external knowledge sourcing. Strategic Management Journal. 38: 395-414.

Gottinger, H.; Umali, C.; and, Floether, F. 2010. Strategic Alliances in Biotechnology and Pharmaceuticals. Nova Science Publishers, Inc. New York.

Hagedoorn, J.; and, Duysters, G. 2002. External Sources of Innovative Capabilities: the preference for strategic alliances or mergers and acquisitions. Journal of Management Studies. 39(2): 167–188.

Haleblian J,; and, Finkelstein S. 1999. The Influence of Organizational Acquisition Experience. Administrative Science Quarterly. 44: 29–56.

Hayward, M. 2002. When Do Firms Learn From Their Acquisition Experience? evidence from 1990 – 1995. Strategic Management Journal. 23(1): 21–39.

Hess, A.; and, Rothaermel, T. 2011. When Are Assets Complementary? star scientists, strategic alliances, and innovation in the pharmaceutical industry. Strategic Management Journal. 23: 895-909.

Higgins, M.; and, Rodriguez, D. 2006. The Outsourcing of R&D Through Acquisitions in the Pharmaceutical Industry. Journal of Financial Economics. 80: 351-383.

Hill, C.; and, Rothaermel, F. 2003. The performance of Incumbent Firms in the Face of Radical Technological Innovation. Academy of Management Review. 28(2): 257-274.

Hoffmann, W.; and, Schaper-Rinkel, W. 2001. Acquire or Ally? a strategic framework for deciding between acquisition and cooporation. Management International Review. 41: 131-159.

Kale P.; Dyer J.; and, Singh H. 2002. Alliance Capability, Stock Market Response, and Long Term Alliance Success: the role of alliance function. Strategic Management Journal 23(8): 747–767.

Katila, R.; Rosenberg, J.; and, Eisenhardt, K. 2008. Swimming With Sharks: technology ventures and corporate relationships. Administrative Science Quarterly. 53: 295-332.

Makri, M.; Hitt, M.; and, Lane, P. 2010. Complementary Technologies, Knowledge Relatedness, and Invention Outcomes in High Technology Mergers and Acquisitions. Strategic Management Journal. 31(6): 602-628.

Powell, W.; Koput, K.; and, Doerr-Smith, L. 1996. Interorganizational Collaboration and the Locus of Innovation: networks of learning in biotechnology. Administrative Science Quarterly. 41(1): 116-145.

Rothaermel, F. 2001. Incumbent's Advantage Through Exploiting Complementary Assets Via Interfirm Cooperation. Strategic Management Journal. 30: 687–699.

Rothaermel, F.; and, Deeds, D. 2004. Exploration and Exploitation Alliances in Biotechnology: a system of new product development. Strategic Management Journal. 25(3): 201-221.

Rothaermel, F.; and, Hess, A. 2007. Building Dynamic Capabilities: innovation driven by individual-, firm-, and network-level effects. Organization Science. 18(6): 898-921.

Schweizer, L. 2005. Organizational Integration of Acquired Biotechnology Companies into Pharmaceutical Companies: the need for a hybrid approach. The Academy of Management Journal. 48(6): 1051-1074.

Shibayama, S.; Tanikawa, K.; Fujimoto, R.; and, Kimura, H. 2008. Effect of Mergers and Acquisitions on Drug Discovery: perspective from a case study of a japanese pharmaceutical company. Drug Discovery Today. 12: 86-93.

Smit, J.; and, Trigeorgis, L. 2004. Quantifying the Strategic Option Value of Technology Investments. Working Paper, Erasmus University Rotterdam & University of Cyprus.

Sytch, M.; and, Bubenzer, P. 2008. Research on Strategic Alliances in Biotechnology: an assessment and review. In H. Patzelt and T. Brenner (eds.), Handbook of Bioentrepreneurship: 105-131. Springer.

Teece, D.; Pisano, G.; and, Shuen, S. 1997. Dynamic Capabilities and Strategic Management. Strategic Management Journal. 18(7): 509-533.

Tushman, L.; and, Anderson, C. 1986. Technological Discontinuities and Organizational Environments. Administrative Science Quarterly. 31: 439–465.

van Beuzekom, B. and Arundel, A. 2009. OECD Biotechnology Statistics. OECD. Paris.

Venhaverbeke, W.; Duysters, G.; and, Noorderhaven, N. 2002. External Technological Sourcing Through Alliances or Acquisitions: an analysis of the application-specific integrated circuits industry. Organization Science. 13(6): 714-733.

Villalonga, B.; and, McGahan, A. 2005. The Choice Among Acquisitions, Alliances, and Divestures. Strategic Management Journal. 26: 1183–1208.

Yin, X.; and, Shanley, M. 2008. Industry Determinants of the 'Merger Versus Alliance' Decision. Academy of Management Review. 33(2): 473-491.

Zhang, J.; and, Baden-Fuller, C. 2010. The Influence of Technological Knowledge Base and Organizational Structure on Technology Collaboration. Journal of Management Studies. 47(4): 679-704.

Zucker, L.; and, Darby, M. 1997. Present at the Biotechnological Revolution: transformation of technological identity of a large incumbent pharmaceutical firm. Research Policy. 26: 429-446.

APPENDIX A – main figures and tables

Table 1 – Sample generating process

The Table describes the sample generating process.

Sample	Sample Description	Dropped vertical deals	Kept vertical deals (MA/SA)	Kept pharma companies	Kept biotech companies	Kept companies	Kept US pharma companies	Kept US biotech companies
A	Baseline	-	690 (178/512)	286	390	676	170	272
B	Deals for which pharma and biotech companies are eligible in terms of parent vs. non-parent	205	485 (108/377)	214	288	502	121	204
C	Deals for which pharma companies are not identified as biotech companies and vice-versa (between deal)	51	434 (99/335)	203	265	468	111	187
D	Bridge to WRDS (North American Compustat data): pharma unrestricted and biotech restricted to existing bridge	224	210 (38/172)	105	96	201	54	82
E	Deals for which biotech companies have non-missing values in core controls (WRDS)	24	186 (31/155)	99	89	188	51	76

Table 2 – Variable definitions

The Table describes the variables. Event year is the year of the vertical deal (deal established between the pharma and the biotech). Horizontal partners are biotech companies engaged with the biotech leg of the vertical deal (through an MA or an SA), within 4 years prior to the vertical link event (excluding the event year). Sample A was used to compute variables 15 and 16.

	Designation	Description
1	Vertical MA	Dummy variable, equal to 1 if the established upstream vertical deal is a MA, and 0 if it is a SA.
2	Upstream Patents	Equals [3]+[4].
3	Biotech Patents	Number of patents granted within 4 years prior to the vertical link event (excluding the event year) that are within the boundaries of the biotech leg of the vertical deal, in the year before the event year.
4	Partners Patents	Number of patents granted within 4 years prior to the vertical link event (excluding the event year) that are within the boundaries of the horizontal partners of the biotech leg of the vertical deal, in the event year.
5	Biotech-to-Upstream Patents Ratio	Equals [3]/[2].
6	Dissimilarity	One minus the ratio of overlap of patents in the same class (*Biotech Patents* vs. *Partners Patents*) weighted by the relative importance of common classes for the biotech leg of the vertical deal (*Biotech Patents*). In other words, it is defined as one minus a measure of similarity proposed by Makri *et al.* (2010). More specifically, Similarity equals the following, where 'overlap' is the sum of eligible *Biotech Patents* and *Partners Patents* for the corresponding criterion. [overlap all patent classes/*Upstream Patents*]*[Biotech patents in common classes/*Biotech Patents*].
7	Complementarity	Ratio of overlap of patents in the same subcategory but in different class (*Biotech Patents* vs. *Partners Patents*) weighted by the relative importance of common subcategories for the biotech leg of the vertical deal (*Biotech Patents*). Analogous to a measure of complementarity proposed by Makri *et al.* (2010). Complementarity equals the following, where 'overlap' is the sum of eligible *Biotech Patents* and *Partners Patents* for the corresponding criterion. [(overlap all patent subcategories-overlap all patent classes)/*Upstream Patents*]*[Biotech patents in common subcategories/*Biotech Patents*].
8	Partners Knowledge Dispersion	Equals to 1 minus the summation of the squared class shares of patents granted, within 4 years prior to the vertical link event (excluding the event year), to the horizontal partners of the biotech leg of the vertical deal (Partners Patents).
9	Partners No Patent	Dummy variable, equal to 1 if the horizontal partners of the biotech leg of the vertical deal have no granted patens (thus no classes) within 4 years prior to the vertical link event (excluding the event year).
10	Biotech Classes	Number of classes in which the biotech leg of the vertical deal was granted at least a patent within 4 years prior to the vertical link event (excluding the event year).
11	Upstream MA	Number of upstream horizontal MA established by the biotech leg of the vertical deal with its horizontal peers, within 4 years prior to the vertical link event (excluding the event year).
12	Upstream SA	Number of upstream horizontal SA established by the biotech leg of the vertical deal with its horizontal peers, within 4 years prior to the vertical link event (excluding the event year). Equals [12]+[13].
13	Upstream HC SA	Upstream SA characterized by some type of joint efforts and not characterized by a licensing agreement.
14	Upstream LC SA	Upstream SA that are not considered Upstream HC SA.
15	Other Vertical SA	Dummy variable, equal to 1 if the biotech leg of the vertical deal established an SA with a distinct pharma leg, within 4 years prior to the vertical link event (excluding the event year), and 0 otherwise.
16	Own Vertical SA	Dummy variable, equal to 1 if the two legs of the vertical deal established an SA with each other, within 4 years prior to the vertical link event (excluding the event year), and 0 otherwise
17	Pharma U.S.	Dummy variable, equal to 1 if the nation of the pharma leg of the vertical deal is the U.S., and 0 otherwise.
18	Biotech U.S.	Dummy variable, equal to 1 if the nation of the biotech leg of the vertical deal is the U.S., and 0 otherwise.
19	Same Nation	Dummy variable, equal to 1 if the legs of the vertical deal have the same assigned nation, and 0 otherwise.
20	Biotech Assets	Total assets of the biotech leg of the vertical deal, reported 1 year prior to the event year.

Table 3 – Descriptive statistics

The Table presents descriptive statistics for each variable of interest and controls, based on sample E. Panel A presents basic descriptive statistics. Panel B presents a correlation matrix. *P<10%; **P<5%; ***P<1%.

Panel A – basic

	Designation	Obs.	Mean	S.D.	Min.	Max.
1	Vertical MA	186	0.167	0.374	0	1
2	Upstream Patents	186	251.8	544.7	0	2965
3	Biotech Patents	186	49.56	157.9	0	1919
4	Partners Patents	186	202.2	487.9	0	2965
5	Biotech-to-Upstream Patents Ratio	125	0.566	0.431	0	1
6	Dissimilarity	186	0.859	0.257	0.184	1
7	Complementarity	186	0.055	0.115	0	0.519
8	Partners Knowledge Dispersion	186	0.275	0.372	0	0.949
9	Partners No Patent	186	0.575	0.496	0	1
10	Biotech Classes	186	7.629	14.09	0	102
11	Upstream MA	186	0.495	0.988	0	5
12	Upstream SA	186	2.715	4.555	0	19
13	Upstream HC SA	186	1.199	2.329	0	10
14	Upstream LC SA	186	1.516	2.385	0	19
15	Other Vertical SA	186	0.581	0.495	0	1
16	Own Vertical SA	186	0.081	0.273	0	1
17	Pharma U.S.	186	0.548	0.499	0	1
18	Biotech U.S.	186	0.774	0.419	0	1
19	Same Nation	186	0.468	0.500	0	1
20	Biotech Assets	186	1458	3518	0.18	23904

Table 3 – Descriptive statistics (cont.)

The Table presents descriptive statistics for each variable of interest and controls, based on sample E. Panel A presents basic descriptive statistics. Panel B presents a correlation matrix. *P<10%; **P<5%; ***P<1%.

