CEO evaluation horizons and innovation∗

Lingtian Kong†

January 18, 2017

Abstract

To assess its impact on innovation, I focus on an important aspect of corporate governance that has not been empirically identifiable – the CEO performance evaluation horizon. Because trial and error is a time-consuming process, evaluation horizon is an important determinant of innovation, an insight provided by Manso (2011). Taking advantage of the 2011 SEC rule change that requires all public US companies to vote on the “say on pay” frequency, I apply sharp regression discontinuity design (RDD) to show that firms with long term evaluation generate low-value innovation at first but more valuable innovation in the long run. Furthermore, my evidence suggests that the underlying mechanism is that long evaluation horizons allow costly “explorations” in a broad spectrum of technological fields in the beginning. Later on, the firm benefits from the exploration by “exploiting” the field that was identified as the most promising during exploration. The patent “truncation problems” that traditionally limit investigations of recent (here 2011 onward) innovation are solved by the collection and usage of patent application data. The differentiation of exploration and exploitation is enabled by two novel measures of exploration.

∗I would like to thank Jose Albuquerque de Sousa, Dion Bongaerts, Mariano Croce, Gelly Fu, Sebastian Gryglewicz, Tim Jenkinson, Stefan van Kampen, Thomas Lambert, Fabrizio Spargoli, Wolf Wagner, Scott Yonker and Eden Zhang for helpful comments.
†PhD candidate, Rotterdam School of Management, Erasmus University, Department of Finance, Burgemeester Oudlaan 50, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: kong@rsm.nl.
1 Introduction

Corporate governance has first order effect on firm innovation. This notion is supported by evidence in various contexts\textsuperscript{1} and underlies the JOBS (Jumpstart Our Business Startups) Act. Among various aspects of corporate governance, CEO performance evaluation horizon is of particular importance, as predicted by theories. Stein (1988) points out that the short term performance pressure may result in CEO Short-Termism, including the abandonment of innovation. Recently, Manso (2011) shows that the incentive scheme that best motivates innovation tolerates short term failure and rewards long term success. High value innovation is often accompanied by great uncertainty. Its completion requires “exploration” as the first stage, which serves to reduce uncertainty but does not generate measurable short term performance. Its great benefit is actualized in the “exploitation” stage when the firm is able to produce tangible innovation outputs (patents, etc.). But, when the evaluation horizon is too short, the shareholders of the firm are unable to determine whether the temporary lack of performance is the result of exploration or shirking. In equilibrium, the friction caused by the short term shareholders oversight may discourage the CEO from exploring, which means the firm is unable to pursue high value innovation. The competing “quiet life theory”\textsuperscript{2} by Bertrand and Mullainathan (2003) makes the effect of evaluation horizon on innovation ultimately an empirical question. But so far it has been difficult to measure the evaluation horizon directly\textsuperscript{3}. The lack of direct measurement makes the impact of evaluation horizon on innovation controversial. For example, this issue is a subject of debate in the 2016 US presidential election\textsuperscript{4}.

In this study, I identify the CEO evaluation horizon by exploring a unique SEC rule change made in 2011. All public firms in the US were required to hold a shareholder vote in 2011 on the frequency of the “say on pay” (SOP) votes. The outcome of this vote determined how frequently the shareholders would discuss and vote to approve/disapprove the executive compensation proposal. To identify the causal effect of interest, I apply “regression discontinuity design” (RDD) to compare

---
\textsuperscript{1}Notable examples are Seru (2014) and Bernstein (2015).
\textsuperscript{2}Without short term evaluation the CEO would simply shirk and not innovate.
\textsuperscript{3}Natural experiments usually generate exogenous variations in the intensity of governance but rarely in the time length of evaluation horizon. The closest proxy for the evaluation horizon have been active institutional ownership and public ownership in general. But these proxies are not able to exclude channels other than the horizon effect. The difference of this paper with another paper on this topic, Gonzalez-Uribe and Xu (2015) will be discussed in the last part of the introduction. The empirical studies on this issue are reviewed by Kerr and Nanda (2014) and Chemmanur and Fulghieri (2014).
\textsuperscript{4}For two opposing views expressed in the Wall Street Journals, see: “The Imaginary Problem of Corporate Short-Termism.” (Roe (2015)) and “Clinton Gets It Right on Short-Termism.” (Galston (2015))
firms whose shareholders voted for a lower frequency of SOP by a small margin to those that voted for a higher frequency by a small margin. Therefore the firms that lie close to either side of the vote share threshold can be considered identical in all aspects but the choice of the evaluation horizons. I use SOP as a proxy for corporate governance in general because its effect extends beyond executive compensation. The SOP usually occurs as the last session of the shareholder annual meetings, by which time the shareholders have heard all items on the agenda and can use the SOP as a tool to punish or reward the CEO for all governance issues. Furthermore, unlike most other shareholder votes, the vote on SOP frequency is not manipulated by the management. In addition, it is implemented 99.7% of the time in my sample. These two features allow for a clean sharp regression discontinuity design (RDD), eliminating concerns of endogeneity.

To address a long lasting challenge in the measurement of innovation, I apply a new methodology and collect new data. Previously, due to “truncation problems” with the patent data, researchers have only been able to evaluate an innovation at least 10 years after the patent is granted (Lerner and Seru (2015)). An analogous problem is also found in the academic publication process. When evaluating a researcher’s recent productivity, one counts not only his published articles, but also his pre-publication working papers, since most of his recent work is still in this stage. Similarly, the pre-grant patent application data is crucial in the case of corporate innovation. However, it has not been collected into one database or matched with company identifiers such as CUSIPs. This makes timely (here, a shock in 2011) evaluation of innovation difficult. Thus, I collect patent application data via large scale data mining and show through backtesting that by adding the application data, I am able to evaluate innovations as recent as three years with accuracy and precision.

Despite the importance of exploration and exploitation in innovation literature, the literature does not provide measures to differentiate between the two processes. I design two novel measures for exploration. The first measure is the Herfindahl dispersion of the technological fields that each firm files patent applications in. Exploration usually requires experimenting with various fields; thus, the field dispersion represents the intensity of exploration. The second measure is the ratio of forward and backward self-citations. By definition, the more a firm explores, the less it relies on its previous knowledge, which is represented by fewer retroactive citations on patents previously filed by itself.

I assemble a sample of 2247 firms that held the say on pay frequency vote

---

5Backward self-citations are the citations given to existing patents of the same firm while forward self-citation are the citations a patent would receive from future patents of the same firm.
and have patenting activities. In 2011, all the firms held\footnote{A small number of firms held it in 2013. I will discuss this in detail in section\textsuperscript{3}.} the SOP frequency vote in their annual meeting to determine the voting frequency of the SOPs. The options given by the SEC were either “every year” (“distracted firms”) or “once every three years” (“undistracted firms”). The first installment of the SOP was held in 2011, and the second installment was to occur in 2012 for the distracted firms and in 2014 for the undistracted firms. Thus, the undistracted firms had a three-year undistracted period from 2011 to 2014, while the distracted firms did not. I apply RDD by comparing the innovation of these two groups of firms year by year throughout the undistracted period (2011 to 2014) for the undistracted firms. Any difference in innovation between these two groups during these years is a result of their different evaluation horizons (one year versus three years).

Consistent with the predictions made by Manso (2011), I find a dynamic relationship between the frequency of SOP voting and the value of innovation outputs. The undistracted firms produce patents that are 7\% less economically valuable in the first year of the undistracted period\footnote{The economic value is measured by the Kogan (2012) measure. It is computed based on the stock market reaction on the date that a patent application is granted.} but in the third year, this trend is reversed. Indeed, at this point in the undistracted period, the undistracted firms apply for patents that are 7\% more economically valuable than the distracted firms. Therefore in the end, the undistracted firms break even relative to the distracted firms, and this does not take into account the potentially greater benefits it may experience after the three year period. The scientific value of these patents, measured by the number of citations received, demonstrates a similar pattern. These results suggest that having the SOP meeting less frequently leads firms to make more valuable innovations in the long run at the expense of short term poor output.

What leads to this dynamic effect? To shed light on the underlying mechanism, I compare the firms’ emphasis on exploration versus exploitation year by year. I find that in the first year, the undistracted firms are more likely to explore a broader spectrum of technological fields, measured by the Herfindahl dispersion defined above. In addition, I find that in the second and third year, the undistracted firms are more likely to have changed the technological fields in which they innovate, compared to their fields before the shock. Such changes are found not to be routine for these firms, indicating an extraordinary and permanent effect of exploration. These results suggest that without short term pressure, firms are first able to spend time exploring a myriad of technological fields. As a result, they are able to zero in on the most lucrative area for innovation later on. These results also confirm the empirical prediction by Manso (2011) that suggests that exploration
manifests its benefit only in the long term and is tolerated only by firms with long evaluation horizons. What is more, the change of technological focus suggests a intrafirm Schumpeterian creative destruction, which on aggregate leads to progress in technology.

