Fall 5-18-2018

Technological Diversity in Finance

Blake K. Rayfield
University of New Orleans, bkrayfie@uno.edu

Technological Diversity in Finance

A Dissertation

Submitted to the Graduate Faculty of the
University of New Orleans
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Financial Economics

Doctor of Philosophy in Financial Economics

by

Blake Rayfield

B.B.A. Florida International University, 2014
M.S. University of New Orleans, 2016

May 2018

Dedication

To my mothers Tina Rayfield and Letty Cortez and my grandmother Sylvia. Without you, I would not be where I am today. Thank you for your love and support over the years in my pursuit of my Ph.D.

Acknowledgement

I would like to thank my peers and faculty for this accomplishment. I am most grateful to my peers Jennifer Brodmann, Hasib Ahmed, and Ben Wuthisatian. Thank you for all the late nights and countless hours of debate and conversation. I am grateful to my family and friends for their support, and kindness. Special thanks go to my dissertation committee, main advisor and co-chairs Dr. M. Kabir Hassan and Dr. Neal Maroney. I would also like to thank two people who go out of their way to make sure I am on track, Dr. Duygu Zirek, and Dr. Omer Unsal. Lastly, I would like to thank Dr. Linxiong Li for being a professor I aspire to be.

Table of Contents

List of Figures

List of Tables

Abstract

The dissertation consists of two chapters on measuring firms technological profile. Patent data can be grouped into two primary generations. The first generation lead by the work of Schmookler (1966), Scherer (1982), and Griliches (1984), and the second generation led by Trajtenberg, Jaffe, and Henderson (1997) and Kogan et al. (2016). When combined, both generations data spans from nearly 1926-2010 and has made a meaningful impact on innovation research. In the first chapter, I propose a third generation of patent data. The third generation of patent data has two distinct contributions. First, it extends patent-firm ownership information beyond 2010 to 2016. The new dataset uses the established connections of previous datasets and builds on that information with additional data on firm names gathered from EDGAR. Second, it takes advantage of the information contained in the text of patents using text analysis. Using text analysis allows for greater flexibility over traditional measures. The second chapter investigates how ownership structure affects firm value. The previous literature has assumed more innovation is better, meaning the more innovation a business creates; the better off it is in the long-run. However, not all innovations are created equal. We contribute to the literature by investigating how institutional investors change future innovation, not in quantity, but diversity. Using several unique measures of technological diversification created from firm-level patent data, we show that institutional investors increase the focus on a firm's future innovation. Our results are robust to the classification scheme. Ultimately, our results indicate institutional investors create value by encouraging firms to build on prior knowledge.

Keywords: Innovation; Corporate Finance; Text Analysis; Ownership Structure; Institutional Investors

Chapter 1: Generality, Originality and Firm Value: A Text Based Approach.

Abstract

This chapter investigates the ability of textual content to classify patents as "general" or "original." We introduce a new measure of patent generality and originality that takes advantage of the information content of the text contained in patents. Our new measure captures two features not previously grasped by similar measures, such as that of Jaffe and Henderson (1997). First, our measure accounts for the distance between technological classes. Second, our measure has the ability to capture the similarity of patents within classes. Lastly, we introduce a revised dataset of firm level patent information that extends previous datasets.

1.1 Introduction

The goal of this paper is to create and describe an improved dataset based on U.S. patents. The dataset provides researchers an expanded perspective on firm innovation that can be used widely for research. Prior data on U.S. patents has become outdated, and researchers are identifying limitations. This paper seeks to overcome those limitations by improving upon the prior datasets. The three primary contributions of this paper are; an updated link between patents and firm owners, a new measure of patent originality based on the patent text, and a widely accessible dataset for use in research.

Patent count as a measure of innovation has captured the imagination of finance and economics researchers because of its unique features. A patent holder has exclusive rights to seek a return on investment; no other firm can produce its product. This exclusive right incentivizes firms to report innovation through patents, further strengthening the argument for patents as a measure of innovation.

In the corporate finance literature, patents proxy for an indirect measure of innovation. However, researchers find patents a more reliable indicator of firm-level innovation than R&D expenditures. To take advantage of the informational content delivered by patents, researchers have transformed patent data into firm-level innovation datasets. These datasets can be classified into two distinct categories: Generation One marked by simple patent counts of firms that were hand collected, and the most recent Generation Two marked by citation weighted patent counts and a more reliable firm-patent connection.