Panel B – correlations

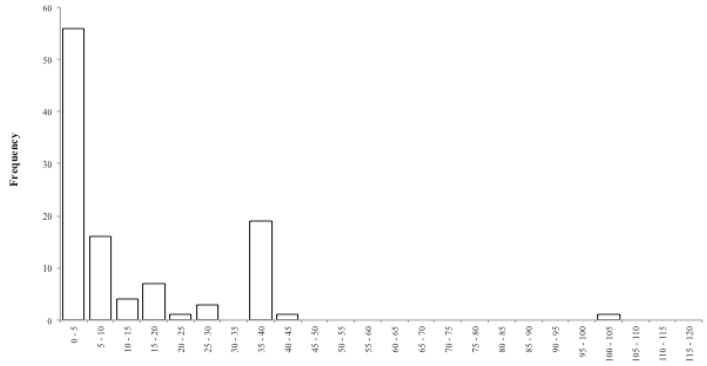
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	-0.16**																		
3	-0.12	0.49***																	
4	-0.14*	0.96***	0.22***																
5	-0.03	-0.51***	-0.01	-0.56***															
6	0.19**	-0.47***	-0.35***	-0.41***	0.38***														
7	-0.19**	0.61***	0.31***	0.58***	-0.39***	-0.37***													
8	-0.14*	0.66***	0.30***	0.64***	-0.83***	-0.68***	0.57***												
9	0.18**	-0.53***	-0.32***	-0.48***	0.77***	0.64***	-0.56***	-0.86***											
10	0.20***	0.61***	0.85***	0.41***	-0.09	-0.60***	0.39***	0.45***	-0.49***										
11	0.01	0.08	0.13*	0.04	-0.09	-0.26***	0.10	0.37***	-0.37***	0.24***									
12	-0.16**	0.52***	0.45***	0.43***	-0.22**	-0.58***	0.28***	0.41***	-0.54***	0.77***	0.23***								
13	-0.13	0.42***	0.44***	0.33***	-0.15*	-0.54***	0.21***	0.34***	-0.50***	0.75***	0.23***	0.97***							
14	-0.17**	0.57***	0.43***	0.50***	-0.29***	-0.59***	0.33***	0.45***	-0.54***	0.75***	0.22***	0.97***	0.87***						
15	-0.14**	0.30***	0.25***	0.25***	-0.09	-0.39***	0.24***	0.29***	-0.36***	0.40***	0.22***	0.39***	0.35***	0.40***					
16	0.19**	0.02	-0.06	0.04	-0.04	0.01	0.20***	0.03	0.01	-0.07	0.01	-0.09	-0.10	-0.06	0.05				
17	0.00	0.02	0.04	0.01	0.07	0.08	-0.09	-0.05	0.10	-0.03	-0.07	-0.04	0.00	-0.07	-0.03	-0.01			
18	0.10	-0.31***	-0.39***	-0.22***	0.22**	0.33***	-0.12*	-0.24***	0.29***	-0.55***	-0.11	-0.63***	-0.61***	-0.60***	-0.15**	0.02	0.00		
19	0.01	-0.06	-0.15**	-0.02	0.07	0.14*	-0.17**	-0.10	0.17**	-0.19***	-0.01	-0.15**	-0.11	-0.17**	-0.08	-0.08	0.68***	0.30***	
20	-0.09	0.44***	0.76***	0.25***	-0.00	-0.40***	0.24***	0.25***	-0.39***	0.86***	0.15**	0.80***	0.81***	0.73***	0.31***	-0.06	0.03	-0.59***	-0.14*

Figure 1 – Level of specialization of biotechnological companies

The Figure shows the histogram for classes within *Biotech Patents* and *Partners Patents*, for our sample E. It excludes cases of zero classes (due to zero patents).

Panel A – biotech

Corresponds to the variable *Biotech Classes*. It excludes 78 cases of zero classes (due to zero patents).



Panel B – partners

It excludes 107 cases of zero classes (due to zero patents and possibly zero deals). Note that unit of analysis here is the pool of partners.

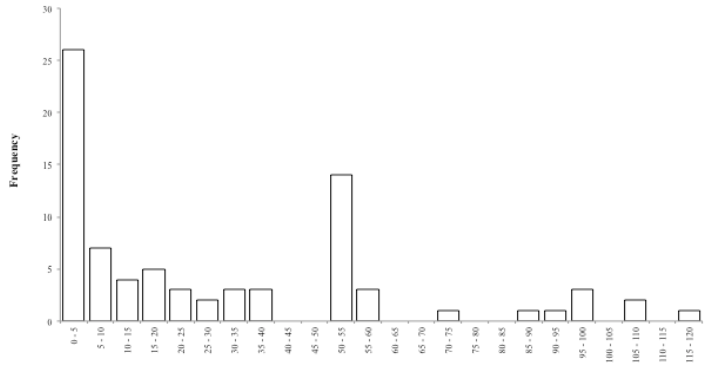


Table 4 – Results

The Table presents the results of alternative probit models. All models control for industry effects and time. Robust standard errors, clustered at the pharma company level, are reported in parenthesis. *P<10%; **P<5%; ***P<1%. Models B exclude cases with zero *Upstream Patents*.

Model	**(A1)**	**(A2)**	**(A3)**	**(A4)**	**(B1)**	**(B2)**
Upstream Patents				0.0004605 (0.00031)	0.0012581 (0.0014917)	-0.0003946 (0.0007087)
Biotech-to-Upstream Patents Ratio				-0.6084625 (0.6580661)	-4.552329** (2.013632)	-2.895428 (2.444808)
Dissimilarity					**35.59048***** **(12.41138)**	
Complementarity						**-6.283886***** **(1.86395)**
Partners Knowledge Dispersion					**13.39961***** **(4.678663)**	**3.371424**** **(1.405349)**
Partners No Patent					2.480849** (1.114179)	3.475881* (2.083252)
Biotech Classes				-0.1054819 (0.0640501)	0.0977621 (0.0849748)	-0.0422711 (0.0893393)
Upstream MA			0.2056085 (0.1258541)	0.7401273** (0.2989838)	4.196928*** (1.613973)	0.6469116** (0.3145717)
Upstream SA			-0.1826186*** (0.0671454)	-0.3276314** (0.1456337)	-2.701515** (1.067067)	-0.3459938* (0.1995506)
Upstream HC SA						
Upstream LC SA						
Other Vertical SA	-0.5665795* (0.316844)	-0.5927921** (0.3139478)	-0.4646432 (0.3089432)	0.2011312 (0.3971517)	1.214552 (0.9267981)	0.4993909 (0.568257)
Own Vertical SA	0.9983417*** (0.2525285)	1.057216*** (0.3153217)	0.9794228*** (0.3427316)	0.8634734** (0.4201371)	8.413154*** (2.581105)	1.14213 (0.7045262)
Pharma U.S.	-0.5237233 (0.3599466)	-0.2440444 (0.2618392)	-0.3201962 (0.291003)	0.311859 (0.6699128)	0.6045358 (3.846964)	0.722788 (0.678309)
Biotech U.S.	1.124578** (0.4525807)	0.9581687** (0.4043911)	0.5410075 (0.3521303)	1.17865 (0.7364819)	12.45168*** (4.051034)	1.546445** (0.6683081)
Same Nation	0.1412451 (0.4011774)	-0.1235507 (0.2588062)	-0.0009105 (0.3070685)	0.1600301 (0.7280088)	-0.2738365 (3.500686)	-0.5116357 (0.72024)
Biotech Assets	0.0000433 (0.0000603)	-0.0000514 (0.000043)	0.0000172 (0.0000361)	0.0003581 (0.0002471)	0.0002823** (0.0001188)	0.0002677*** (0.0000924)
Constant	-0.9042972 (0.6248261)	-3.464825*** (0.5682005)	-3.232207*** (0.5124773)	-4.581558*** (1.062518)	-63.95065*** (21.08339)	-6.98576*** (1.336719)
Area Under ROC Curve	0.9311	0.8966	0.9095	0.9519	0.9766	0.9629
McFadden's Pseudo R^2	0.4793	0.3783	0.4091	0.5719	0.6997	0.6221
Prob>Chi^2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Clusters (pharma)	85	99	99	82	82	82
Observations (deals)	157	186	186	125	125	125

APPENDIX B – secondary figures and tables

Table 5 – Robustness to alternative knowledge controls

The Table is analogous to Table 4 (models B), except that variable *Biotech Patents* replaces the variable *Upstream Patents*. *P<10%*; **P<5%; ***P<1%.

Model	(B1)	(B2)
Biotech Patents	0.0072349 (0.0082925)	-0.0003223 (0.0124506)
Biotech-to-Upstream Patents Ratio	-4.979207** (2.27157)	-2.727916 (2.494728)
Dissimilarity	**36.1446***** **(11.8611)**	
Complementarity		**-5.873368***** **(1.814128)**
Partners Knowledge Dispersion	**14.05075***** **(4.718917)**	**3.123145**** **(1.42284)**
Partners No Patent	2.945722** (1.313618)	3.225577 (2.101892)
Biotech Classes	0.0902127 (0.1016238)	-0.0457169 (0.1180491)
Upstream MA	3.989538*** (1.341867)	0.6783155** (0.3071863)
Upstream SA	-2.572518*** (0.8816755)	-0.3863572** (0.1819195)
Upstream HC SA		
Upstream LC SA		
Other Vertical SA	1.166251 (0.8734766)	0.5213559 (0.5235258)
Own Vertical SA	8.657354*** (2.498953)	1.170504 (0.7151435)
Pharma U.S.	-4.67829 (3.250625)	0.5523605 (0.5978552)
Biotech U.S.	14.97915** (5.884755)	1.607239** (0.7361025)
Same Nation	5.177952 (3.372954)	-0.3873933 (0.659965)
Biotech Assets	0.0002626** (0.0001144)	0.0002757*** (0.0000907)
Constant	-67.31037*** (22.13312)	-6.90756*** (1.368463)
Area Under ROC Curve	0.9753	0.9622
McFadden's Pseudo R^2	0.6986	0.6210
Prob>Chi^2	0.0000	0.0000
Clusters (pharma)	82	82
Observations (deals)	125	125

Table 6 – Robustness to alternative deal controls

The Table is analogous to Table 4 (models B), except that variable *Upstream HC SA* and *Upstream LC SA* replace the variable *Upstream SA*. *P<10%*; **P<5%; ***P<1%.

Model	**(B1)**	**(B2)**
Upstream Patents	0.0006191 (0.0014176)	-0.0007355 (0.0009758)
Biotech-to-Upstream Patents Ratio	-7.051704*** (2.204536)	-4.164205* (2.204536)
Dissimilarity	**36.81558***** **(10.96287)**	
Complementarity		**-6.758551***** **(2.154692)**
Partners Knowledge Dispersion	**14.85915***** **(3.987402)**	**5.019122***** **(1.601679)**
Partners No Patent	4.246713*** (1.531644)	5.108408** (2.125122)
Biotech Classes	0.2499233** (0.1002764)	-0.0069137 (0.0564745)
Upstream MA	3.927232*** (1.275999)	0.5175611* (0.2830072)
Upstream SA		
Upstream HC SA	-3.768397*** (1.084137)	-0.8902064*** (0.3001662)
Upstream LC SA	-2.372548** (0.9393867)	-0.1298979 (0.2562074)
Other Vertical SA	0.8424977 (0.8820536)	0.5553574 (0.6064981)
Own Vertical SA	9.751172*** (2.268078)	1.597725** (0.6695509)
Pharma U.S.	-6.016753 (3.689073)	0.634207 (0.8369923)
Biotech U.S.	15.50904*** (4.559765)	1.587203** (0.7174402)
Same Nation	6.154392* (3.647569)	-0.3486459 (0.67554)
Biotech Assets	0.0002628** (0.0001117)	0.0002605** (0.0001112)
Constant	-69.61162*** (19.22503)	-8.196646*** (1.502705)
Area Under ROC Curve	0.9794	0.9677
McFadden's Pseudo R^2	0.7133	0.6415
Prob>Chi^2	0.0000	0.0000
Clusters (pharma)	82	82
Observations (deals)	125	125

Table 7 – Robustness to inclusion of pharma deal experience

The Table is analogous to Table 4 (model B2), except that we add variables *Pharma MA* – number of vertical MA established by the pharma leg of the vertical deal, within 4 years prior to the vertical link event (excluding the event year) – and *Pharma SA* – number of vertical SA established by the pharma leg of the vertical deal, within 4 years prior to the vertical link event (excluding the event year). Sample A was used to compute these variables. Results for model B1 are not presented, as convergence was not achieved. *P<10%*; **P<5%; ***P<1%.