Finally, I turn my attention towards the distracted firms to validate an assumption made by Manso (2011) that if given frequent opportunities to evaluate the CEO, the shareholders indeed punish the CEO with the SOP vote should initial exploration results in poor output at first. I find that SOP vote is less in favor of a compensation plan proposed by the management, an act of punishment, if the economic value of patents are poor in that year, regardless of whether this is a result of exploration. This suggests that the shareholders do not observe the effort exerted in exploration and instead respond to short term performance. It should be noted that this last result may suffer from selection bias since it is not generated from RDD but a multiple regression with possible omitted variables.

As already explained throughout the above text, this paper makes three important contributions. Not only is it the first study to empirically identify the causality of CEO evaluation horizon on innovation, but it develops a data set and methodology to enable the study of recent innovations. Finally, it designs new measures to differentiate exploration from exploitation.

This paper relates to three strands of literature. First, it joins the debate on the role of corporate governance and ownership structure in innovation. The empirical evidence on this issue is divided. For example, Atanassov (2013) finds that innovation falls in the states that pass antitakeover laws while Chemmanur and Tian (2016) find that innovation is higher in firms with more antitakeover defenses. Similarly, Acharya and Krishnamurthy (2009) find that lenient bankruptcy laws promote innovation, suggestive of a greater willingness to take on risky projects with stronger downside protection. However, Mann (2015) finds that stronger creditor rights are associated with greater innovation by firms in his sample. My findings establish that the effect of corporate governance depends on its horizon.

Other important works in this area include Gonzalez-Uribe and Xu (2015), who look at the job security within contract duration on innovation. There are two aspects of incentive structure that stimulates innovation, as predicted Manso (2011): short term tolerance of failure and long term rewards for success. Each of these two predictions do not hold individually but only work when they are together. So it is important to test both of them simultaneously. Gonzalez-Uribe and Xu (2015)’s paper looks at the impact of CEO contract length on innovation,

---

8Without the tolerance for early failure, CEOs do not make risky but potentially valuable innovations, without reward for long-term success, CEOs will not have an incentive to innovate and simply shirk. It is only when both methods are adopted that CEOs make important innovations.
which only measures short term tolerance for failure but not long term reward for success, while my paper tests both together.

Second, with regards to the underlying mechanism of this differing effect, this study relates to the literature on the trade-off between two types of innovations: exploration and exploitation. Exploration refers to the process in which the innovator experiments with various potential fields or strategies to identify the most promising one. In contrast, exploitations build on the expertise and knowledge acquired from explorations. Recent findings highlight the important role of exploration in generating fundamental innovations. (Kerr, Ramana and Matthew (2014) and Manso (2011)) In particular, the theoretical model proposed by Manso (2011) predicts that the most effective contract to motivate innovation tolerates short term failures and rewards long term success. This study tests this prediction with rich dynamism enabled by novel measures of exploration and the time length specific identification strategy.

Third, by using say on pay as a proxy for corporate governance, this paper relates to the literature of the effect of say on pay votes. For example, Cuñat, Gine and Guadalupe (2012) and Cuñat, Gine and Guadalupe (2015) use RDD to study the stock market reactions to say on pay vote results. In contrast, my study is about the vote on SOP frequency not SOP vote itself. The variation in frequency is necessary to identify the governance horizon. I also supplement this literature by revealing dynamic actions the firms take between two consecutive SOPs.

The rest of the paper proceeds as follows. Section 2 describes the data and measures. Section 3 demonstrates the validity of the identification strategy. Section 4 provides the main results. Section 5 concludes.

2 Data and measures

2.1 Patent Data

2.1.1 Truncation problems with the patent data

This paper introduces the usage of patent application data into the study of corporate innovation. Traditionally, researchers have relied only on granted patent data. But due to “truncation problems” inherent in this data, a firm’s innovation output can only be assessed 10 years after it is produced which makes it impossible to evaluate impacts of recent events affecting innovation. With the addition of application data, however, I show that innovation can be measured with reasonable accuracy, even for events that have occurred as recent as 2013. In this section, I will explain what I mean by “truncation problems” and describe how patent
application data can solve them.

The formal patenting process begins with the filing of an application. Patent status is awarded or denied based on an investigation into the originality of the invention, which takes 34 months on average. Because of this extended processing time, a company’s most recent inventions typically remain in the application stage and do not appear in the patent database. In this sense, the patent count, which represents a company’s innovation, appears truncated in the database, as demonstrated by the left panel of figure 3. In the figure, the x-axis represents the year in question and the y-axis the average number of patents per firm that were applied in the year in question and granted by the end time of the database. The red dots represent the number of patents counted in 2006, whereas the black dots represent the patents counted in 2016. Hall, Jaffe and Trajtenberg (2005) demonstrate that inventions that are granted patent status receive their status within ten years after the initial application. Therefore for the same firm at the same filing year, the black line represents the eventual accurate patent count, whereas the red line represents what researchers could construct with data available in 2006. The widening gap between the two plots appears as if the firms’ innovative activities declined as the sampling period came to an end, but this truncation is actually an artifact. The company’s recent patent patent applications have not been granted patent status yet, and thus are not collected by the patent database.

A related truncation problem exists in regards to citations that a patent receives, which are often used as a measure of the scientific value of the patent. Patents garner citations over time, and a large number of citations they receive in the patent’s infancy come from other applications that have not been awarded patent status; though, citations are also received from other patents. In addition, a patent starts to receive citations when it is still an “application”, from both other applications and other patents. In total, there are four sources of citations that a firm’s inventions receive in a certain year (Figure 2), but without the application data, researchers are only able to observe those from the first source (arrowhead A). This is especially an issue for the recent innovative activities, since most of the innovation outputs are still in the “application” stage thus missing from the database. The right panel of Figure 3 illustrates the truncation problem with the citations. One can control for the time fixed effect to eliminate resulted biases, but the small number of citations and patents counted make the measurement very noisy. Traditionally, researches have waited at least 10 years after a patent application was filed to assess its importance, by which time the small sample problem has been compensated by the accumulated citations over the years.
2.1.2 Adding the application data

There are two challenges in using the application data. First, the application filings published by the USPTO were in the form of around six million individual webpages rather than one dataset, and the company names, if provided, are not standardized and may include typos. For example, the company “Hewlett-Packard” has as many as 34 name variants, including “HP”, “HP INC.” and “Hewett-Packerd”. Second, these patents are not matched with company identifiers such as the CUSIP, making it difficult to merge them with financial databases.

I overcome these challenges using Python scripts to perform web crawling to obtain patent and application information. I then match the names with CUSIPs using the pre-existing matches provided by the NBER patent database (constructed by Hall et al. (2005)). I also use name matching algorithms to overcome incomplete names from the patent database. Further details on the process are described in the appendix.

To confirm that the added application data allows me to accurately measure recent innovation, I conducted “backtesting”. This methodology has been used primarily in asset pricing studies to test whether an investment strategy will be profitable in the future by assessing whether it would have been profitable had it been applied in the past. An important caveat lies in that the only information that can be used in the testing is the information available at the time the testing is intended to emulate. I use the same principle to conduct backtesting of the innovation measures in three steps. First, for all the firms, for the years 2001 – 2006, I construct measures both with the application data (the “supplemented measure”) and without (the “raw measure”). Additionally, only the information available in 2006 was used. In other words, any application that was granted after 2006 is counted as an application rather than a patent. Similarly, any citations that occurred after 2006 are not counted. Then for the same firms of the same years (2001 – 2006), I construct the innovation measures using only patent data, but using the information that is available as of 2016. In other words, I count the applications that eventually get patent status as well as citations these patents received after 2006. The measure constructed in this step, which allows for a 10-year accumulation period is what have been used in the literature. It has the disadvantage of a long wait, but should be precise and accurate. Thus it serves as the benchmark against which the supplemented and raw measures are evaluated. Finally I compute the Pearson correlation coefficients between the supplemented and raw measures and this benchmark. The difference between these two correlations signifies how accurate the supplemented measure (including the application data) is compared to the raw measure (without the application data).
that this paper creates is relative to the raw measure. If this advantage is big enough, we can be confident to extrapolate that, the supplemented measures that I compute for the year 2011, with the data available as of 2016, accurately predict the citations and patent counts that this firm will receive in the long run, say, by 2026.