Generation Two adds value to the patent data transforming the data to create new measures. Specifically “generality” and “originality,” as introduced by Trajtenberg, Jaffe, and Henderson

(1997) are created from the link between patent citations. These measures, along with others, have made a significant contribution to the literature. However, more recently, researchers have identified several weaknesses that plague the interpretation as well as the reliability of this measure. For example, "generality" and "originality" rely on heterogeneous technological classifications that are subject to evolution over time. We introduce a new measure of patent generality and originality that takes advantage of the information content of the text contained in patents. This new measure does not rely on patent classification schemes, nor is it affected by changing technology over time.

Trajtenberg (1990) introduced the use of weighted citations as a measure of economic value, arguing that patents cited by innovations in various asset classes have greater "generality." Citation weighted measures are more reliable, as they better represent innovation output, whereas a simple patent count represents input. Hall, Jaffe, and Trajtenberg (2005) find weighted citation measures to be more correlated with firm value than patent counts.

To improve upon Jaffe and Henderson's (1997) measure of "generality" and "originality," we use the textual information contained in patent documents. These measures rely on the simple assumption that similar patents use similar language. Because our measure does not rely on the classification of technological classes by the US patent office, it has two distinct advantages in comparison to Jaffe and Henderson's (1997) measure. First, our measure accounts for the distance between technological classes. Second, our measure has the ability to capture the similarity of patents within classes.

In the next section, we investigate the data available for patent researchers. Next, we proceed with a review of patent and text analysis literature, as well as its growing importance in finance. We then proceed to review the methodology used to create our measure of generality as well as our extended patent-ownership dataset. Finally, we illustrate the effectiveness of our measure by demonstrating its ability to explain the similarity between technological classes as well as within the class.

1.2 Patents as a Data Source

The use of patents as a data source is not new. Patents have long been identified as an important data source for measuring innovation impact, as indicated by the several researchers committed to creating, expanding, and maintaining patent datasets including Griliches (1984); Hall, Jaffe, and Trajtenberg (2001); Kogan, et al. (2016); Li, et al. (2014) among others.

The reliable use of patents as a proxy for innovation grew from the unreliability of R&D expenditures. R&D is considered unreliable for several reasons. Those highlighted in Lerner and Seru (2015) include the issue firstly that firms only need report R&D if expenditures are material. This can lead to uncertainty and heterogeneous reports across firms and industry because the definition of material is left to interpretation by the parent firm. Secondly, R&D expenditures are not broken down by product line. Lastly, patents, as the product of a firm's R&D, are more representative of innovation output than input. Beyond the nature of R&D reporting, the integrity of the data can also be unreliable.

Patents as a data source can be grouped into two distinct generations as we discuss in the forthcoming section. Generation One relied upon simple patent count. Therefore, the data used by

early patent researchers only used the names on patents and the quantity assigned to each firm. Later, in the second generation, researchers begin to take advantage of additional information, such as citation, patent class, among others. The improvements between generations can be broadly summarized by taking advantage of the additional information contained in the patent. In the next section, we describe in the process firms to be awarded a patent, as well as the information provided in a patent.

1.2.1 The Patent Award Process and the Information They Contain

Nearly all firms and industries hold patents. There are three different types of patents. Utility, Design, and Plant patents. As prior studies have, this study will focus on the Utility patent. The USPTO describes the Utility Patent as "[a] new and useful process, machine, article of manufacture, or compositions of matters, or any new useful improvement thereof."

A Utility patent has several key sections. The sections this study focuses on include: The Patent Number, Application Date, Award Date, Inventor, Assignee, References Cited, Abstract, Claims, and Description. Each patent has an Inventor, the individual whom files the patent, while many (but not all) have an Assignee, the firm or individual which the intellectual property belongs[1]. Citations, play a crucial role in determining the impact of innovations, as indicated by the earlier research. These citations play a critical legal role, as they define the scope of an invention as well as having possible legal implications in the future. Alcacer and Gittelman (2006) find that more than 60% of citations are selected by examiners. Patents that cite other patents in a broader array of technology classes are often viewed as having more "originality." Patents that are themselves

[1] A detailed description of the sections in a Utility patent may be found at https://www.uspto.gov/patents-getting-started/patent-basics/types-patent-applications/nonprovisional-utility-patent#heading-18 .

cited by a more technologically dispersed array of patents are viewed as having greater "generality." In the following section, we create a measure of originality and generality that uses text analysis to circumvent the need for technology classes. This study focuses on the inclusion of the data contained in the Abstract, Claims, and Description sections.

Patent applications submitted to the U.S. Patent and Trademark Office (USPTO) consist of claims and other supporting documentation. Patent applications have three primary targets for text analysis; Abstract, Claims, and Description. Both the Abstract and Description give broad descriptions of the individual intellectual property claim. Where the section, Claim describes in detail precisely what process, material, or other is being claimed. Some of the claims in a patent application will be cast in concrete terms; others may be sweeping.