Model	**(B2)**
Upstream Patents	-0.0015758* (0.0008893)
Biotech-to-Upstream Patents Ratio	-2.522594 (2.406654)
Dissimilarity	
Complementarity	**-6.675304*** (2.119292)**
Partners Knowledge Dispersion	**4.418576*** (1.505755)**
Partners No Patent	3.746152* (2.075916)
Biotech Classes	-0.1044899 (0.0930259)
Upstream MA	0.7880926** (0.3641742)
Upstream SA	-0.2014026 (0.2092514)
Upstream HC SA	
Upstream LC SA	
Other Vertical SA	0.2050096 (0.5228911)
Own Vertical SA	-0.9269861 (0.8782442)
Pharma MA	0.8591314*** (0.3238183)
Pharma SA	-0.1395421 (0.1396416)
Pharma U.S.	1.150504* (0.6620684)
Biotech U.S.	1.797834** (0.7731881)
Same Nation	-1.312133* (0.7172104)
Biotech Assets	0.0003275** (0.0001302)
Constant	-7.825538*** (1.517522)
Area Under ROC Curve	0.9698
McFadden's Pseudo R^2	0.6508
Prob>Chi^2	0.0000
Clusters (pharma)	82
Observations (deals)	125

Table 8 – Testing the alternative explanations – pharma financial constraints and size

The Table is analogous to Table 4 except that we focus on variables *Pharma Assets* – total assets of the pharma leg of the vertical deal, reported 1 year prior to the event year – and *Pharma OCF* – operating cash flow of the pharma leg of the vertical deal, reported 1 year prior to the event year. Note that the observations in models C1 to C3 depart from baseline sample with further additional constraints on data availability from WRDS (North American Compustat data). Model C4 departs from sample E with further constraints on data availability from WRDS (North American Compustat data) at pharma variables *Pharma Assets* and *Pharma OCF*. *P<10%*; **P<5%; ***P<1%.

Model	**(C1)**	**(C2)**	**(C3)**	**(C4)**
Upstream Patents				-0.0173207 (0.0276303)
Upstream MA			-0.070661 (0.1745629)	0.1328951 (0.4619224)
Upstream SA			-0.1804755*** (0.0667835)	-0.6553494*** (0.2442326)
Other Vertical SA			-0.6974715 (0.4431186)	-0.7830496 (0.5535991)
Own Vertical SA			0.3015001 (0.6947488)	0.2204666 (0.8093782)
Pharma U.S.			1.180769* (0.6491745)	1.337908 (0.9771603)
Biotech U.S.			-0.6613857 (0.5586514)	-1.271078 (0.9389101)
Same Nation			0.719692 (0.6922558)	1.381137 (0.9333125)
Biotech Assets		-0.000101*** (0.0000374)	-0.000000 (0.0000429)	0.0009301 (0.0006677)
Pharma Assets	**-0.0001118** (0.0000504)**	**0.0000571 (0.0000605)**	**0.0000767* (0.0000462)**	**0.000405 (0.0000896)**
Pharma OCF	**0.0005399 (0.0003422)**	**-0.0002551 (0.0002899)**	**-0.0005131 (0.0003272)**	**0.0002678 (0.0008022)**
Constant	-1.220263*** (0.2189548)	-2.192018*** (0.7273224)	-3.134976*** (1.138975)	-6.059128*** (1.372889)
Area Under ROC Curve	0.6988	0.8484	0.9153	0.9562
McFadden's Pseudo R^2	0.0992	0.2641	0.4318	0.6539
Prob>Chi^2	0.0000	0.0000	0.0002	0.0000
Clusters (pharma)	71	40	40	26
Observations (deals)	226	100	100	67

The Effect of SOX Section 404 on Long-Term Investment of Small Innovative Companies

Fernando Santos
Universidad Carlos III de Madrid
Department of Business Administration
fjasantos@gmail.com | fasantos@porto.ucp.pt

PhD BF | Working Paper
Supervisors: Andrea Fosfuri and Neus Palomeras

ABSTRACT

In this paper we examine the impact of SOX Section 404 on long-term investment of small innovative companies. We hypothesise that R&D intensity increases the impact of SOX 404 on long-term investment. Making use of a quasi-natural experiment, our results suggest that the impact of SOX Section 404 on companies' long-term investment is uneven, favouring R&D intensive companies. This may call for a re-centring of policy discussion around the distribution of the net benefit of coercive financial disclosure programs *versus* the overall economic impact of those programs.

INTRODUCTION

Is long-term investment affected by coercive financial disclosure? We find arguments pointing in opposing directions. On the one hand, coercive financial disclosure may foster companies' long-term investment due to a decrease in information asymmetry between insiders and outsiders. However, on the other hand, coercive financial disclosure may hinder long-term investment due to an increase in direct and indirect costs of compliance.

Public companies are associated with information asymmetry between insiders and outsiders. In the financial markets, outsiders' confidence on financial reporting is an essential element for a proper functioning of financial markets because their investment decisions rely on disclosed information. Coercive financial disclosure arises by the need to improve or restore public confidence in financial reporting. From the late 1990's to the early 2000's a series of accounting and auditing scandals eroded public confidence – of which Enron and Worldcom are the most preeminent examples – and motivated to the enactment of Sarbanes-Oxley Act (SOX), the most significant change in securities' regulation since Securities Act 1933.

Section 404 of SOX is one of the most important and contentious components SOX [Zhang (2007); and, Coates and Srinivasan (2014)]. SOX 404 Section– designated "management assessment of internal controls" – requires most publicly registered companies to include on their annual report a management's assessment on the effectiveness of the company's internal control over financial reporting along with an external auditor's verification on this assessment. This is intended to improve quality of financial disclosure and thus, it is expected to impact the confidence and the monitoring power of external investors, solving adverse selection and moral hazard problems. There is evidence consistent with the positive impact of SOX 404 on the quality of firms' information environment [Alexander *et al.* (2013)][1]. Accordingly, we find proponents of the idea that SOX 404 would ultimately foster companies' long-term investment, through a decrease in the cost of capital driven by a decrease in information asymmetry between insiders and outsiders [Ashbaugh-Skaife *et al.*

[1] *"SEC administered the survey between December of 2008 and January of 2009. Managers from 2,901 unique U.S. public companies participated in the survey […] 80% of respondents ascribe some benefits to Section 404 compliance. Large fractions of respondents cite a positive impact of compliance on the quality of their firm's internal control structure (73%), their audit committee's confidence in the ICFR (71%), the quality of their firm's financial reporting (48%), and their firm's ability to prevent and detect fraud (47%)"* [Alexander et al. (2013)].

(2009); Singer and You (2011); Arping and Sautner (2013); Andrade *et al.* (2014); and, Albuquerque and Zhu (WP2014)].

However, we also find proponents of a concurring idea. SOX 404 may hinder long-term investment because it increases direct and indirect costs of compliance [Bargeron *et al.* (2010); and, Kang *et al.* (2010)]. Regarding direct costs, the argument is that the increase in compliance costs implies more severe financial constraints (namely, to long-term investment). There is evidence consistent with the positive impact of SOX 404 on audit fees, internal labor costs, outside consultants and other miscellaneous expenses [Alexander *et al.* (2013)]. Regarding indirect costs, the argument is that risky projects (namely, long-term investments) may be sacrificed, as they are associated with more weaknesses in financial reporting – which increase direct costs of compliance if weaknesses are addressed or stock price decline and/or litigation if weaknesses are not addressed. These potential unintended consequences of SOX echoed on policy and business domains. Allan Greenspan stated, *"corporate executives and boards of directors are seemingly unclear, in the wake of the recent intense focus on corporate behaviour, about how an increase in risk-taking on their part would be viewed by shareholders and regulators. As a result, business leaders have been quite circumspect about embarking on major new investments"*[2]. William Donaldson, Chairman of SEC stated, *"I worry about the loss of risk-taking zeal. […] [as it results in a] huge preoccupation with the dangers and risks of making the slightest mistake, as opposed to a reasonable approach to legitimate business risk"*[3]. Finally, Alan Eisenberg, on behalf of the Biotechnology Industry Organization, stated, *"many emerging biotech companies are directing precious resources from core research and development of new therapies for patients due to overly complex controls or unnecessary evaluation controls"*[4].

It is not surprising that the attention of scholars has been gravitating on the costs and benefits and on indirect consequences of SOX 404 implementation – in particular, the impact on corporate long-term investment. To our knowledge, the existing evidence suggests that compliance is (on average) positively associated with corporate investment [Albuquerque and Zhu (WP2014)]. While we find this compelling

[2] Testimony of Chairman Allan Greenspan before the Committee on Financial Services, U.S. House of Representatives, July 15, 2003.
[3] Adrian Michaels. After a year of US corporate clean-up, William Donaldson calls for a return to risk-taking. FinancialTimes.com, July 24, 2003.
[4] Comment letter to the SEC, July 12, 2007.

evidence, we think it is likely, as with virtually any policy intervention, that the net benefits of coercive financial disclosure – namely, those of SOX Section 404 – are heterogeneous across companies. To our knowledge, we are the first to investigate heterogeneity in this context. We hypothesise and find supportive evidence that R&D intensity increases the impact of SOX 404 on long-term investment.

Our findings are mostly relevant because they contribute to enrich the discussion around SOX. The suggestion that coercive financial disclosure (at least of this type) is an "uneven game" that favours R&D intensive companies may call for a re-centring of policy discussion around the distribution of the net benefit of coercive financial disclosure programs *versus* the overall economic impact of those programs. Moreover, we find no empirical support to the arguments relating an increase in direct and indirect costs of compliance with a decrease in corporate investment, and find empirical support to those relating a decrease in information asymmetry with an increase in corporate investment. This motivates further research as, for instance, other dimensions of information asymmetry could be relevant.

HYPOTHESIS DEVELOPMENT

Decrease in asymmetric information

Disclosure plays a role on solving moral hazard and adverse selection problems associated with companies' financing decisions. Moral hazard arises with the separation between ownership and control. In these circumstances managers may take investment or financing decisions that are not in the best interest of shareholders – a typical agency problem [Jensen and Meckling, (1976)]. Even in the absence conflicting interests between managers and shareholders, moral hazard may also arise with co-existence of debt and equity. In these circumstance companies may take investment or financing decisions that are not in the best interest of debt holders [Jensen and Meckling, (1976), Smith and Warner (1979)]. In both situations the moral hazard arises because, after a financing contract is established (either equity or debt), companies' decision makers may take further decisions that deviate and harm other contract parties. Adverse selection, on the other hand, arises from *ex ante* information asymmetry between contract parties of a financing contract, and corresponds to a typical lemons problem. It has been argued that optimal contracts between insiders (managers) and outsiders (investors) is a potential solution for both lemons and agency problems, because they will frequently require insiders to reveal relevant information to outsiders, allowing more effective monitoring (which eases moral hazard) and mitigating misvaluation problems (which eases adverse selection) [Healy and Palepu (2001)]. In line with this, a vast literature suggests that higher disclosure (in general) and a higher quality financial reporting (in particular) decreases the cost of (external) capital [Diamond and Verrecchia (1991); Welker (1995); Botosan (1997); Healy *et al.* (1999); Leuz and Verrechia (2000); Bushman and Smith (2001); Healy and Palepu (2001); Botosan and Plumlee (2002); Bens and Monahan (2004); and, Lambert *et al.* (2007)][5]. Under this perspective, investors demand a premium for bearing information risk[6], which decreases with disclosure [Healy and Palepu (2001)]. This is consistent with the literature suggesting that SOX (in general) and Section 404 (in particular) reduced the information asymmetry between insiders and outsiders – through a more reliable/accurate financial reporting – leading to a decrease in the cost of capital [Ashbaugh-Skaife *et al.* (2009); Singer and You (2011); Arping and

[5] They mainly focus on voluntary disclosure.
[6] This is in line with the notion of 'estimation risk' [Barry and Brown (1985)].