The backtesting shows that the supplemented measure computed with the added application data features less bias and less noise compared to the raw measurement that does not include application data. This improvement in accuracy is demonstrated in Figure 4 by the alleviated truncations (green dots) for both patent counts (upper panel) and citation counts (lower panel), relative to their raw counterparts (red dots). Aside from accuracy, the assessment of precision is reported in Table 5, which tabulates the Pearson correlation coefficients between the two tested measures and the benchmark. As the literature rightly worries, the raw measure correlates poorly with the true value when the patent is very recent, which is represented by its low correlation with the true value for all five tested years, especially the last three. In contrast, the supplemented measures correlate very well with the benchmark, especially for the first three years. Repeating this backtesting with information sets available in 2007, ..., 2015 respectively (results not shown), it can be shown that the “complete measure” computed for the 2001 – 2003 using the information available as of 2006 is as accurate as the raw measure computed as late as 2011. This makes us confident that the supplemented measures computed in this paper for the years 2010 – 2013 is accurate. The measure for 2014 is noisier but still would have a high accuracy.

2.2 Measures of innovation

In this subsection I describe rigorously how the supplemented measures are computed with the added application data, and why they are less biased and less noisy.

2.2.1 “Supplemented measures” of patent counts and patent citations

Before explaining these two supplemented measures, I first describe the measures that have been used in the literature, which I refer to as the “raw measures”. These measures are precise only when they are computed for patents and applications filed at least 10 years before the end of the sample period of a studyLerner and Seru (2015). They serve as the benchmark to which the supplemented measures are compared and contrasted.

“Raw patent count” $P_{j,t}$ is simply the total number of patents that firm $j$ filed
in year $t$, that are granted patent status as of year $T$, the last year of the available patent data:

$$P_{j,t} \equiv \sum_{p \in P_{j,t}} 1$$

where $P_{j,t}$ represents the set of granted patents that firm $j$ filed in year $t$. This measure has been used to represent the quantity of the innovation output that a firm produces in a certain year. But, it suffers from the truncation problem, since as $t$ is very close to the end of the sample period $T$, most patents are still at the stage of “applications”, and are thus accounted for by this measure.

“Raw average citation count” $\bar{C}^{P,P}_{j,t}$ is the average normalized number of citations that each patent in $P_{j,t}$ receives from other granted patents as of the end of the sampling period $T$:

$$\bar{C}^{P,P}_{j,t} \equiv \sum_{p \in P_{j,t}} \frac{C^{P,P}_{p}}{\bar{C}_{f,t}} P_{t,t}$$

where “$P, P$” super scripts represent the fact that the citations counted are only those from granted patents to granted patents. Thus, $C^{P,P}_{p}$ represents the citations that patent $p$ receives from other patents. $\bar{C}^{P,P}_{f,t}$ represents the average number of citations that are received by other patents filed in the same technological field in the same year. The division by this cohort average serves to control for the substantial cross field and time series variation of the patent citations Lerner and Seru (2015).

As alluded to in the previous section, the patent-to-patent citations counted by this measure only account for one of the four types of citations, which is represented by the $A$ type in Figure 2. As $t$ approaches the end of the sampling period, most of the patents are still applications, which results in this measure ignoring a large part of total citations.

In summary, since they are constructed using only the patent data, the raw measures suffer from small sample problems when constructed for recent patent and applications, and this reduces the measure’s statistical precision.

Because of the limitation inherent in using only patent data, I add application data to my analytic repertoire. However, patent applications should not be given the same weight as granted patents in quantifying innovation. After all, only 56% of all applications are eventually granted patent status. This implies that on average, applications are of less value than granted patents. Thus, to the extent that the

---

9In this paper, I use “Holy Roman bold” font to represent sets.
number of eventually granted patents is the most accurate measure of innovation, the supplemented measure would generate a positive bias if the applications were used without any discount. This is analogous to the fact that when evaluating someone’s research output, her working paper is less of a signal than her published papers, *ceteris paribus*.

To implement this discount to the applications, I derive the “application discount factor” $\beta_{\tau,f}$. It is the conditional probability that any patent application that hasn’t been approved in $\tau$ years will be approved eventually. It is estimated by:

$$\beta_{f,\tau} = \Pr(\mathcal{A} \cup \mathcal{B} | \bar{\mathcal{A}}) = \frac{\Pr(\mathcal{B})}{1 - \Pr(\mathcal{A})}$$  \hspace{1cm} (3)$$

$$= \sum_{l=\tau+1}^{\infty} g^f(l) \frac{1}{1 - \sum_{l=0}^{\tau} g^f(l)}$$  \hspace{1cm} (4)

where $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ represent three disjoint events. $\mathcal{A}$ is that a patent application filed in year $t$ would be approved as of year $T$, $\mathcal{B}$ is that it would be approved after year $T$, and $\mathcal{C}$ is that it would never be approved eventually. Thus by Bayes’ rule, formula 3 represents the conditional probability that any patent application that hasn’t been approved in $\tau$ years will be approved eventually. This probability is estimated by formula 4 where $g^f(l)$ is the number of patents in technological field $f$ that are granted $l$ years after the application was filed. The estimation assumes different application-grant lag distribution for patents in different technological field, a phenomenon reported in Hall et al. (2005). Any patent that shows up in the patent database must have been approved prior to $T$, then by this formula, it has a discount factor that equals 1.

With the discount factors applied to applications, I am able to compute the unbiased supplemented version of the two raw measures.

The supplemented measure for patent count is defined as the total number of patents and applications a firm $j$ filed in year $t$, with the applications discounted:

$$PA_{j,t} = \sum_{p \in P_{j,t}} 1 + \sum_{a \in A_{j,t}} \beta_{F(a),t}$$  \hspace{1cm} (5)$$

$$= P_{j,t} + \sum_{a \in A_{j,t}} \beta_{F(a),t}$$  \hspace{1cm} (6)$$

raw  \hspace{1cm} discounted

patent \hspace{1cm} application

count  \hspace{1cm} count
where $F(a)$ is the technological field to which the application $a$ belongs. $A_{j,t}$ is the set of applications filed by firm $j$ in year $t$ but have not been approved as of $T$. Each application is not counted as one like the patents, but is discounted by the probability $(\beta_{F(a),t})$ that it will eventually turn into a patent, given that it hasn’t made it as of $T$. As a result, it can be shown that this measure converges to the benchmark. In contrast, the raw measures would underestimate the benchmark for recent patents because it is composed only of the first term in formula $[6]$. The undiscounted supplemented measure, on the other hand, would overestimate it because it does not apply the discount factors that are smaller than one. This convergence translates into the accuracy of the supplemented measure that has been illustrated in Figure 3.

The “supplemented average citation count measure” is defined as the average number of citations made to each patent or application that a firm filed at year $t$, normalized by the mean citations to other patents or applications in the same cohort:

$$
\bar{C}_{j,t} = \frac{1}{PA_{j,t}} \left[ \sum_{p \in F_{j,t}} \frac{C_{p,P}^{P,F}}{C_{F(p),t}} + \beta_{F(c),T(c)} \frac{C_{P,A}^{P,F}}{C_{F(p),t}} + \sum_{a \in A_{j,t}} \beta_{F(a),t} \frac{C_{A,P}^{A,F}}{C_{F(a),t}} + \sum_{a \in A_{j,t}} \beta_{F(a),t} \frac{C_{A,A}^{A,F}}{C_{F(a),t}} \right]^{(7)}
$$

Similar to the supplemented measure for patent count, this measure can be shown to converge with the “long waited measure for patent count”. In contrast, the raw measures would underestimate the “long waited measure” for recent patents, because it is composed only of the term $A$ in formula $[7]$. The terms $B$, $C$, and $D$ correspond to the other three sources of citations in Figure 2 that are not utilized in the raw measure. The undiscounted supplemented measure, on the other hand, would overestimate it because it doesn’t apply the discount factors that are smaller than one. This convergence translates into the accuracy of the supplemented measure illustrated in Figure 3.

10In this paper, upright bold letters represent operators that take a patent or an application as the argument, and generate a certain characteristic of that patent or application.
2.2.2 Measure of economic value of patents

It is important to differentiate between the economic value and scientific value of patents. Although it has been shown that these two values are correlated, they are not always aligned (Kogan (2012)). While it is the scientific value that is important for society, most shareholders prioritize economic value. Therefore it should be the economic value not scientific value that shareholders provide incentive for. Additionally, the CEO should also arrange the innovation based on this value.

The economic value of a patent can be measured based on how the patent affects the value of the inventing company, which is computed based on changes in stock price on the day the patent is granted (Kogan (2012)). It is assumed that the shareholders know the true economic value of the patent before the patent is granted. The granting of the patent reveals information to the stock market because it corresponds to a discrete increase in the shareholders’ beliefs that this invention can reach patent status, which guarantees the company a monopoly rent for a long period of time, thus holding the patent can have a substantial impact on the value of the firm. Refer to the appendix for further details on the computation of this economic value measure.