1.3 Literature Review

The use of patent data in finance began with the early work by Schmookler (1966), Scherer (1982), and Griliches (1984). This literature was groundbreaking, as it began the use of citation counts as a proxy for industry and firm innovation activity. However, researchers soon began to realize that the drawback of simple patent citation counts was two-fold; first, simple citation count was unable to capture the extreme heterogeneity displayed by true innovation value. Second, simple citation counts are only a fraction of the total information available in patent documents, according to Griliches, Hall and Pakes (1987).

These drawbacks lead to the development of new measures, such as the weighted citation measure of Trajtenberg (1990). The weighted citation measure was developed to specifically address the problem of heterogeneous patent and citation activity across industry. Trajtenberg shows how

referenced citation counts can proxy for the importance of a patent in a specific technology field, or with alterations, counts can capture technology spillover.

To further expand the usefulness of information contained in patent data Trajtenberg, Jaffe and Henderson (1997) created measures of "generality" and "originality." These two new measures were created with the intention of further capturing the importance of innovation on a patent and firm level. For example, Jaffe and Henderson define generality as:

$$Generality_t = 1 - \sum_{j}^{n_i} s_{ij}^2 \quad (1)$$

Where s_{ij}^2 is the squared percentage of citations received by patent *i* that belong to patent class *j*, out of n_i asset classes. This measure is known as a forward-looking measure, as it considers the impact of the patent post-filing date. Jaffe and Henderson's measure for originality is defined similar to Equation 1, however, it is a backwards-looking measure that takes into account citations made by the target patent.

Both measures drive and inspire a diverse set of literature in corporate finance. Bernstein (2015) uses both measures to show how public firm innovation becomes less novel after IPO, but that firms begin to acquire innovation rather than create it. Acharya and Xu (2016) investigate the relation between innovation and a firm's financial dependence using a sample of privately held and publicly traded US firms. Amore, Cedric, and Zaldokas (2013) investigate the impact of interstate banking deregulation on innovation activity, where they find that interstate banking deregulation has had a significant, positive effect on the originality of patents.

Our methodology is not unlike the early work of Scherer (1982), who develops and uses a classification scheme by manually analyzing patent text and creating a technology flow matrix.

However, in contrast to Scherer (1982), we use a computerized algorithm to compute backward (originality) and forward (generality) similarity measure using text analysis. Our new measure of forward and backward similarity does not rely on changing the definition/the changing definitions of technology sector and accounts for heterogeneous distances between sectors, unlike the measure proposed by Trajtenberg, Jaffe and Henderson (1997).

The use of text analysis in financial economics can be found in an extensive survey by Loughran and McDonald (2016). In short, it is the process of converting the informational content contained in text to quantitative data. Researchers have used text analysis to investigate asset returns, such as Frazier et al. (1984), Antweiler and Frank (2004), Das and Chen (2007), Tetlock (2007), and Li (2008), readability Jones and Shoemaker (1994) and Li (2008), and industry groupings Hoberg and Phillips (2010).

Li (2008) connects the readability of a firm's financial reports to that firm's performance. This paper measures the FOG index, a measure of readability, of a firm's financial reports. The author finds financial reports with more complex language belong to firms with lower profitability ratios, as compared to firms with more simple reports and higher returns.

The most closely related paper to our research is that of Hoberg and Phillips (2010). The authors segment firms by product markets, based on key-words in the business description section of their 10-K filings. We propose that our measure of originality also captures the spirit of Hoberg and Phillips (2010). Patents represent real options of future product creation activity, so firms that hold a diverse set of patents hold a diversified set of real options. For example, differentiation in product markets produces less volatile cash flows, and reduces firms' risk and expected returns (Hou and Robinson (2006)).

1.4 Methodology and Data

1.4.1 Patent Similarity

Our new measure of patent generality addresses a complication of patent data as raised by Lerner and Seru (2015), namely, "the failure to adjust for the technological class of the discovery." The USTPO currently hosts approximately 475 total classes and 165,000 total subclasses. The technology of each industry and class evolves over time, as can be seen using the text analysis application by Packalen and Bhattacharya (2012). However, its classification remains the same.

This complication proves particularly challenging when considering how weighted patent counts are constructed. For example, consider Trajtenberg's (1990) simple linear weighted patent count measure:

$$WPC_t = \sum_{i=1}^{n_t} (1 + C_i) \tag{2}$$

Where *n* is the total number of patents issued during year *t* in one specific product class. Because of the ever-changing nature of technology, using this measure across decades becomes challenging.

To alleviate this complication, we introduce a simple linear measure of likeness that can be used to measure a patent's generality and originality.