Sautner (2013); Andrade *et al.* (2014); and, Albuquerque and Zhu (WP2014)]. Accordingly, one would expect that by lowering the cost of capital, Section 404 of SOX would ultimately foster companies' long-term investment. Because R&D intensive companies have higher levels of information asymmetry [Aboody and Lev (2000); and, Iliev (2010)], keeping everything else constant, we would expect them to experience a relative higher decrease in the cost of capital.

Increase in direct and indirect costs of compliance

However, a concurring argument has been presented in the SOX 404 context. According to this alternative perspective, SOX 404 may hinder long-term investment because it increases direct costs and indirect costs of compliance.

Direct costs

According to the Securities Exchange Commission (SEC), Section 404 of the Sarbanes-Oxley Act requires that public companies' annual reports should include companies' own assessment of internal control over financial reporting, and that this is attested by an auditor[7]. Management must identify the financial reporting risks, the controls that address them and provide evidence (documentation) to provide support for its assessment on whether internal control is effective. Assessment is based on control deficiencies identified and the extent to which they constitute a material weakness[8]. Part of the controversy around Section 404 is that its compliance is costly and that this burdens companies distinctively. The discussion around Section 404 of SOX indicates that size and organizational complexity are the most significant

[7] According to SEC Release 33-8238 and 34-47986, 14 August 2003, the report *"contains:* [1.] *a statement of management's responsibility for establishing and maintaining adequate internal control over financial reporting for the company;* [2.] *a statement identifying the framework used by management to conduct the required evaluation of the effectiveness of the company's internal control over financial reporting;* [3.] *management's assessment of the effectiveness of the company's internal control over financial reporting as of the end of the company's most recent fiscal year, including a statement as to whether or not the company's internal control over financial reporting is effective. The assessment must include disclosure of any 'material weaknesses' in the company's internal control over financial reporting identified by management. Management is not permitted to conclude that the company's internal control over financial reporting is effective if there are one or more material weaknesses in the company's internal control over financial reporting; and,* [4.] *a statement that the registered public accounting firm that audited the financial statements included in the annual report has issued an attestation report on management's assessment of the registrant's internal control over financial reporting."*

[8] According to SEC Release 33-8810 and 34-55929, 27 June 2007, a material weakness is *"a deficiency, or a combination of deficiencies, in ICFR [Internal Control over Financial Reporting] such that there is a reasonable possibility that a material misstatement of the registrant's annual or interim financial statements will not be prevented or detected on a timely basis."*

compliance cost determinants[9]. Based on this, we assume that the increase in direct costs of compliance is homogeneous across levels of R&D intensity.

Indirect costs

It has been argued that investing in risky projects increases the likelihood of reporting material weaknesses, and that these can lead to stock price decline and/or litigation[10]. According to this perspective, to minimize material weaknesses companies would avoid risky investments [Bargeron *et al.* (2010); Kang *et al.* (2010); and, Albuquerque and Zhu (WP2014)].

Since long-term investment of R&D intensive companies is likely to involve riskier projects, this perspective suggests that the long-term investment of these companies to be particularly hindered by SOX 404. However, if investing in risky projects increases the likelihood of reporting material weaknesses, an alternative plausible solution (to material weaknesses avoidance) would be that companies solve material weaknesses through a more effective (and arguably costly) internal control over financial reporting. We assume that companies' willingness to minimize material weaknesses through avoidance only increases if solving material weaknesses is costly. In that case, we would expect a clear association between direct costs of compliance and level of risky project on the extensive discussion around direct compliance cost determinants. We don't find that. As such, we interpret this absence of discussion as evidence of either: investing in risky projects does not increase the likelihood of reporting material weaknesses; or, it does, but material weaknesses of this type do not lead to stock price decline and/or litigation.

Hypothesis

While SOX 404 may hinder long-term investment due to increase in direct and indirect costs of compliance, we assume this impact is homogeneous across levels of R&D intensity, and hypothesise the following.

Hypothesis – R&D intensity increases the impact of SOX 404 on long-term investment.

---

[9] According to SEC Release 33-8238 and 34-47986, 14 August 2003, *"many commenters indicated that even the more limited definition related to financial reporting that we proposed will impose substantial reporting and cost burdens on companies […] We believe that there will be a marked disparity of burdens and costs resulting from the new internal control requirements between the largest and smallest reporting companies. […] This burden will also vary among companies based on the complexity of their organization and the nature of their current internal control procedures."*

[10] Triggered by (outside) investors.

EMPIRICS

We implement a natural quasi-experiment to find further evidence on the impact of financial disclosure on long-term investment of small (non-financial, U.S. incorporated) innovative companies. In particular, we implement a regression discontinuity design using Section 404 of SOX to assess whether R&D intensity increases the impact of SOX 404 on long-term investment.

Context

Passed in 2002, SOX introduced a set of new or expanded disclosure requirements for public companies, aiming to improve quality of financial reporting and, in turn, outside investor confidence. One of the most debated items of SOX is Section 404. According to the Securities Exchange Commission (SEC), Section 404 of the Sarbanes-Oxley Act requires that public companies' annual reports should include companies' own assessment of internal control over financial reporting, and that this is attested by an auditor. Hereafter, we refer to this as management report.

Implementation of Section 404 for small firms is different from that of large firms. This is because these requirements have important compliance costs that impact smaller firms disproportionately (due to non-negligible fixed costs involved). While most U.S. public companies had to file their management report with audit assessment along with annual reports for the fiscal year ending on or after November 15, 2004, small firms were under an exception rule as follows. Companies with a public float[11] above $75 million in their reports for fiscal years 2002 (from November 2002 to October 2003), 2003 (from November 2003 to October 2004) or 2004 (from November 2004 to October 2005) were required to comply with Section 404[12]. Hereafter, we refer to these firms as MR companies. Companies with a public float under $75 million in their reports for fiscal years 2002, 2003 or 2004 were not required to comply with Section 404[13]. Hereafter, we refer to these firms as non-MR

[11] Public float is the part of equity that is not held by managers or large shareholders. It is reported on 10-k filing front page.

[12] These are companies considered accelerated filers by SEC. Accelerated filler status was introduced in December 15, 2002 by SEC. Once a company is classified as such, it is classified as such thereafter (with very few exceptions, not present in our sample). Accelerated filers have to file the 10-k report within 75 days of their fiscal year-end, while non-accelerated filers have to file the 10-k report within 90 days of their fiscal year-end. Every year the status is defined based on public float threshold.

[13] These are companies considered non-accelerated filers by SEC.

companies. These companies finally had to submit an MR without audit assessment in the fiscal year 2007, and finally had to submit an MR with audit assessment in 2010[14].

Identification

We make use of this compliance rule for small companies to implement a regression discontinuity design[15]. In concrete, we compare the outcomes of companies in the neighbourhood of the $75 million threshold – between $50 million and $100 million – using the following model.

$$Y = \beta_0 + \beta_1 . MR + \beta_2 . R\&DI + \beta_3 . MR . R\&DI + \sigma . PFL\ Terms + \alpha . X + \varepsilon$$

MR is a dummy variable that identifies companies that filed a management report. *R&DI* is a measure of R&D intensity, specified on data section. *MR.R&DI* is a term that interacts *MR* and *R&DI*, and the coefficient associated to it is the coefficient of interest. *PFL Terms* is a vector of the public float terms (polynomial of degree 3)[16], with *PFL* being the public float reported at the event fiscal year (2004). *Y* is the dependent variable and *X* is a vector of industry and company level controls, both specified below[17, 18].

Limiting the analysis to observations that lie within the neighbourhood of the threshold is appealing because enhances the likelihood that assignment to treatment is (as good as) random.

[Figure 1 around here]

While the compliance rule implies a clear discontinuity in the probability of submitting a management report at the threshold of $75 million, internal validity also requires that the threshold and *PFL* are determined independently of each other. In this particular empirical context, this condition may be jeopardized because some companies may manipulate *PFL* after knowing the threshold, choosing the most favourable side of the threshold. Figure 1 shows evidence that manipulation may be in

[14] See Iliev (2010) for a more detailed description of this context.
[15] The implementation of a regression discontinuity design in this context is not new. Iliev (2010) and Albuquerque and Zhu (WP2014) use similar identification strategies.
[16] To control for any direct effects of public float on the outcome variables.
[17] Section "Variables and descriptive statistics".
[18] We also report results for a difference-in-differences approach.

place. In the event year (2004) there is a significant jump in the density at the threshold, when compared to years 2002 and 2003. Some companies seem to be evading filling a management report by manipulating their public float[19]. In this case, it is reasonable to suspect that the OLS will be biased (not capturing the true effect of rule compliance) because compliers and non-compliers could differ in unobserved factors that would impact *Y*.

We use an instrumental variables approach to overcome this empirical challenge, using the following two-stage model.

$$MR = \delta_0 + \delta_1 . PFL752002 + \delta_2 . R\&DI + \delta_3 . MR . R\&DI + \gamma . X + \nu$$

$$Y = \beta_0 + \beta_1 . \widehat{MR} + \beta_2 . R\&DI + \beta_3 . MR . R\&DI + \alpha . X + \varepsilon$$

MR is a dummy variable that identifies companies that filed a management report. *PFL752002* is a dummy variable that identifies companies that reported a public float higher than $75 million in the fiscal year 2002. *R&DI* is a measure of R&D intensity, specified below[20]. *MR.R&DI* is a term that interacts *MR* and *R&DI*, and the coefficient associated to it on the second stage regression is the coefficient of interest. *Y* is the dependent variable and *X* is a vector of controls, both specified below[21].

The validity of the instrument is assured because: *PFL752002* strongly predicts MR by design of the rule – companies that reported a public float higher than $75 million in the fiscal year 2002 immediately become accelerated fillers and had to submit a management report in the fiscal year 2004; and, *PFL752002* has no direct effect on *Y* – it is unlikely that this dummy variable would impact companies' long-term investment outcomes (like CAPEX and R&D), particularly in the neighbourhood of $75 million; *PFL752002* is not subject to manipulation – the rule was defined after companies defined the public float of 2002[22].

An accelerated filler status implies two simultaneous treatments: file a management report and to file the 10-k report sooner. We assume that this last difference has no

[19] Companies can do so by increasing large shareholders and/or management holdings, and/or by decreasing the value of equity outstanding (through share repurchase) [Gao *et al.* (2009), Iliev (2010), and Nondorf *et al.* (2012)]. Evidence of manipulation was already found by Iliev (2010). Yet, we provide further evidence on its drivers on section "Results".

[20] Section "Variables and descriptive statistics".

[21] Section "Variables and descriptive statistics".

[22] While SEC introduced accelerated filler status in December 15, 2002, the rule of compliance with a management report was published in June 2003. Public float in year 2002 is the public float observed in their second fiscal quarter-end.

impact on companies' long-term investment outcomes (like CAPEX and R&D). Lastly, it must hold that companies are following the rule. Iliev (2010) provides strong evidence that they do[23].