2.2.3 Measures to differentiate exploration and exploitation

In order to test the shift of focus between exploration and exploitation, I need proxies to measure these two types of innovations. Surprisingly, despite the burgeoning literature on exploration and exploitation, there are not any measures designed to the best of my knowledge. Therefore in this session, I have designed measures that differentiate these two types of innovation.

In the innovation literature, exploration usually refers to the endeavor of exploring an area new to the innovator. I identify the flow of heritage within a firm by the citation it gives to previously granted patents, as the identities of the cited and citing patent firm have been shown to represent knowledge flow (Jaffe, Trajtenberg and Fogarty (2000)). To be specific, for any patent, the “exploration/exploitation quotient (EEQ)” of any patent is defined as the number of forward citations divided by the backward citations. Here forward citations refer to the citations that this patent receives from the future patents of the same firm, and backward citations refer to the citations that it gives to the previous patents of the same firm. The more a patent cites previous work of the same firm, the less it is exploring unfamiliar areas. The number of backward citations, however, is affected by company size (Hall, Jaffe and Trajtenberg (2001)). As companies grow in size, they are more likely to cite their own patents more frequently. To control this difference, I normalize the number of forward citation by dividing the number
of backward citations. The rationale is that companies that have more forward
citations should also have more backward citations, unless they are in a exploration
stage.

Another important manifestation of exploration is the diversity of the fields
experimented in. Due to inherent uncertainty a firm does not know whether it is
going to get the information it needs before setting out to explore an unfamiliar
area. Therefore it is very likely that it will explore a second area. As a result,
the more aggressive a company is exploring, the more fields it is likely to engage
itself in. Therefore the second measure for exploration is the Herfindahl dispersion
(“Herf”) of the technological fields in which a firm $j$ files patent in year $t$:

$$Herf_{j,t} = 1 - \sum_{f=1}^{7} \theta_{f,j,t}^2$$

where $f = 1, 2, ..., 7$ is the seven-category division of technological fields proposed
by Hall et al. (2001). This categorization has been shown in the literature to capture
the bulk of cross field variations in patenting activities. $\theta_{f,j,t}$ is the share of patent
applications in each of the seven fields that company $j$ files in year $t$. The value of
this measure increases as the fields a firm engages in get more diversified.

### 2.3 Voting data

Implemented in the spirit of the Dodd-Frank Act, on Jan. 25th., 2011, the SEC
issued a new rule requiring that all reporting companies have their shareholders
vote on the frequency of the “say on pay” (SOP) meetings to be held in the future.
The outcome of this vote determines how frequent the Say On Pay votes are to
be held in the subsequent years. The first SOP vote is also required to be held
in 2011, at the same time with the vote on its frequency.(See[1] for a rough time
line.) These SOP meetings would be held as a session in the annual meetings,
during which the shareholders vote to decide whether to approve the executive
compensation packages proposed by the management[11]. The companies are given
three choices of the frequency: once per year, once every two years and once every
three years[12]. The SEC also required that this voting take place in the first annual
meeting after the rule change. The companies with market capitalizations smaller

---

[11]It’s important to clarify that there are two types of voting involved here, the voting to occur
in the SOP meetings and the voting beforehand on the frequency of these meetings. The second
type of voting provides me with the RDD identification, and the first type is involved in producing
only one set of results. The first type is used in Cunat el al 2015 but not the second type. I will
refer to the first type of voting as “SOP voting” and the second type as “SOP frequency voting”.

[12]I address later how RDD can be applied to voting with three possible outcomes.
than 75 million USD were the only ones that were granted a later deadline, which was set in their annual meetings in 2013. Each company is also required to disclose the voting result and whether it will be implemented in its 8−K filings. I obtained the SOP frequency voting results from the ISS, which includes the year of the voting, the share of votes for each of the three choices and whether the voting result was implemented. I also obtained from ISS the information regarding the SOP meetings themselves, including the meeting date and the vote share for and against the compensation plan proposed by the management.

In addition, I obtained information pertaining to the company fundamentals from Compustat as well as stock prices and market index prices from CRSP.

3 The identification strategy

It is challenging to establish the causal relationship between the frequency of corporate governance and corporate innovation. There might be underlying factors, such as the innovation strategy of the company, that drive both variables. In addition, the shareholders may anticipate the innovation intensity of the firm and choose governance frequency accordingly. As mentioned in the introduction, these underlying factors, most of which relate to the propensity to innovate, are not observable, and thus, are difficult to control for.

To overcome this challenge and prove causality, this paper exploits the 2011 SEC mandated SOP frequency voting as an exogenous source of variation in the governance frequency. I apply regression discontinuity design (RDD) methodology, effectively comparing the innovation outcomes of the firms whose shareholders voted for “once per year” with a small margin over 50% to those that are marginally in favor of “once every three years.” As long as these two groups of firms are identical in all other aspects, we can attribute all later differences in innovation to the different SOP frequencies adopted by the firms. This identifying assumption is indeed satisfied, as is demonstrated later in the paper. For this and other reasons listed below, the regulatory mandate exploited provides a unique opportunity for clean identification.

3.1 No manipulation and the balance check

The SOP frequency vote is rarely manipulated by management. If the management was able to influence votes cast by the shareholders, the “everything else equal” assumption would be violated. This is because the firms that narrowly voted for “once every three years” would have been more likely to possess characteristics that make this voting outcome appealing to the management so that management would
try hard to sway the votes to obtain this result. Empirical testing shows that this concern is not valid. Had manipulation occurred, there would be a discontinuity at the 50% cutoff in the vote share distribution. Figure 6 and Figure 7 suggest that this is not the case.

To evaluate the probability of manipulation rigorously, I applied the test of discontinuity in distribution developed by McCrary (2008). The point estimate of the discontinuity is 0.1109 with a standard error of 0.1556, thereby making it insignificant. Figure 7 plots the probability density estimated by this test and the discontinuity is shown to be almost nonexistent.

The absence of manipulation observed here is very rare in management initiated shareholder voting. In the US, most voting is manipulated by the management, as reported in Yermack (2010). The absence of manipulation in the SOP frequency voting is probably due to the pressure for prudent governance in the immediate aftermath of the 2008 financial crisis.

The absence of manipulation alone guarantees the “everything else the same” condition in a large sample asymptotically. I nonetheless compare important ex ante firm characteristics between the treatment and control groups to provide additional validation, as reported in table 2. Firm characteristics that have been shown to affect innovation do not differ significantly across the two groups. In the results section, I take this notion further and show that the various innovation measures do not differ significantly across the treatment and the control before the treatment took place. Taken together, this is strong evidence that the treatment and the control groups differ only in whether the treatment is received.

### 3.2 Near 100% implementation

The “cause” in the causal relationship of interest is the implemented frequency of the SOP rather than the mere voting frequency. If the voting result was not always implemented, I would have needed to adjust the directly estimated effect size by the implementation rate, introducing noise. But strikingly, the implementation of the SOP frequency voting results are almost 100%, allowing for an intuitive and precise sharp RDD. This feature of near perfect implementation is illustrated by figure 9.

---

13Imbens and Lemieux (2008) shows that the RDD is valid for identification as long as the management cannot manipulate the outcome perfectly. But in that case the interpretation of the causal effect would be less straightforward and its estimate less accurate. The SOP voting, which has no manipulation at all, provides cleaner identification.

14The vote share in this figure appears asymmetric, representing the fact that in most firms, the shareholders voted for “once per year”. This, however, does not affect the validity of the RDD in any way. This asymmetry is probably a result of the elevated consciousness of governance after the financial crisis.
A near perfect implementation is unique to this particular voting occurrence, as previous work has found that the management only implements a fraction of these non-binding voting results (Yermack (2010)). This is again probably due to the stringent application of corporate governance in the wake of the 2008 crisis.

### 3.3 Exogeneity of the participation and voting time

As explained earlier, all the firms that report to the SEC are required to hold this voting procedure, and none of them have the luxury of doing so on their own timeline. All the “big” firms were required to vote in their annual meeting in 2011, while all “small” firms had to vote in their 2013 meetings, with the criteria for each category ("big or small") set arbitrarily by the SEC. Had the participation been voluntary, the causal effect measured would have been the “local average” for the firms motivated enough to hold such a voting, decreasing the external validity of the identification. If the companies had had the freedom to choose the time of the voting, the problem would have been even more complicated because it is well known that on the aggregate level, innovations come in waves over time (Lerner and Seru (2015)). Thus, it is possible that the companies might time the SOP frequency voting to accommodate these innovation waves. Trying to overcome this problem by adding time fixed effects would have biased the effect size estimation as the byproduct by abandoning meaningful time series variations as pointed out by Hall et al. (2005).

Fortunately, this paper does not suffer from these concerns. Due to the mandated participation and timing of this voting, all relevant variations can be controlled for. This exogenous participation and timing is rare for shareholder voting, because most shareholder voting is called not randomly, usually if and when needed.

### 3.4 SOP frequency representing governance frequency in general

“Say on pay” meetings are about more than just “pay”. They have profound influence on corporate governance in general. These SOP sessions are part of the shareholders’ annual meetings, during which other important issues are communicated and discussed. Typically, the SOP session is the last item on the agenda. The timing of the vote makes the SOP session a powerful tool of the shareholders when negotiating issues beyond executive compensation. After listening to reports from management about past performance and future plans, the shareholders can use the SOP vote to punish or reward the management for all issues discussed in the annual meeting. Furthermore, what is at stake for the management is arguably
what they care the most: their compensation. It must be noted that the result of the vote is not binding\textsuperscript{15}, so the management can adopt pay scheme that deviates from the shareholders’ vote. But Cuñat et al. (2012) shows that they nonetheless “get the message” and would modify behaviors shareholders signal disapproval of. Ultimately, the frequency of the SOP vote substantially represents the frequency of the governance actions that the shareholders get to take on the management.

3.5 RDD applicable to voting with three possible outcomes

RDD is usually applied to votes with only two choices on the ballot. In the case of SOP frequency vote the shareholders have three choices: once per year, once every two years and once every three years. In the appendix, I prove econometrically that the identification assumptions carry over to the majority vote cases with three possible outcomes, as long as the one of the voting result is not used in the analysis, and the vote shares used are the ones for the remaining two choices. Loss of observations presents the only downside to this method, but in this case less than 5% of the companies voted in favor of “once every two years” so abandoning this outcome is not a significant loss. The other two options typically receive most of the votes in my sample, an outcome common to the plural majority votes.

3.6 External validity

RDD is known to have very strong internal validity but relatively weak external validity. However, I argue that this does not apply to the RDD with vote share as the running variable.

RDD would have had limited external validity had the running variable be a fundamental characteristic of the firm. For example, if the treatment is applied only to firms larger than a size threshold, then the RDD estimate is only locally valid for the companies with sizes very close to the threshold. Vote share, however, is not a fundamental feature of the company, so the firms close by the vote share threshold may be anywhere in the spectrum of firm characteristics, unless the vote share is strongly correlated with a certain firm characteristic. Table \textsuperscript{3} shows that the distribution of firm fundamentals is very broad in the sub-sample of close-call votes.

\textsuperscript{15}The SOP frequency voting are effectively binding, as explained above, but not necessarily the SOP voting.
4 Results

4.1 Empirical specification

Throughout the analyses, I use regression discontinuity design (RDD) to estimate the following varying weight regression, based on the methodology developed by Imbens and Kalyanaraman (2012):

\[ Y_j = \alpha + \beta I_{3yr} + P_l(v, c) + P_r(v, c) + \gamma I_{2011} _j + \epsilon_j \]  

(9)

where \( j \) index the firm that holds the “say on pay” (SOP) frequency vote. \( I_{3yr} \) is the indicator variable that equals 1 if and only if the vote result is in favor of “once every three years”. \( I_{2011} _j \) is an indicator variable of whether the company held frequency vote in 2011 or 2013. \( c \) represents the vote share threshold 50% above which the “once every three years” is passed and put in practice. \( P_l(v, c) \) is a flexible polynomial function for observations on the left-hand side of the threshold \( c \); \( P_r(v, c) \) is a flexible polynomial function for observations on the right-hand side of the threshold \( c \) with different polynomial orders; \( v \) is the total vote share (percentage of votes in favor). This equation is estimated with more weight put on the observations closer to the cutoff. This ensures that the analysis relies mostly on the observations near the continuity, which leads to a ceteris paribus comparison. Following the convention of the literature, I report throughout the results section the polynomial specification of the fourth order, but the results are not substantially different when other orders are applied.

The dependent variable \( Y_j \) represents the five innovation measures described in section 2: log patent count \( \log(1 + AP_{jt}) \) measure the quantity of innovation output; average normalized citation count \( \bar{C}_{jt} \) measures its quality in terms of scientific value, market valuation of the patent \( MV_{jt} \) measures the economic value of an innovation, Herfindahl dispersion of the technological fields involved: \( Herf_{jt} \) represents how broad a company explores, and the exploration/exploitation index: \( EEP_{jt} \) measures how a firm explores areas that it did not have expertise in before. These measures are time varying, and I track their evolution within the three year undistracted period of the undistracted firms benchmarked against the distracted firms. Therefore I apply the one, two and three year lagged values of these dependent variables after the SOP frequency vote/first installment of the SOP on the left hand side of the regression.

Indicator variable \( I_{2011} _j \) controls the difference in timing of when the SOP vote takes place. Self selection of the participation time can result in endogeneity that is difficult to control because the innovation comes in waves and the self selection may cater to this time trend. This concern is not necessary in this setup,
because the SEC mandates that companies with market capitalization larger than 75 million dollars hold the SOP frequency vote in their annual meetings in 2011 while those smaller than 75 million vote in 2013. This arbitrary assignment of treatment time makes the control simple because the variable that needs control is well defined and observable. Aside from this covariate representing heterogenous in the vote mechanism itself, the RDD identification assumptions validated in section 3 ensure that the firms whose shareholders vote in favor of “once every three years” by a small margin of votes is otherwise identical (on average) to firms whose shareholders vote in favor of “once per year” by a small margin. The only difference between these two groups of firms is that the undistracted firms would have the SOP once every three years while the second group would have it every year. Thus, all differences in innovative outcomes in subsequent years can be attributed to their different governance horizon.

4.2 Innovation quantity and quality

I find that compared with the control group, firms that hold say on pay (SOP) votes once every three years generate patents with less economic value in the first year but patents with greater economic value in the third year.

The economic value of a patent is measured based on the stock price reactions on the date \([t, t + 2]\) of this patent was granted. Compared to the scientific value measured by the number of citations, the economic value is more relevant to shareholders and corporate governance. It contributes to the firm value directly and may be used by shareholders as implicit performance targets to evaluate the CEO.

Column 1 of table 4 reports the aforementioned difference in economic value. Undistracted companies produce patents that are on average 1.39 million less valuable than the distracted companies in the first year after the SOP frequency vote. The first installment of the SOP is required by SEC to be held immediately after the SOP frequency vote, therefore this year is also the first year in the three-year “undistracted period” for the firms that have SOPs once every three years. From the second year on, however, this difference starts to be reversed. Column 2 shows that in the second year, the undistracted companies caught up. They start to produce patents that are not significantly different in value from the control firms. The comparison is completely reversed in the third year, by which time the undistracted firms produce patents that are 1.36 million more valuable than the other group. This difference is robust with alternative orders of polynomials in equation 9. This difference can also be seen in the upper panels of Figure 10. The sharp discontinuity at the vote cut-off represents the effect size of the low frequency
of the SOP.

The consequence of the less frequent SOP vote is significant. The 1.36 million represents a 7% change of average market value of a patent (19.99 million). From another perspective, multiplied by the average number of patents per firm (22), the total value lost in the first year and the value gained in the third year are both about 30 million for each undistracted firm. Gain and loss of this magnitude may cause serious concerns of the shareholders in the first year and great relief in the third year. However, when shareholders commit to not holding an SOP vote in the first year, then the management of undistracted firms remain insulated from short term pressure applied by corporate governance.

This drastic time varying effect also demonstrates the importance of a clean identification of the treatment time in innovation research. Since the firms can change their output of innovation in matters of two years, any measurement error in the treatment time that is larger than a year effectively produces a moving average of the time varying effects with opposite signs. This may explain why the existent studies find contradictory results when quantifying this relationship.

Compared with the economic values, the comparison of the distracted and undistracted firms in scientific values tell a slightly different story. The effect in the third year is similar to that of the economic value. The undistracted firms file patents that later garner significantly more citations relative to those filed by the control firms. In the first year, their patents are not less cited than the control firms. The difference in dynamism between the scientific value and economic value demonstrates that while for the private benefit of shareholders, the cost and benefit of less frequent SOPs does not break even until the third year, from a societal point of view, less frequent SOPs lead to a net increase in the quality of intellectual output produced.

In contrast with the rich dynamism of the patent quality in terms of economic and scientific values, the patent quantity is not significantly different between the distracted and undistracted firms throughout this period. The exploration - exploitation literature is agnostic about the implication on the quantity of the innovation output.

As explained above, the pecuniary value of patents produced does not break even until the last year of the evaluation period. At the time of the writing, the firms are still in their fifth year after the SOP frequency vote took place, and as shown in section 2 I am not able to measure their innovation accurately after 2014. Therefore I am not able to determine whether the advantage the undistracted firms started to accumulate is long term. If the effects uncovered above represent a temporary change that would be reversed in the future, then the undistracted
firms would not have a net economic gain from this three-year undisputed period. However, the next section reveals that fundamental changes occur in the underlying technological focus of these firms, which likely positions them in a long term strategic advantage.