Sample

Table 1 describes the sample generating process of the main (long-term) analysis sample. We depart from a database developed by Iliev (2010) that comprises a sample of 1492 non-financial U.S. incorporated companies with market equity between $30 and $330 million[24]. We dropped 653 companies because they do not fulfil our eligibility criteria in terms of R&D and CAPEX – our main long-term investment outcomes. In particular, we collect annual data on Compustat of both R&D and CAPEX between January 1 2002 and December 31 2010, and keep 839 companies that reported R&D and CAPEX both before and after the event year (2004)[25]. Next we excluded 627 companies that did not fulfil our bandwidth criterion, and keep 212 companies that reported public float between $50 and $100 million in the event year (2004). Next, we dropped 29 companies for which the reported a public float in 2004 did not correspond to their second fiscal quarter-end, and 1 firm that could be considered an outlier in terms of the assets reported in 2004[26]. After employing these further restrictions, we keep 182 companies on our main (long-term analysis) sample.

Some robustness tests involve 2 alternative measures of the dependent variables of interest, each of which potentially corresponding to 2 different samples: the short-term analysis sample, and the alternative long-term analysis sample. As we explain below, our short-term analysis sample has further restrictions related with the operationalization of the dependent variables. This implies dropping only 1 additional company from our main (long-term analysis) sample. Our alternative long-term analysis sample has fewer restrictions related with the operationalization of the dependent variables. Yet, this has no implications on sample size.

---

[23] As explained in the data section, we depart from database developed by Iliev (2010), and focus on a subset of the sample where he tests rule implementation. As such, evidence could be even higher in our case.

[24] Iliev (2010) uses Compustat for market capitalization and accounting data and collects data on public float and management reports from companies' annual 10-k fillings. This sample of 1492 companies excludes financial companies because these have already similar regulation in place.

[25] This is further explained below (section: variables and descriptive statistics).

[26] While low public float is supposed to be associated with small firms, we add this further constraint to foster the precision on identifying small firms. In any case, results do not change significantly if this company is included.

[Table 1 around here]

[Table 2 around here]

Variables and descriptive statistics

Table 2 describes the variables of the main (long-term) analysis sample. Dependent variables are outcomes of the long-term investment measured after (and including) the event year (2004), namely: (1) *CAPEX plus R&D after*, (2) *CAPEX after*, and (3) *R&D after*[27]. We measure R&D intensity as the ratio *R&D Before* to assets reported in event year (2004)[28]. We control for the following company level characteristics: (8) *Assets* to account for potentially remaining differences in size; (9) *Leverage Ratio* to account for possible effects of financing decisions on long-term investment decisions; and, to account for persistent factors that affect long-term investment decisions, we include the long-term investment outcome corresponding to the dependent variable measured before (and excluding) the event year (2004), namely: (4) *CAPEX plus R&D before*, (5) *CAPEX before*, and (6) *R&D before*.

[Table 3 around here]

Table 3 presents descriptive statistics for these variables. On average, R&D is higher than CAPEX both before and after. The basic descriptive statistics of *Assets* further justifies the reason why we dropped one firm with assets higher than $2000 million. Moreover, there is a significant negative correlation between R&D and CAPEX, both before and after, and significant positive lagged correlation for each. Comparing the means between MR and non-MR companies is further informative. The differences in means of most outcomes of the long-term investment are statistically significant, both before and after, further justifying controlling for pre-event long-term investment

[27] In line with the literature, we define long-term investment as expenses designed to generate benefits over the long-term and, as such, include in the analysis both capital expenditures (CAPEX) and research and development expenses (R&D). It is worth noting that *CAPEX* and *R&D* are usually subject to a distinct accounting treatment – typically, CAPEX are capitalized (thus, an asset is created when expensed), and R&D is expensed as incurred (thus, no asset is created when expensed). The argument behind this distinction is that the benefits from R&D are highly uncertain, when compared to CAPEX.

[28] While this may be a relative consensual measure of R&D intensity, we try alternative measures for which results are presented on section "Results".

outcome corresponding to the dependent variable. Figure 2 further explores the differences in means between groups of dependent variables and highlights no evident changing pattern around the event year (2004).

[Figure 2 around here]

We are aware of the problems associated to missing data. In our case, this is a particular sensitive issue on R&D data. We do have a lot of missing data for R&D between January 1, 2002, and December 31, 2010[29]. We decided restrict the analysis to non-missing data observations[30], and thus our analysis is conditional on reporting at least one non-missing value for R&D and CAPEX before and after the event year (2004).

Results

Estimation results for main long-term analysis are provided on Tables 4 and 5. If impact of SOX 404 on non-innovators (R&D intensity equals zero) is captured by the coefficient associated to *MR*, IV estimations suggest that SOX Section 404 impacts negatively on total investment (*CAPEX plus R&D*), with no positive impact on *R&D*. Moreover, the differences to OLS suggest that non-innovative companies that expected a significant negative impact on total investment evaded regulation[31].

The coefficient associated to $MR.R\&DI$ is positively and statistically significant on all IV estimations, supporting our hypothesis. Taken together, this suggests that the impact of SOX Section 404 is contingent on the level of R&D intensity, and that the overall impact on companies' long-term investment can vary from negative to positive. As evidenced in Figure 3, our IV results predict a negative impact of SOX 404 for more than 80% of small innovative companies in the sample (those with lower level of R&D intensity).

[Table 4 around here]

[29] For a discussion on missing R&D data see Koh and Reeb (2015).
[30] As opposed to other empirical alternatives, such as: replacing missing R&D with zero; replacing R&D industry average; or, replacing R&D with some historical value from prior years.
[31] Results of cross section OLS are not clearly robust to an alternative approach – Difference-in-differences – as reported in Table 11.

[Table 5 around here]

[Figure 3 around here]

Robustness tests

Our first set of robustness tests involve 2 alternative measures of the dependent variables of interest: [1] in the short-term analysis the dependent variables of interest are defined for a narrower time window, just including the event year (2004); and, [2] in the alternative long-term analysis the dependent variables of interest are defined for a broader time window, including data till December 31, 2010. As mentioned above, this imposes few restrictions on sample size, and these are confined to the short-term analysis. Tables 6 and 7 show that results are mostly insensitive to these alternative measures of the dependent variables.

[Table 6 around here]

[Table 7 around here]

We also find that results are robust if we alternatively measure *R&DI* by scaling *R&D Before* by sales (as opposed to assets) in the event year (2004)[32]. Table 8 reports results for this alternative.

[Table 8 around here]

To minimize the possibilities of omitted-variable bias, we further include the following controls. First, since it can be argued that investment depends on companies' profitability or growth opportunities, we include the variables *CFO* and *MTB*[33], corresponding to cash flow from operations and market to book (value of equity) ratio, measured in the event year (2004), respectively. Second, SEC guidance[34] suggests that larger and more complex companies have higher likelihood

[32] Scaling R&D by sales has been extensively used as a measure of R&D intensity – among others, Katila and Ahuja (2002).
[33] *MTB* is the inverted BE/ME (book value of equity divided by the market value of equity), as defined by Iliev (2010). *CFO* is scaled by previous-year total assets, also in line with Iliev (2010).
[34] SEC Release 33-8810 and 34-55929; June 27, 2007.

of reporting financial misstatements. To the extent that this may impact direct and indirect compliance costs, it may impact our key dependent variables. While we focus on small companies and still control for size, in order to control for complexity, we include the variables *Business Segments* and *Geographic Segments*[35], corresponding to the number of business segments and the number of geographic segments, respectively. Results are mostly insensitive to the inclusion of these additional controls.

[Table 9 around here]

We assess whether results are sensitive to bandwidth definition. As reported in Table 10, results are robust to a narrower bandwidth and not to a wider bandwidth. We attribute this last result to the fact that an increase in the bandwidth may be associated with an increase in bias from unobservable factors that might be correlated with both dependent and independent variables.

[Table 10 around here]

Finally, we assess whether results are sensitive to the instrument used (*PFL2002* – public float reported in 2002 – instead of *PFL752002* – a dummy variable that identifies companies that reported a public float higher than $75 million in the fiscal year 2002). Results are robust[36].

[35] This is also in line with Iliev (2010).
[36] These are enabled upon request.

DISCUSSION

The purpose of this study was to examine the impact of SOX Section 404 on long-term investment of small innovative companies. In particular, we address the research question of whether R&D intensity increases the impact of SOX 404 on long-term investment. Combining Iliev (2010) database with further Compustat data, our main results support the hypothesis and suggest that the impact of SOX Section 404 on companies' long-term investment can vary from negative to positive, positively depending on the level of R&D intensity.

We think this is compelling evidence for several reasons. First, it suggests that both arguments – asymmetric information and costs of compliance – are in play. Second, it suggests that companies subject to coercive financial disclosure are forced to "play different games", and that R&D intensive companies are relatively benefited. Third, it suggests that the argument 'costs of compliance' dominates the argument 'asymmetric information' for nearly 90% of our sample of small innovative companies – challenging the findings that compliance is (on average) positively associated with corporate investment [Albuquerque and Zhu (WP2014)]. Finally, evidence of an "uneven game" suggests a re-centring of policy discussion around the distribution of the net benefit of coercive financial disclosure programs *versus* the overall economic impact of those programs.

This motivates further research as, for instance, other dimensions of information asymmetry could be relevant. Furthermore, future research could focus on including acquisitions (as alternative investment mechanism), alternative empirical strategies to deal with missing values on R&D, address investment efficiency[37], and disentangle accounting from investment decisions.

[37] Biddle and Hilary (2006); and, Biddle *et al.* (2009).

REFERENCES

Aboody, D.; and, Lev, B. 2000. Information Asymmetry, R&D, and Insider Gains. The Journal of Finance. 55(6): 2747-2766.

Albuquerque, A.; and, Zhu, J. 2014. Has Section 404 of the Sarbanes-Oxley Act Discourage Corporate Investment? new evidence from a natural experiment. Working paper.

Alexander, C.; Bauguess, S.; Berline, G.; Lee, Y.; and Marietta-Westberg, J. 2013. Economic Effects of SOX Section 404 Compliance: a corporate insider perspective. Journal of Accounting and Economics. 56: 267-290.

Andrade, S.; Bernile, G.; and, Hood III, F. 2014. SOX, Corporate Transparency, and the Cost of Debt. Journal of Banking and Finance. 38: 145-165.

Arping, S.; and, Sautner, Z. 2013. Did SOX Section 404 Make Firms Less Opaque? evidence from cross-listed firms. Contemporary Accounting Research. 30(3): 1133-1165.

Ashbaugh-Skaife, H.; Collins, D.; Kinney JR., W.; and, Lafond, R. 2009. The Effect of SOX Internal Control Deficiencies on Firm Risk and Cost of Equity. Journal of Accounting Research. 47(1): 1-43.

Bargeron, L.; Lehn, K.; and, Zutter, C. 2010. Sarbanes-Oxley and Corporate Risk-Taking. Journal of Accounting and Economics. 49: 34-52.

Barry, C.; and, Brown, S. 1985. Differential Information and Security Market Equilibrium. The Journal of Financial and Quantitative Analysis. 20(4): 407-422.

Bens, D.; and, Monahan, S. 2004. Disclosure Quality and Excess Values of Diversification. Journal of Accounting Research. 42(4): 691-730.

Biddle, G.; and, Hilary, G. 2006. Accounting Quality and Firm-Level Capital Investment. The Accounting Review. 81(5): 963-982.

Biddle, G.; Hilary, G.; and, Verdi, R. 2009. How Does Financial Reporting Quality Relate to Investment Efficiency? Journal of Accounting and Economics. 48(2-3): 112-131.

Botosan, C. 1997. Disclosure Level and the Cost of Equity Capital. The Accounting Review. 72(3): 323-349.

Botosan, C.; and, Plumlee, M. 2002. A Re-examination of Disclosure Level and the Expected Cost of Equity Capital. Journal of Accounting Research. 40(1): 21-40.

Bushman, R.; and, Smith, A. 2001. Financial Accounting Information and Corporate Governance. Journal of Accounting and Economics. 32: 237-333.