4.3 Exploration - exploitation dynamics

How are the undistracted firms able to produce more valuable patents if given more time? Why does this require a trade-off of producing less valuable patents in the first year? Theories provided by Manso (2011) and Kerr et al. (2014) predict that the results in subsection 4.2 can be explained by the roles of exploration and exploitation, two integral stages necessary to generate innovations with great value. “Exploration” surveys all possible solutions to a problem or all possible areas to do R&D in. The information collected in the exploration stage guides the innovator’s direction in the “exploration” stage. They argue that the exploration stage is especially necessary for fundamental innovations that venture into unchartered territories. But exploration is expensive and its progress is hard to measure in the short term. These features may lead to friction in corporate governance, especially when the horizon of the governance is too short to appreciate the long term benefit of exploration. Therefore a possible explanation of the previous results suggests that the undistracted firms do not have short term performance pressure and are thus able to participate in exploration as the first stage. The short term drop in patent value may reflect the time lag of the benefit of exploration. To validate this mechanism, I search for evidence of more explorative activities from the undistracted firms in the first year of the undistracted period.

The results suggest that the undistracted firms are more explorative in the first year of the undistracted period compared with the distracted firms. They file for patents in significantly broader technological fields (Figure 12) and cite significantly less of their previous work (Figure 12). In the second and third year, they do less exploration and more exploitation, evidenced by the higher EEQ. This suggests that these firms enter into the exploitation stage when they cash in on the information collected in the previous stage. The breadth of technology that the undistracted firms explore then drops to the same level as the control firm suggesting that the undistracted firms regained focus.

4.4 Permanent impact of exploration

Does exploration have a permanent impact on innovation? In the current exploration - exploitation framework developed by Manso (2011) and Kerr et al. (2014),
exploration refers only to the search of new solutions to problems in the field that 
the firm is already in. Here I show that exploration can also include the search of 
ew and promising fields for future innovation. This type of exploration may shape 
the trajectory of innovation permanently for a firm. However, within a corporation, 
this type of exploration is even more limited by the short term governance pressure 
than the tool-searching type of exploration because the information asymmetry 
between the shareholders and the management is likely greater on which field is 
more promising for the future. In addition, this question is relevant on a macro level 
because big technological regime shifts at the macro level are achieved through the 
decentralized actions taken by individual firms in an attempt to explore new fields 
and to identify the field that is most promising for future innovation. The answer 
to this question links Schumpeterian creative destruction of old technological fields 
with the effect of long horizon corporate governance at the micro level.

To test this hypothesis, I investigate whether the undistracted firms are more 
likely to change their fields of technological focus. In subsection 4.3 I show that they 
expand their technological horizon in the first year of the cycle and then regained 
focus in the third year. But the subject of the refocus may not necessarily be in 
the area that the firm initially specializes in before the exploration. To test whether 
the field changes are more likely to occur for the undistracted firms, I construct a 
measure of technological migration, which is the sum of square of the change of 
the distribution of the patent applications in the seven broad technological fields:

$$TM_{j,t} = \sum_{f=1}^{7} (avg_{s=t+2}^{t+3} \phi_{j,f,s} - avg_{s=t-5}^{t-1} \phi_{j,f,s})^2$$

(10)

where $j$ index the firm, $t$ index the year, $f$ index one of the seven broad 
technological fields categorized by Hall et al. (2001). $\phi_{j,f,s}$ is the fraction of patents 
firm $j$ filed in year $s$ that are in field $f$. The distribution is averaged over a two 
year period after the SOP frequency vote, and five years before.

RD of this measure reveals that technological migration is indeed more likely to 
occur in these undistracted firms, as demonstrated by figure [13]. The quantitative 
result from RD point estimate is 0.464 with a standard error of 0.197. Not only is 
this result statistically significant, it is also presents a striking magnitude. Consider 
an example: Firm A previously generate innovations in machinery. But, in three 
years, its patent portfolio lay half in machinery and half in hydrolics. This drastic 
change would result in a tech migration measure of 0.516. The significant result 
presents the opportunity to watch Schumpeterian creative destruction unfold at

\[16\] I also look for evidence of tech migration for years before the SOP frequency vote, but didn’t 
find it. This suggests that such migration is not routine.
the micro level, which is allowed by the long governance horizon.

4.5 Short term pressure in the firms with frequent SOP

The concern over short-termism presented by Stein (1988) and Manso (2011) rests on the assumption that if governance horizon is short, shareholders deliver incentives based only on short term performance. However, it is possible that the shareholders would be able to forecast the long term benefits of the explorative activities with the information available in the short term. I test whether this is the case by investigating what the shareholders might do to management in the SOP meetings if the exploration conducted generated low value patents in the short term.

I test this using a multiple regression analysis:

\[ v_{j,t}^S = \alpha + \beta_1 EEP_{j,t} + \beta_2 MV_{j,t} + \gamma ISSrec_{j,t} + \zeta Z_{j,t} + \epsilon_{j,t} \] (11)

where \( v_{j,t}^S \) is the SOP vote share in favor of management sponsored compensation proposal of firm \( j \) at year \( t \). \( MV_{j,t} \) is the average market value of patents filed in year \( t \) by firm \( j \). \( EEP_{j,t} \), the exploration measure is included in the regression to test whether shareholders are able to infer the exploration activity the firm is undertaking, beyond the part of it that is reflected in patent value \( MV_{j,t} \).

\( ISSrec_{j,t} \) is the vote recommendation by the Institutional Shareholders Services, a voting advisor. It is included to capture the non-incentive aspects of the compensation proposal. For example, companies with subsidiaries in Indonesia that recently suffered a big loss due to tsunami may not approve a pay raise proposal from the management. This variation in vote share is unrelated to innovation or the CEOs effort, and it can be captured by the vote recommendation of the ISS. It is true that \( ISSrec_{j,t} \) may also be based on outputs of innovation, and in this case the causal effect of patent value \( MV_{j,t} \) will be negatively biased. So the coefficient on \( MV_{j,t} \) probably gives us a conservative estimation of the true effect size.

Last but not least, vector \( Z \) which includes institutional ownership and leverage, controls for the variation of vote share resulted from the ownership structure of a firm. For example, institutional investors usually have a fixed alliance either with the management (banks, insurance companies) or the shareholders (activist hedge funds). Thus their ownership may affect the vote outcome for reasons other than innovation outputs. The institution ownership variable, however, is a mixture of these institutions with heterogenous alliances. Thus the leverage variable is introduced in the hope of singling out the banks as the pro-management voters.

\[ \text{Note that this vote is different from the vote on SOP frequency.} \]
Results presented in table 7 confirm the assumption that underlies Manso (2011). The relationship between patent market value and SOP vote share is statistically significant. In terms of economic significance, for every million dollar less market value per patent produced, the shareholders vote 0.2% less in favor of a compensation plan proposed by the management. Multiplying this by 1.39 million, the average under-performance of patents in terms of market value in the short term for firms that explore, I estimate that the distracted firms gain 0.3% more votes on average in the SOP by not exploring. But the effect of SOP offers more than the pecuniary ramifications it directly entails. Disapproval of a management proposal of compensation may signal strong discontent, an attitude that is difficult to credibly express by means other than voting or selling shares. In addition, as discussed earlier, the causal effect of patent value on vote share is probably underestimated by this specification. In summary, this result suggests that shareholders do deliver incentive to the management for the patent value created, if they get to perform SOP every year.

The insignificant coefficient on exploration measure $EPP_{j,t}$ demonstrates that the shareholders are not sophisticated enough to infer the exploration activity that the firm is undertaking beyond what is already reflected in the patent market value.

Taken together, these results validate the assumptions previously made about the short term incentive that the shareholders would deliver if SOP is annual, as well as his inability to observe exploration.

5 Conclusion

This paper provides strong evidence that commitment to governance with long term performance evaluation period encourages exploration which leads to innovation with less economic value in the short term, but innovation with more economic value and scientific value in the long term. In addition, exploration demonstrates permanent effect of shifting the technological focus of a firm to the more promising fields in the future.

These findings suggests the importance of matching governance horizons with the natural periodicity of the innovation activities that the firms meant to undertake. In sub-section 4.3 we observe shifting from exploration to exploitation in the second and third year of the three year evaluation period. Though the outcome is superior than SOPs once a year, this shift is not necessarily optimal. It is possible that the shift to exploitation is still too early. This results advocates the SEC’s allowance of options of even lower frequency of the SOP. A concern might be that an option of lower frequency may open the Pandora’s box for the management to manipulate.
the shareholders into choosing a frequency that is too low for their best interest. But the absence of manipulation in the 2011 and 2013 votes is reassuring.