Coates, J. 2007. The Goals and Promise of the Sarbanes–Oxley Act. Journal of Economic Perspectives. 21(1): 91-116.

Coates, J.; and, Srinivasan, S. 2014. SOX After Ten Years: a multidisciplinary review. Accounting Horizons. 28(3): 627-671.

Diamond, D.; and, Verrecchia, R. 1991. Disclosure, Liquidity, and the Cost of Capital. The Journal of Finance. 46(4): 1325-1359.

Gao, F.; Wu, J.; and, Zimmerman, J. 2009. Unintended Consequences of Granting Small Firms Exemptions from Securities Regulation: Evidence from the Sarbanes-Oxley Act. Journal of Accounting Research. 47(2): 459-506.

Healy, P.; Huttom, A.; and, Palepu, K. 1999. Stock Performance and Intermediation Changes Surrounding Sustained Increases in Disclosure. Contemporary Accounting Research. 16(3): 485-520.

Healy, P.; and, Palepu, K. 2001. Information Asymmetry, Corporate Disclosure, and the Capital Markets: a review of the empirical disclosure literature. Journal of Accounting and Economics 31(1-3): 405-440.

Iliev, P. 2010. The Effect of SOX Section 404: costs, earnings quality, and stock prices. Journal of Finance. 65(3): 1163-1196.

Jensen, M.; and, Meckling, W. 1976. Theory of the Firm: managerial behavior, agency costs and ownership structure. Journal of Financial Economics. 3(4): 305-360.

Kang, Q.; Liu, Q.; and, Qi, R. 2010. The Sarbanes-Oxley Act and Corporate Investment: a structural assessment. Journal of Financial Economics. 96: 291-305.

Katila, R.; and, Ahuja, G. 2002. Something Old, Something New: a longitudinal study of search behaviour and new product introduction. Academy of Management Journal. 45(6): 1183-1194.

Koh, P-S.; and, Reeb, D. 2015. Missing R&D. Journal of Accounting and Economics. 60(1): 73-94.

Lambert, R.; Leuz, C.; and, Verrecchia, R. 2007. Accounting Information, Disclosure, and the Cost of Capital. Journal of Accounting Research. 45: 385-420.

Leuz, C.; and, Verrecchia, R. 2000. The Economic Consequences of Increased Disclosure. Journal of Accounting Research. 38: 91-124.

Nondorf, M.; Singer, Z.; and, You, H. 2012. A Study of Firms Surrounding the Threshold of Sarbanes–Oxley Section 404 Compliance. Advances in Accounting. 28(1): 96-110.

Singer, Z; and, You, H. 2011. The Effect of Section 404 of Sarbanes-Oxley Act on Earnings Quality. Journal of Accounting, Auditing and Finance. 26(3): 556-589.

Smith, C.; and, Warner, J. 1979. On Financial Contracts: an analysis of bond covenants. Journal of Financial Economics. 7: 117-161.

Welker, M. 1995. Disclosure Policy, Information Asymmetry, and Liquidity in Equity Markets. Contemporary Accounting Research. 11(2): 801-827.

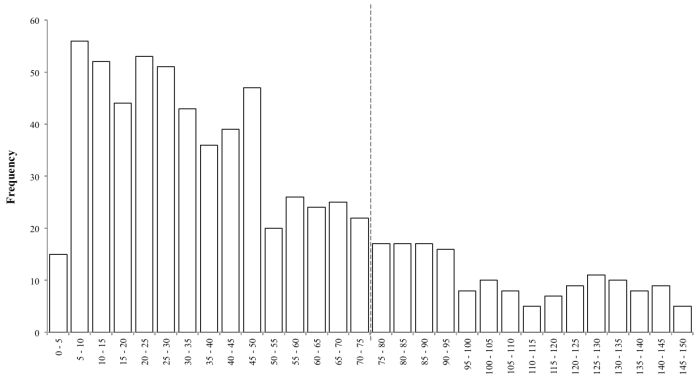
Zhang, I. 2007. Economic Consequences of the Sarbanes-Oxley Act of 2002. Journal of Accounting and Economics. 44: 74-115.

APPENDIX A – main figures and tables

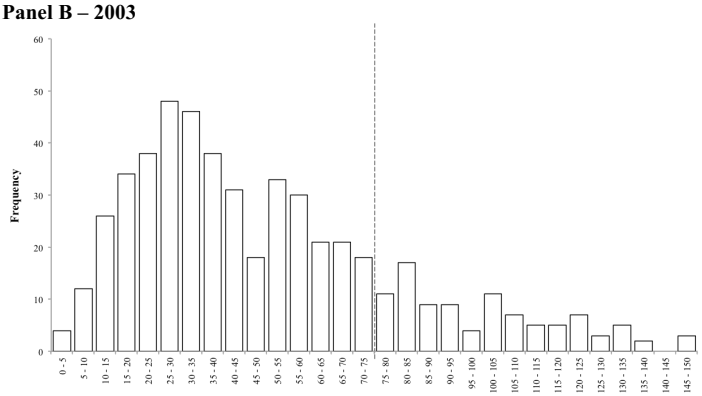
Figure 1 – Evidence of PFL manipulation

The figure shows the histogram for PFL till $150 million for firms with PFL lower than $75 million in years before. Firms included are eligible firms in terms of R&D and CAPEX. In general, these are firms for which we observe both R&D and CAPEX before and after the event year (2004). The eligibility criterion is defined on data section in more detail. Panel A has 710 observations, Panel B has 516 observations, and panel C has 390 observations. Companies above the threshold in Panel A are not part of Panel B, as they become accelerated filers in year 2002. Companies above the threshold in Panel B are not part of Panel C, as they become accelerated filers in that 2003. Fiscal year 2002 comprises reports from November 2002 to October 2003. Fiscal year 2003 comprises reports from November 2003 to October 2004. Fiscal year 2004 comprises reports from November 2004 to October 2005.

Panel A – 2002



Panel B – 2003



Panel C – 2004

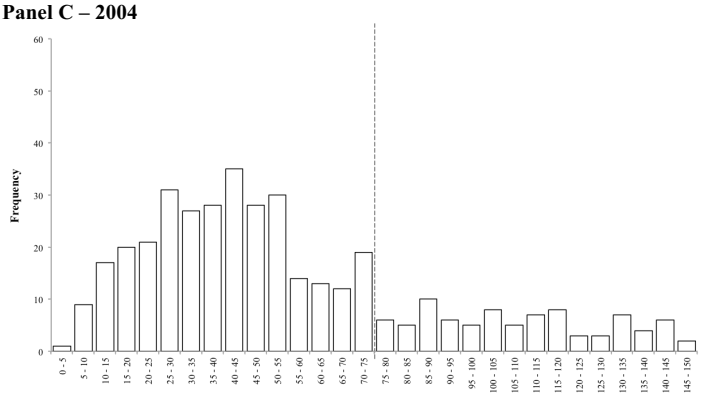


Table 1 – Sample generating process

The Table describes the sample generating process of the main (long-term) analysis sample. We depart from a database developed by Iliev (2010) that comprises a sample of 1492 non-financial U.S. incorporated companies with market equity between $30 and $330 million. After employing further restrictions, we keep 182 companies on our main (long-term analysis) sample. Our short-term analysis sample has further restrictions related with the operationalization of the dependent variables. This implies dropping only 1 additional company from our main (long-term analysis) sample. Our alternative long-term analysis sample has fewer restrictions related with the operationalization of the dependent variables. Yet, this has no implications on sample size.

Sample Description	Dropped firms	Kept firms
Non-financial with U.S. incorporation companies with market equity between $30 and $330 million	-	1492
Companies eligible in terms of R&D and CAPEX	653	839
Companies with reported public float in 2004 between $50 and $100 million	627	212
Companies with reliable reported public float in 2004	29	183
Companies with assets in 2004 lower than $2000 million	1	182

Table 2 – Variable definitions

The Table describes the variables of the main (long-term) analysis sample. Our short-term analysis sample has further restrictions related with the operationalization of the dependent variables (1) to (4). The time window is narrower and just includes the event year (2004). Our alternative long-term analysis sample has fewer restrictions related with the operationalization of the dependent variables (1) to (4). The time window is broader and includes data till December 31, 2010. CAPEX are the total funds used for additions to property, plant and equipment, excluding those associated with acquisitions (of companies).

	Designation	Description
1	CAPEX plus R&D after	Average of the sum of CAPEX and R&D reported between the event year (2004) and December 31, 2006, including the event year, considering only those observations for which there is non-missing values for both R&D and CAPEX.
2	CAPEX after	Average CAPEX reported between the event year (2004) and December 31, 2006, including the event year, considering only those observations for which there is non-missing value for CAPEX.
3	R&D after	Average R&D reported between the event year (2004) and December 31, 2006, including the event year, considering only those observations for which there is non-missing value for R&D.
4	CAPEX plus R&D before	Average of the sum of CAPEX and R&D reported between January 1, 2002 and the event year (2004), excluding the event year, considering only those observations for which there is non-missing values for both R&D and CAPEX.
5	CAPEX before	Average CAPEX reported between January 1, 2002 and the event year (2004), excluding the event year, considering only those observations for which there is non-missing value for CAPEX.
6	R&D before	Average R&D reported between January 1, 2002 and the event year (2004), excluding the event year, considering only those observations for which there is non-missing value for R&D.
7	R&DI	Equals [6]/[8].
8	Assets	Assets in the event year (2004).
9	Leverage Ratio	Leverage Ratio in the event year (2004).

Table 3 – Descriptive Statistics

The Table presents descriptive statistics for each variable of interest and controls, for the main (long-term analysis) sample. Panel A presents basic descriptive statistics. Panel B presents a correlation matrix. Panel C presents a t-test to the differences in means between MR and non-MR companies. *P<10%; **P<5%; ***P<1%.

Panel A – basic

	Designation	Obs.	Mean	S.D.	Min	Max
1	CAPEX plus R&D after	182	11.6	9.17	0.35	50.73
2	CAPEX after	182	4.24	6.91	0.02	50.08
3	R&D after	182	7.38	8.21	0	48.81
4	CAPEX plus R&D before	182	11.3	9.21	0.09	44.20
5	CAPEX before	182	3.89	5.94	0	37.57
6	R&D before	182	7.41	8.42	0	40.36
7	R&DI	182	0.15	0.20	0	0.975
8	Assets	182	112	107	2.86	645.7
9	Leverage Ratio	182	0.10	0.16	0	0.892

Panel B – correlations

	1	2	3	4	5	6	7	8
2	0.51***							
3	0.69***	-0.28***						
4	0.69***	0.23***	0.58***					
5	0.33***	0.73***	-0.24***	0.45***				
6	0.52***	-0.27***	0.81***	0.78***	-0.21***			
7	0.18**	-0.34***	0.49***	0.36***	-0.31***	0.60***		
8	0.26***	0.60***	-0.22***	0.22***	0.62***	-0.20***	-0.44***	
9	0.11	0.45***	-0.25***	0.13*	0.49***	-0.20***	-0.22***	0.65***

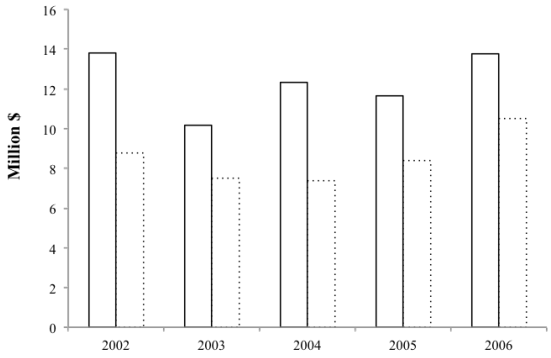
Panel C – between group means differences

	Obs.	1	2	3	4	5	6	7	8	9
Non-MR	72	8.52	4.34	4.21	8.04	3.42	4.62	0.13	95.7	0.10
MR	110	13.6	4.18	9.45	13.4	4.20	9.23	0.16	124	0.10
Difference		-5.12***	0.16	-5.25***	-5.40***	-0.79	-4.61***	-0.04	-28.2*	0.00

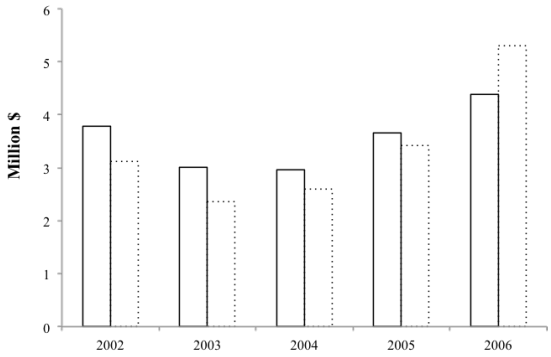
Figure 2 – Trends in annual means – differences between groups

The Figure presents the annual means of the variables of interest, around the event year (2004). Note that in each panel the observations are those available for the 182 companies in the main (long-term analysis) sample, for those companies that have non-missing values for all years (2002 to 2006). In black, MR companies. In white, non-MR companies.