The SOP frequency votes suggests a decentralized way of assigning appropriate time length for governance horizon. A one size fit all governance frequency is obviously not optimal. (Kerr and Nanda (2014)) Mandating shareholders to vote on this issue serves to utilize the information the shareholders have about the idiosyncracy of the innovation that the firm is conducting. To the extent that the shareholders can rationally determine \textit{ex ante} the appropriate governance horizon, voting serves as a commitment device. Therefore testing whether the SOP frequency votes choose the optimal frequency for each firm can be a fruitful area of future research.

References


26


Mann, W.: 2015, Creditor rights and innovation: Evidence from patent collateral. Available at SSRN 2356015.


A Figures

Figure 1: Diagram of governance horizon

Usually corporate governance is interpreted as a mapping from the observable value change of a firm’s to rewards or punishments to be delivered. “Governance horizon” (a.k.a. results evaluation horizon) is an implicit parameter of this mapping because an evaluation horizon needs to be specified for the computation of the change of the observable firm value. The curve in this diagram represents the exploration-exploitation stages of the innovation process. Short term governance horizon would result in the delivery of punishment due to the lack of short term return for the exploration.
This figure illustrates the cause of truncation problems. The solid boxes and solid arrowhead represent the information available in the patent database, and the dashed boxes and dashed arrowheads represent information only available from the application data. Each arrowhead represents one source of citations. For example, arrowhead B represents the citations made to pre-grant applications from granted patents. The patent data does not allow researchers to count applications for the patent count measure. This is the first type of truncation problem. It does not allow researchers to count sources B, C and D for the citation count measure. This is the second type of truncation problem.
Figure 3: Truncation problems of patent data - accuracy

This figure illustrates how truncation problems with patent data cause biases of the raw measures for evaluating recent innovations. In both panels, the x-axis represents the year the patent applications were filed. For the left panel, the y-axis represents the number of patents applied in year x that is granted as of 2016 (the black dots), or as of 2006 (the red dots). For the right panel, the y-axis represents the number of citations received by the patents that were filed in year x, and granted as of 2016 (the black dots), or as of 2006 (the red dots). Measures constructed after a ten year wait period suffers little from truncation problems, as most applications in figure 2 are either granted patent status, or are effectively dormant at that point. (Lerner and Seru (2015)) Thus, the black dots represent the benchmark to which short term measures compete to estimate. Therefore the vertical gap between black dots and red dots represent how much the amount of bias in the raw measure. The figure demonstrates that the bias is substantial for recent innovations (from the perspective of a researcher in 2006), and the bias increases as year x approaches 2006. The detailed procedure of the computation of the raw measure is described in section 2.

(a) A: Log patent count

(b) B: Ave. citations
Figure 4: Truncation problems solved by adding patent application data - accuracy

This figure illustrates how the truncation problem can be solved by including patent application data. The x, y axes, black dots and red dots are defined the same as figure 3. In both panels, the green dots represent the two measures computed using both patent and application data with the data available as of 2006 (called the “supplemented measures”). The vertical gap between each 2006 measure (red and green) and the benchmark represent how biased each measure is for patents filed in year x.

(a) C: Log patent count

(b) D: Ave. citations

Figure 5: Truncation problems solved by adding patent application data - precision

This table quantifies and contrasts the precision of the raw and supplemented measures by their Pearson correlations with the benchmark measure. p-value is reported in the parenthesis.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Correlation with the benchmark</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patent count</td>
</tr>
<tr>
<td></td>
<td>2001  2002  2003  2004  2005  2006</td>
</tr>
<tr>
<td>Supplemented Measure</td>
<td>0.9944 0.9888 0.9948 0.9965 0.9956 0.9947</td>
</tr>
<tr>
<td>(0.00)</td>
<td>(0.00) (0.00) (0.00) (0.00) (0.00) (0.00)</td>
</tr>
<tr>
<td>Raw measure</td>
<td>0.9898 0.9709 0.9434 0.9051 0.8201 0.6765</td>
</tr>
<tr>
<td>(0.00)</td>
<td>(0.00) (0.00) (0.00) (0.00) (0.00) (0.00)</td>
</tr>
<tr>
<td></td>
<td>Average normalized citation count per patent</td>
</tr>
<tr>
<td></td>
<td>2001  2002  2003  2004  2005  2006</td>
</tr>
<tr>
<td>Supplemented Measure</td>
<td>0.8646 0.7718 0.7304 0.6654 0.6155 0.5758</td>
</tr>
<tr>
<td>(0.00)</td>
<td>(0.00) (0.00) (0.00) (0.00) (0.00) (0.00)</td>
</tr>
<tr>
<td>Raw measure</td>
<td>0.4533 0.3048 0.1625 0.0427 0.0046 0.0108</td>
</tr>
<tr>
<td>(0.02)</td>
<td>(0.02) (0.02) (0.08) (0.85) (0.68)</td>
</tr>
</tbody>
</table>
Table 1: Summary statistics

This table reports the summary statistics for variables constructed based on the sample of U.S. public firms from 2011 to 2015. Panel A reports the summary statistics of innovation measures. Panel B reports the summary statistics of the percentage of votes that are for “once every three years”. Panel C presents the year distribution of SOP freq. voting. Panel D reports the industry distribution of firms with shareholder proposals. Panel E reports the summary statistics of firm characteristics that have shown to affect innovations.

Panel A: Innovation measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patents</td>
<td>22.37</td>
<td>0.37</td>
<td>162.54</td>
<td>20,223</td>
</tr>
<tr>
<td>Norm. Cites/Patent</td>
<td>0.20</td>
<td>0</td>
<td>0.44</td>
<td>20,223</td>
</tr>
<tr>
<td>Field Dispersion</td>
<td>0.76</td>
<td>0.85</td>
<td>0.26</td>
<td>11,602</td>
</tr>
<tr>
<td>Explor. Quotient</td>
<td>1.10</td>
<td>1</td>
<td>0.59</td>
<td>20,223</td>
</tr>
<tr>
<td>Market Reaction (Mil.$)</td>
<td>19.99</td>
<td>0</td>
<td>75.36</td>
<td>16,251</td>
</tr>
</tbody>
</table>

Panel B: Percentage of Votes that are for “once every three years”

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>25th</th>
<th>Median</th>
<th>75th</th>
<th>Passage Rate</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>All SOP freq. votes</td>
<td>31.2%</td>
<td>7.3%</td>
<td>15.6%</td>
<td>54.3%</td>
<td>28.3%</td>
<td>4,884</td>
</tr>
<tr>
<td>Those filed patents</td>
<td>29.2%</td>
<td>7.7%</td>
<td>14.7%</td>
<td>48.2%</td>
<td>27.2%</td>
<td>2,247</td>
</tr>
<tr>
<td>Close-call ATP proposals</td>
<td>47.5%</td>
<td>44.1%</td>
<td>49.7%</td>
<td>55.1%</td>
<td>51.5%</td>
<td>249</td>
</tr>
</tbody>
</table>

Panel C: Year Distribution of Shareholder Proposal Voting

<table>
<thead>
<tr>
<th>Year</th>
<th>All SOP freq. votes</th>
<th>Those filed patents</th>
<th>Close-call SOP freq votes</th>
<th>Ave. firm size Mil.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>3,328</td>
<td>1,673</td>
<td>162</td>
<td>1,665</td>
</tr>
<tr>
<td>2012</td>
<td>13</td>
<td>8</td>
<td>1</td>
<td>660</td>
</tr>
<tr>
<td>2013</td>
<td>929</td>
<td>340</td>
<td>71</td>
<td>181</td>
</tr>
<tr>
<td>2014</td>
<td>272</td>
<td>85</td>
<td>9</td>
<td>424</td>
</tr>
<tr>
<td>2015</td>
<td>204</td>
<td>75</td>
<td>9</td>
<td>440</td>
</tr>
<tr>
<td>2016</td>
<td>138</td>
<td>64</td>
<td>4</td>
<td>NA</td>
</tr>
<tr>
<td>Total</td>
<td>4,884</td>
<td>2,247</td>
<td>249</td>
<td>1,271</td>
</tr>
</tbody>
</table>
Panel D: Industry distribution of firms that voted on SOP frequency

<table>
<thead>
<tr>
<th>SIC</th>
<th>Description</th>
<th>All proposals</th>
<th>ATP-related Proposals</th>
<th>Close-call ATP Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Agriculture</td>
<td>11</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>Mining</td>
<td>266</td>
<td>55</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Light manufacturing</td>
<td>668</td>
<td>458</td>
<td>73</td>
</tr>
<tr>
<td>3</td>
<td>Heavy manufacturing</td>
<td>976</td>
<td>781</td>
<td>91</td>
</tr>
<tr>
<td>4</td>
<td>Transportation</td>
<td>326</td>
<td>115</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Wholesale trade</td>
<td>342</td>
<td>102</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Finance</td>
<td>1,016</td>
<td>148</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>Services</td>
<td>489</td>
<td>272</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>Health services</td>
<td>162</td>
<td>64</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>Public Administration</td>
<td>15</td>
<td>7</td>
<td>3</td>
</tr>
</tbody>
</table>

Panel E: Distribution of firm characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assets (Mil. $)</td>
<td>1,700</td>
<td>262</td>
<td>5547</td>
</tr>
<tr>
<td>R&amp;D/Assets (%)</td>
<td>2.3</td>
<td>1.1</td>
<td>7.14</td>
</tr>
<tr>
<td>PPE/Assets (%)</td>
<td>72.22</td>
<td>59.63</td>
<td>140.74</td>
</tr>
<tr>
<td>CapExp/Assets (%)</td>
<td>16.65</td>
<td>5.23</td>
<td>170.70</td>
</tr>
</tbody>
</table>
**Figure 6:** Internal validity - manipulation - density of vote shares

This figure plots the histogram of the distribution of the percentage of votes for “once every three years” SOP in our sample across 40 equally-spaced bins. The x-axis is the percentage of votes favoring a “once every three years” SOP frequency. The y-axis represents the fraction of firms whose shareholders votes in favor of this. Votes for “once every two years” are not counted. SOP frequency voting results are obtained from the ISS.