Panel A – CAPEX plus R&D



Panel B – CAPEX



Panel C – R&D

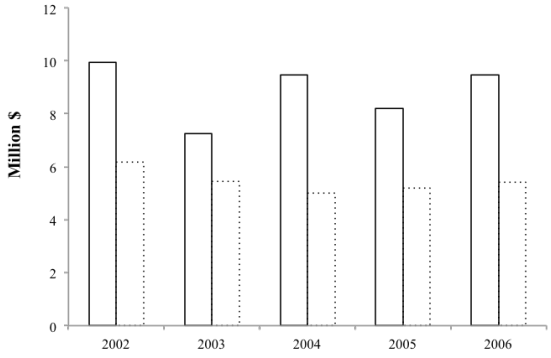


Table 4 – OLS results – focal dependent variables – main long-term analysis

The table presents results for the focal dependent variables from OLS approach. *Ybefore* is the corresponding dependent variable measured before (and excluding) the event year (2004). *MR* is a dummy variable equal to one if the company filed a management report in 2004. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel A – OLS and second-stage IV

Specification	**Without *MR.R&DI***			**With *MR.R&DI***		
Dependent Variable	**CAPEX**	**R&D**	**CAPEX plus R&D**	**CAPEX**	**R&D**	**CAPEX plus R&D**
MR	-1.919587** (0.7488118)	1.36252 (0.9832975)	-0.2967501 (1.339995)	-2.571218** (1.044792)	0.2839871 (1.010195)	-2.060995 (1.573386)
R&DI	-0.9874827 (1.254374)	-2.320417 (3.784041)	-2.173618 (4.296004)	-3.269669** (1.604164)	-6.052904* (3.2253)	-8.33218** (3.924891)
MR.R&DI				**3.969426 (2.624617)**	**6.802372* (3.754547)**	**10.89552** (4.824401)**
Ybefore	0.6591339** (0.2711632)	0.7180763*** (0.1028334)	0.6715246*** (0.1128551)	0.66385** (0.2701519)	0.7111035*** (0.1001953)	0.6677082*** (0.1086846)
Assets	0.0111994 (0.0086446)	0.0004986 (0.0037967)	0.0128465 (0.0086213)	0.0112485 (0.0087015)	0.0009008 (0.0035689)	0.0133959 (0.0085876)
Leverage ratio	0.4829434 (3.49564)	-3.529772 (2.536369)	-3.538085 (4.480416)	-0.7751547 (3.520206)	-3.044121 (2.429979)	-2.671423 (4.462904)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
PFL terms	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.6257	0.6930	0.5257	0.6266	0.6993	0.5407
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	182	182	182	182	182	182

Table 5 – IV results – focal dependent variables – main long-term analysis

The table presents results for the focal dependent variables from IV approach. *Ybefore* is the corresponding dependent variable measured before (and excluding) the event year (2004). *MR* is the predicted treatment based on the first-stage regression. The IV estimation uses the instrument *PFL752002*, a dummy equal to one if the company public float was above $75 million in 2002. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel A – second-stage IV

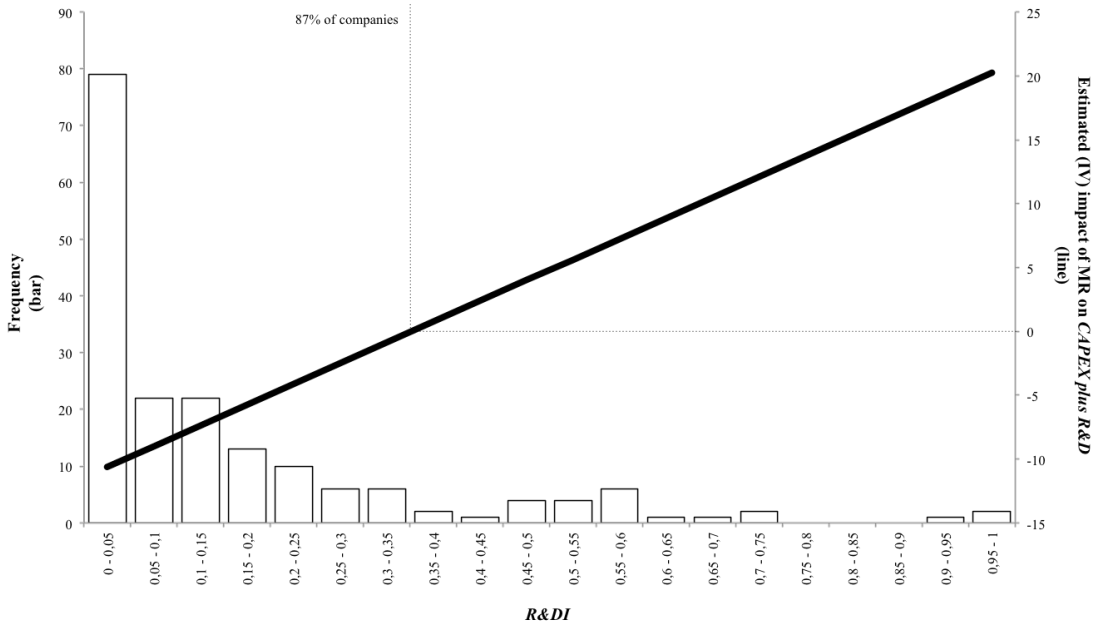
Specification	**Without *MR.R&DI***			**With *MR.R&DI***		
Dependent Variable	**CAPEX**	**R&D**	**CAPEX plus R&D**	**CAPEX**	**R&D**	**CAPEX plus R&D**
MR	-5.366473*** (1.768703)	-0.8948996 (2.07621)	-6.574427** (3.061102)	-7.745526*** (2.502018)	-2.680727 (2.79754)	-10.61478** (4.170464)
R&DI	0.1979409 (1.814472)	-2.555396 (4.076757)	-2.571936 (5.054919)	-10.00792*** (3.611927)	-10.51793* (5.415004)	-20.94136*** (7.704745)
MR.R&DI				**17.43365*** (5.806039)**	**14.66095** (6.731214)**	**32.4624*** (9.973644)**
Ybefore	0.7025287*** (0.2500443)	0.7800878*** (0.1077383)	0.7907841*** (0.1218928)	0.7253993*** (0.2436506)	0.7547796*** (0.0996082)	0.7635757*** (0.1103086)
Assets	0.0172037* (0.009921)	0.0036345 (0.0047891)	0.0195743* (0.0108025)	0.0172159* (0.0103757)	0.0044105 (0.0050411)	0.0211783* (0.011795)
Leverage ratio	-2.031948 (3.665405)	-4.607087 (2.995848)	-7.165451 (5.229458)	-0.8599279 (3.852823)	-3.6685 (2.759732)	-4.694465 (5.22707)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
PFL terms	No	No	No	No	No	No
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.5371	0.6621	0.3752	0.5080	0.6642	0.2589
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	182	182	182	182	182	182

Panel B – first-stage IV

Dependent Variable	**MR**			**MR**		
PFL752002	0.4711756*** (0.0779356)	0.4280121*** (0.0767117)	0.4254543*** (0.0783202)	0.3424738*** (0.0671679)	0.3275424*** (0.0664091)	0.322032*** (0.0673962)
R&DI	Yes	Yes	Yes	Yes	Yes	Yes
MR.R&DI	No	No	No	Yes	Yes	Yes
Ybefore	Yes	Yes	Yes	Yes	Yes	Yes
Assets	Yes	Yes	Yes	Yes	Yes	Yes
Leverage ratio	Yes	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.2007	0.2340	0.2229	0.4370	0.4462	0.4446
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	182	182	182	182	182	182

Figure 3 – Estimated impact of SOX Section 404 on long-term investment by levels of R&D intensity

The figure shows the estimated (IV) impact of *MR* on *CAPEX plus R&D* by levels of *R&DI*.



APPENDIX B – secondary figures and tables

Table 6 – Robustness to dependent variables measurements – short-term analysis

The Table is analogous to Tables 4 and 5 (full models), except that dependent variables of interest are defined for a narrower time window, just including the event year (2004). Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel A – OLS and second-stage IV

Specification	**OLS**			**IV**		
Dependent Variable	**CAPEX**	**R&D**	**CAPEX plus R&D**	**CAPEX**	**R&D**	**CAPEX plus R&D**
MR	-1.714883 (1.128775)	1.219578 (1.000276)	-0.475052 (1.737109)	-5.608355** (2.25623)	-4.281157 (4.000095)	-10.05637** (4.748717)
R&DI	-2.090817 (1.328782)	-5.900201 (4.26102)	-8.012073* (4.6084)	-6.714954** (2.758372)	-13.77485 (9.708319)	-21.45171** (10.53637)
MR.R&DI	**3.624537 (2.948552)**	**3.043189 (4.664291)**	**6.460598 (5.485813)**	**12.97723** (5.355377)**	**16.19965* (9.042728)**	**29.13403*** (11.06776)**
Ybefore	0.8248169*** (0.2884465)	0.7266561*** (0.0919208)	0.7341681*** (0.1016799)	0.85998*** (0.2723585)	0.797714*** (0.110592)	0.8366288*** (0.1094686)
Assets	0.0090494 (0.0092123)	0.0064756 (0.0065018)	0.0172344 (0.0117975)	0.0127466 (0.0099354)	0.0109273 (0.0103609)	0.0231972 (0.0151806)
Leverage ratio	-0.2817447 (2.771223)	-5.853113 (4.15033)	-5.566074 (5.637741)	-0.9715009 (3.247226)	-6.097645 (4.782441)	-6.614037 (6.707793)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
PFL terms	Yes	Yes	Yes	No	No	No
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.7264	0.4351	0.4224	0.6709	0.3832	0.2969
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	181	181	181	181	181	181

Panel B – first-stage IV

Dependent Variable	**MR**		
PFL752002	0.3466442*** (0.0678704)	0.3297335*** (0.0671753)	0.3241937*** (0.068369)
Adjusted R^2	0.4355	0.4441	0.4425
Prob>F	0.0000	0.0000	0.0000
Observations	181	181	181

Table 7 – Robustness to dependent variables measurements – alternative long-term analysis

The Table is analogous to Tables 4 and 5 (full models), except that dependent variables of interest are defined for a broader time window, including data till December 31, 2010. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel A – OLS and second-stage IV