**Figure 7:** Internal validity - manipulation - McCrary test

This figure reports the result of the McCrary test (McCrary (2008)) of discontinuity of vote share density at 50%. The x-axis represents the percentage of votes favoring “once every three years” SOP. The y-axis represents the density estimate. The dots depict the density estimate, and the solid line represents the fitted density function of the x (the vote share) with a 95% confidence interval around the fitted line. The votes in favor of “once every two years” are not counted. SOP frequency voting results are obtained from the ISS.
Table 2: Internal validity - discontinuity of *ex ante* innovation measures

This table supports the validity of the RDD design by showing that *ex ante* values of innovation measures do not have discontinuities around the 50% vote share. I use 4-th order polynomial with 5th-order correctional term with an optimal bandwidth to estimate the discontinuous jump in the *ex-ante* outcome, following Imbens and Kalyanaraman (2012). Standard errors are presented in brackets. */**/*** indicates significant at the 10%, 5% and 1% levels, respectively.

<table>
<thead>
<tr>
<th>Log patent count</th>
<th>Ave. normalized citation count</th>
<th>Market reaction on grant date</th>
<th>Dispersion of tech. field</th>
<th>Exploration Quotient</th>
<th>Log Total Compensation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>-0.5042</td>
<td>-0.645</td>
<td>-0.112</td>
<td>0.089</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>[0.576]</td>
<td>[1.168]</td>
<td>[0.140]</td>
<td>[0.114]</td>
<td>[0.145]</td>
</tr>
<tr>
<td>Obs</td>
<td>2056</td>
<td>2056</td>
<td>878</td>
<td>1120</td>
<td>1226</td>
</tr>
</tbody>
</table>

Figure 8: Internal validity - discontinuity of *ex ante* innovation measures - RD plots

This figure is the RD plot generated during the procedure described table 2. The x axis represents vote shares in favor of “once every three years”. The y axis represents the density of each vote share. Each dot represents the density averaged over a 0.25% bin. The interval around it is the 95% confidence interval of the mean. The curve is the fitted 4th order polynomial on either side of the cutoff.
This figure plots the implementation of the SOP frequency voting results. The x-axis is the percentage of votes favoring “once every three years” SOP. The y-axis is the SOP frequency implemented, where “1” represents “once per year” and “3” represents “once every three years”. Each dot represents the SOP frequency voting of one firm in my sample. The votes in favor of “once every two years” are not counted. SOP frequency voting results are obtained from the ISS. Implementation data is collected by ISS from SEC 8-K filings.
Table 3: External validity - distribution independence of vote share and firm characteristics

This table reports the distribution of firm characteristics within the sub-sample of close-call firms. Comparison with the Panel E of table [1] provides suggestive evidence that the distribution of firm fundamental characteristics is similar in the close-call vote sample with that of the entire sample.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assets (Mil. $)</td>
<td>915.8</td>
<td>56.6</td>
<td>5527</td>
</tr>
<tr>
<td>R&amp;D/Assets (%)</td>
<td>3.1</td>
<td>1.8</td>
<td>6.20</td>
</tr>
<tr>
<td>PPE/Assets (%)</td>
<td>117.7</td>
<td>45.13</td>
<td>263.4</td>
</tr>
<tr>
<td>CapExp/Assets (%)</td>
<td>8.82</td>
<td>3.61</td>
<td>22.65</td>
</tr>
</tbody>
</table>
Table 4: Dynamics of patent value

This table is generated following the same RD measure as table 2. The measures used are the market value of patents and average normalized citation count that are filed in the first, second and third year after the SOP frequency vote and SOP vote.

<table>
<thead>
<tr>
<th>Patent Market Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>first year</td>
</tr>
<tr>
<td>$\beta$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Obs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ave. Norm. Citation Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>first year</td>
</tr>
<tr>
<td>$\beta$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Obs</td>
</tr>
</tbody>
</table>

Figure 10: Dynamics of Patent Value

This figure is the RD plot generated during the procedure described table 2. The x axis represents vote shares in favor of “once every three years”. The y-axis represents the market value of patents and average normalized citation count that are filed in the first, second and third year after the SOP frequency vote and SOP vote.
Table 5: Dynamics of log patent count

This table is generated following the same RD measure as Table 2. The measure tested is the log patent count.

<table>
<thead>
<tr>
<th></th>
<th>first year</th>
<th>second year</th>
<th>third year</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.074</td>
<td>0.245</td>
<td>0.422</td>
</tr>
<tr>
<td></td>
<td>[0.455]</td>
<td>[0.359]</td>
<td>[0.513]</td>
</tr>
<tr>
<td>Obs</td>
<td>2056</td>
<td>2353</td>
<td>2060</td>
</tr>
</tbody>
</table>

Figure 11: Dynamics of log patent count

This figure is the RD plot generated during the procedure described table 2. The x axis represents vote shares in favor of “once every three years”. The y-axis represents the number of patents that are filed in the first, second and third year after the SOP frequency vote and SOP vote.
The x axis represents vote shares in favor of “once every three years”. The y-axis represents the Herfindahl dispersion of the technological fields in which the firm files patent, as well as the ratio of the number the forward self-citation divided by the backward self-citation.

<table>
<thead>
<tr>
<th>Herf</th>
<th>first year</th>
<th>second year</th>
<th>third year</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>0.223*</td>
<td>0.108</td>
<td>-0.112</td>
</tr>
<tr>
<td></td>
<td>[0.129]</td>
<td>[0.080]</td>
<td>[0.140]</td>
</tr>
<tr>
<td>Obs</td>
<td>763</td>
<td>1610</td>
<td>1125</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Explore</th>
<th>first year</th>
<th>second year</th>
<th>third year</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>0.233*</td>
<td>-.009</td>
<td>-0.254</td>
</tr>
<tr>
<td></td>
<td>[0.135]</td>
<td>[0.153]</td>
<td>[0.166]</td>
</tr>
<tr>
<td>Obs</td>
<td>1688</td>
<td>2100</td>
<td>2100</td>
</tr>
</tbody>
</table>

This figure is the RD plot generated during the procedure described table 2. The x axis represents vote shares in favor of “once every three years”. The y-axis represents the Herfindahl dispersion of the technological fields in which the firm files patent, as well as the ratio of the number the forward self-citation divided by the backward self-citation.
Figure 13: Shifting of the technological focus

This figure is the RD plot generated during the procedure described in Table 2. The x-axis represents vote shares in favor of “once every three years”. The y-axis represents the $L^2$ norm of the change of distribution of the patents in technological fields.
Table 7: Incentive enacted by SOP

This table is the results of estimation of regression equation: \( v_{j,t}^S = \alpha + \beta_1 EEP_{j,t} + \beta_2 MV_{j,t} + \gamma ISSrec_{j,t} + \zeta Z_{j,t} + \epsilon_{j,t} \).

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Inst. Ownership</th>
<th>leverage</th>
<th>ISSRec</th>
<th>Exploration</th>
<th>Cite Count</th>
<th>Market Value of Patent</th>
</tr>
</thead>
<tbody>
<tr>
<td>coefficient</td>
<td>0.00**</td>
<td>0.009</td>
<td>0.033***</td>
<td>-0.000</td>
<td>-0.000*</td>
<td>0.002*</td>
</tr>
<tr>
<td>p-value</td>
<td>(0.037)</td>
<td>(0.646)</td>
<td>(0.000)</td>
<td>(0.919)</td>
<td>(0.084)</td>
<td>(0.100)</td>
</tr>
</tbody>
</table>