Specification	**OLS**			**IV**		
Dependent Variable	**CAPEX**	**R&D**	**CAPEX plus R&D**	**CAPEX**	**R&D**	**CAPEX plus R&D**
MR	-2.512081** (1.081836)	0.6768071 (1.179741)	-1.543142 (1.745173)	-7.944831** (3.228988)	-2.273438 (3.051052)	-10.53774** (5.049432)
R&DI	-2.214665 (1.675263)	-7.543074* (4.372064)	-8.086121 (5.078583)	-9.503849** (4.227326)	-11.65693* (6.059563)	-21.32298** (8.917034)
MR.R&DI	**2.394322 (2.875263)**	**8.436142* (4.372064)**	**11.20483** (5.574837)**	**17.3808** (7.257576)**	**15.93995** (7.928953)**	**34.45413*** (12.42302)**
Ybefore	0.5723** (0.2763398)	0.6706381*** (0.1116586)	0.589794*** (0.1252033)	0.636757*** (0.2402311)	0.7219223*** (0.112644)	0.6968344*** (0.1267657)
Assets	0.0187573** (0.0090279)	0.0028403 (0.0042904)	0.0215546** (0.0104075)	0.0257522** (0.0116357)	0.0060318 (0.0053384)	0.0301021** (0.0142055)
Leverage ratio	-0.0225027 (4.412029)	-4.445079* (2.679626)	-4.664553 (5.463693)	-2.086417 (4.158034)	-5.013966* (2.952468)	-7.145398 (5.841884)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
PFL terms	Yes	Yes	Yes	No	No	No
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.5366	0.6812	0.4491	0.4189	0.6450	0.2589
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	182	182	182	182	182	182

Panel B – first-stage IV

Dependent Variable	**MR**		
PFL752002	0.3424738*** (0.0671679)	0.3275424*** (0.0664091)	0.322032*** (0.0673962)
Adjusted R^2	0.4370	0.4462	0.4446
Prob>F	0.0000	0.0000	0.0000
Observations	182	182	182

Table 8 – Robustness to *R&DI* measurement

The Table is analogous to Tables 4 and 5 (full models), except that *R&DI* variable is defined by scaling *R&D Before* by the logarithm of sales (as opposed to assets) in the event year (2004), and that panel C presents the basic descriptive statistics for this alternative measure of *R&DI*. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel A – OLS and second-stage IV

Specification	**OLS**			**IV**		
Dependent Variable	**CAPEX**	**R&D**	**CAPEX plus R&D**	**CAPEX**	**R&D**	**CAPEX plus R&D**
MR	-2.008895*** (0.7528934)	1.224948 (0.9609053)	-0.5414143 (1.360538)	-5.678583*** (1.891418)	-1.651561 (2.04735)	-7.626845** (3.198813)
R&DI	-0.0033595 (0.0026933)	-0.0076751 (0.0049753)	-0.0103154* (0.0054383)	-0.0149743** (0.0059146)	-0.0163296* (0.0090891)	-0.0325226** (0.0131312)
MR.R&DI	**0.0033574 (0.0026932)**	**0.0079929 (0.0049066)**	**0.0106443** (0.0053475)**	**0.0150363** (0.0059305)**	**0.0166355* (0.009043)**	**0.032879** (0.0130973)**
Ybefore	0.6583723** (0.270825)	0.6711602*** (0.0791023)	0.6386621*** (0.0929009)	0.7059084*** (0.2501321)	0.7417963*** (0.084361)	0.7661849*** (0.1067422)
Assets	0.0118894 (0.0082506)	0.0029635 (0.0050039)	0.0157595** (0.0079131)	0.0171917* (0.0094157)	0.0068355 (0.0061251)	0.0232795** (0.0105748)
Leverage ratio	0.336501 (3.516726)	-4.48779 (2.892175)	-4.210014 (4.6448)	-2.102607 (3.693982)	-5.693112 (3.466856)	-8.102108 (5.568499)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
PFL terms	Yes	Yes	Yes	No	No	No
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.6254	0.7009	0.5351	0.5288	0.6576	0.3520
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	182	182	182	182	182	182

Panel B – first-stage IV

Dependent Variable			**MR**		
PFL752002			0.4652867*** (0.0800831)	0.4276079*** (0.0769469)	0.4176259*** (0.078441)
Adjusted R^2			0.1973	0.2497	0.2370
Prob>F			0.0000	0.0000	0.0000
Observations			182	182	182

Panel C – basic descriptive statistics

Variable	**Obs.**	**Mean**	**S.D.**	**Min**	**Max**
R&DI	182	316.08	2581	0	29780

Table 9 – Robustness to the inclusion of additional controls

The Table is analogous to Tables 4 and 5 (full models), except that we further include variables to control for profitability, growth opportunities and business complexity. Additional controls are *CFO* (cash flow from operations scaled by assets), *MTB* (market to book ratio), the number of *Business Segments,* and the number of *Geographic Segments*. Panel C presents the basic descriptive statistics for these. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel A – OLS and second-stage IV

Specification	**OLS**			**IV**		
Dependent Variable	**CAPEX**	**R&D**	**CAPEX plus R&D**	**CAPEX**	**R&D**	**CAPEX plus R&D**
MR	-2.67036** (1.220407)	0.3919551 (0.9644397)	-2.025048 (1.706321)	-8.756513*** (2.919434)	-3.352158 (2.792591)	-12.05719*** (4.423977)
R&DI	-0.8605219 (2.687508)	-8.937333*** (3.305788)	-8.813417* (4.716292)	-11.69649* (6.774934)	-15.63091** (7.189762)	-27.06882** (12.40522)
MR.R&DI	**5.885864* (3.288523)**	**5.046523 (4.337225)**	**11.52211* (5.919032)**	**23.68168*** (8.824243)**	**16.46557* (8.529793)**	**40.25562*** (14.05471)**
Ybefore	0.5974519** (0.2744172)	0.6634866*** (0.1023783)	0.6054523*** (0.1185285)	0.6713015*** (0.2453281)	0.696137*** (0.1000304)	0.6816677*** (0.1164059)
Assets	0.01228 (0.0094392)	-0.0003347 (0.0037231)	0.0135786 (0.0102622)	0.0182253 (0.0116178)	0.0040621 (0.0058974)	0.0227754 (0.0143239)
Leverage ratio	4.023682 (4.697983)	-3.188652 (2.450558)	0.2192605 (5.63225)	2.441921 (4.852349)	-3.599397 (3.010932)	-1.344798 (6.182996)
CFO	4.165073* (2.186671)	-5.986975*** (2.129769)	-1.753895 (2.977504)	4.251163* (2.542764)	-5.358844** (2.123236)	-1.148678 (3.463199)
MTB	0.0066135 (0.0042576)	-0.0062826 (0.0057079)	0.0011449 (0.0077257)	0.0149434* (0.0088669)	-0.0031835 (0.0074118)	0.011752 (0.0139395)
Business Segments	-0.0270439 (0.3772479)	-0.2220008 (0.25476)	-0.2577389 (0.449432)	-0.2701775 (0.3824782)	-0.3870018 (0.3589036)	-0.6554979 (0.5550219)
Geographic Segments	0.0539236 (0.2326069)	0.1257415 (0.1749766)	0.1775833 (0.2801413)	0.2515184 (0.3151542)	0.2207935 (0.2049075)	0.467539 (0.3992336)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
PFL terms	Yes	Yes	Yes	No	No	No
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.6273	0.6894	0.4984	0.4781	0.6372	0.2498
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	166	166	166	166	166	166

Panel B – first-stage IV

Dependent Variable	**MR**		
PFL752002	0.3248799*** (0.0696693)	0.321011*** (0.0692872)	0.3168994*** (0.0702857)
Adjusted R^2	0.4218	0.4233	0.4235
Prob>F	0.0000	0.0000	0.0000
Observations	166	166	166

Panel C – basic descriptive statistics

Variable	**Obs.**	**Mean**	**S.D.**	**Min**	**Max**
CFO	182	-0.06	0.25	-0.93	0.49
MTB	166	5.43	27.0	0.43	347
Business Segments	182	1.71	1.30	0	6
Geographic Segments	182	2.08	1.90	0	13

Table 10 – Robustness to bandwidth

The Table is analogous to Tables 4 and 5 (full models), except that we vary the bandwidth. Results from first stage are omitted but similar to those in Table 5. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel A – bandwidth $60-$90 million

Specification	**OLS**			**IV**		
Dependent Variable	**CAPEX**	**R&D**	**CAPEX plus R&D**	**CAPEX**	**R&D**	**CAPEX plus R&D**
MR	-1.184895 (0.8531355)	1.137816 (1.491564)	0.287624 (1.937238)	-4.348328* (2.399502)	-3.380399 (3.723287)	-7.847457* (4.418236)
R&DI	-3.15876* (1.884331)	-15.31037** (6.00316)	-16.96324*** (6.060289)	-9.060428** (4.201893)	-22.90996** (10.77138)	-32.68151*** (12.16221)
MR.R&DI	**2.408066 (2.246929)**	**8.837064 (5.453936)**	**11.32168* (6.222967)**	**11.59637* (6.0078)**	**18.52895** (8.905308)**	**30.32694*** (10.98777)**
R^2	0.7760	0.7105	0.6146	0.6959	0.6814	0.5127
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	101	101	101	101	101	101

Panel B - bandwidth $40-$110 million

Specification	**OLS**			**IV**		
Dependent Variable	**CAPEX**	**R&D**	**CAPEX plus R&D**	**CAPEX**	**R&D**	**CAPEX plus R&D**
MR	-2.390027*** (0.9081219)	-0.046005 (1.121429)	-2.02666 (1.377162)	-2.578843 (2.204858)	-0.1807708 (1.615853)	-2.125079 (2.870259)
R&DI	0.1251455 (0.8331279)	-0.1969165 (2.921784)	1.685397 (2.721629)	-0.8834892 (1.791488)	-1.223917 (3.438367)	-0.2143828 (4.042928)
MR.R&DI	**1.877226 (1.320484)**	**-0.5131986 (3.412144)**	**2.50115 (3.417048)**	**3.674965 (3.674965)**	**1.284521 (4.583412)**	**5.748323 (5.902166)**
R^2	0.6297	0.6851	0.5953	0.6162	0.6725	0.5682
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	270	270	270	270	270	270

Table 11 – Robustness to approach

The table presents results for the focal dependent variables, from Difference-in-Difference approach. Panel A and B present the results for model without and with the interaction of interest (*After.MR.R&DI*), respectively. *After* is a dummy variable equal to 1 if the observation corresponds to the period after 2004. Other variables, as before. Panel B further includes *After.R&DI* and *MR.R&DI*. Robust standard errors are reported in parenthesis. *P<10%; **P<5%; ***P<1%.

Panel A – without interaction of interest

Dependent Variable	CAPEX	R&D	CAPEX plus R&D
MR	-0.2522697 (0.6739047)	4.781163*** (1.143875)	4.522544*** (1.354938)
After	0.9201273 (0.7647868)	-0.4153657 (0.7462729)	0.4881574 (1.077373)
After.MR	-0.9427894 (0.9795549)	0.636343 (1.201523)	-0.2732604 (1.562165)
R&DI	0.0440708 (0.9626845)	20.18253*** (3.012181)	20.22729*** (3.290023)
Assets	Yes	Yes	Yes
Leverage ratio	Yes	Yes	Yes
Constant	Yes	Yes	Yes
PFL terms	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes
R^2	0.5237	0.4845	0.3337
Prob>F	0.0000	0.0000	0.0000
Observations	364	364	364

Panel A – with interaction of interest

Dependent Variable	CAPEX	R&D	CAPEX plus R&D
MR	-0.0969122 (1.510852)	4.048027*** (1.265419)	3.94383** (1.580155)
After	1.510852 (1.048324)	0.7708694 (0.8461273)	2.258748 (1.325773)
After.MR	-1.414688 (1.452196)	-0.0209551 (1.367195)	-1.396825 (1.985414)
After.MR.R&DI	**3.967808 (3.615835)**	**6.206486 (9.946164)**	**10.12821 (10.6534)**
R&DI	2.105953 (1.550578)	21.16886*** (7.332661)	23.25681*** (7.690653)
Assets	Yes	Yes	Yes
Leverage ratio	Yes	Yes	Yes
Constant	Yes	Yes	Yes
PFL terms	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes
R^2	0.5259	0.4961	0.3495
Prob>F	0.0000	0.0000	0.0000
Observations	364	364	364