To compute a new text-based measure of patent similarity, we use text analysis as a technique aimed at converting text data into qualitative data that can be more easily used by the researcher.

First, we gather the full text of all U.S. utility patents granted from 1926 to 2010 from the U.S. Patent and Trademark Office (USPTO). All patent information gathered is publicly available information and contained in the USPTO's bulk data files[2]. While data is available from 1975 or earlier, the text is scanned using an Optical character recognition process and is extremely unreliable and unstructured. Previous researchers have used data pre-1975, such as Packalen and Bhattacharya (2012). Our sample comprises 4,131,597 patents spanning 1976 thru 2010.

Once all patent text is gathered, it is processed, filtered through extensive cleaning processes, and converted into tokenized text. The tokenized text is the process of converting documents (patents) into word vectors and stripped of all punctuation. For example, the simple sentence "Mary had a little lamb." would be converted into the following tokenized text.

(3) [Mary, had, a little, lamb]

We then further select words and phrases made up of only noncapitalized English nouns, and we remove all words lacking information content, called stop words. Stop words are devoid of any informational material (such as: and, it, or) and therefore contribute very little to the understanding of similarities between two documents. Revisiting the example (3), the final output would now look like (4)

(4) [little, lamb]

[2] https://www.uspto.gov/learning-and-resources/bulk-data-products

We then further remove patents containing less than 20 unique words because patents that contain less may not have sufficient content to justify similarity.

Once the text is cleaned, it is then converted into vectors, where each word receives a count for its frequency of use in each document. Each vector is then normalized to unit length; then the cosign similarity is calculated from the dot product of two vectors:

$$SIM_{i,j} = V_i \cdot V_j' \tag{4}$$

Where V_i represents the term frequency normalized vector for patent i and $SIM_{i,j}$ is the cosign similarity between documents i and j.

Cosign similarity is a measure of similarity between two vectors, as it measures the cosign of the angle between them. The resulting value bound by [1,-1] where 1, the two documents are the most similar and -1, the two documents are complements. Because we have created vectors using only positive counts of n-grams words, our cosign similarity is bound by [1,0].

To illustrate the final result, Figure 1 shows the mean similarity score for patent citations over time, as well as the number of citations that occur after a patent's filing date. The number of citations after the filing date resembles a Poisson distribution curve. A Poisson type shape occurs because of truncation by the start of the sample--there is lag between filing and granting a patent that is currently between 2-5 years.

Figure 1-1: Mean Similarity for sample overtime (Forward Looking)

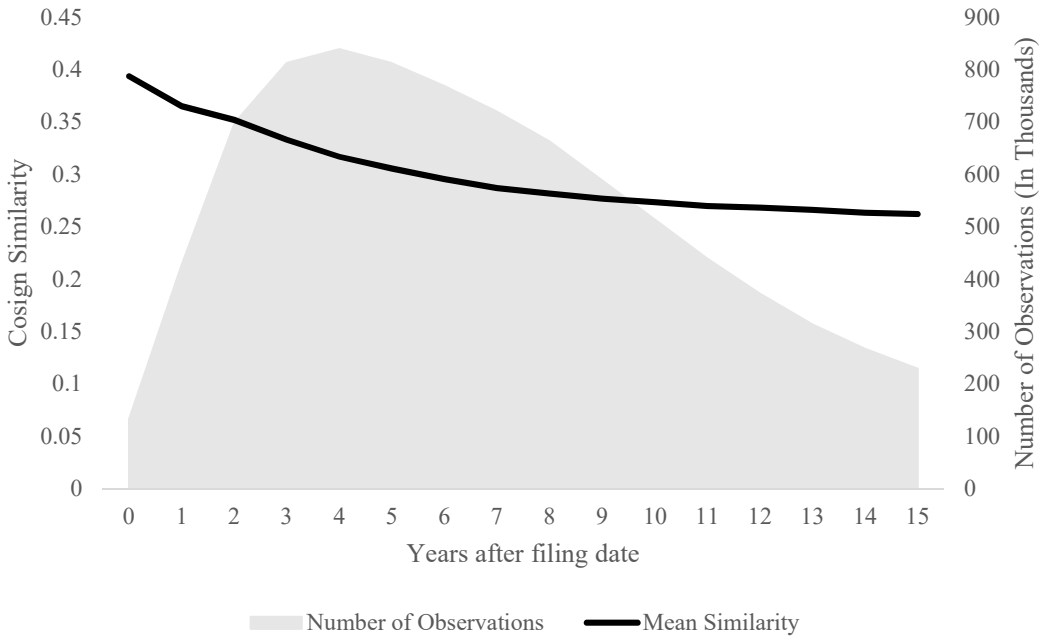
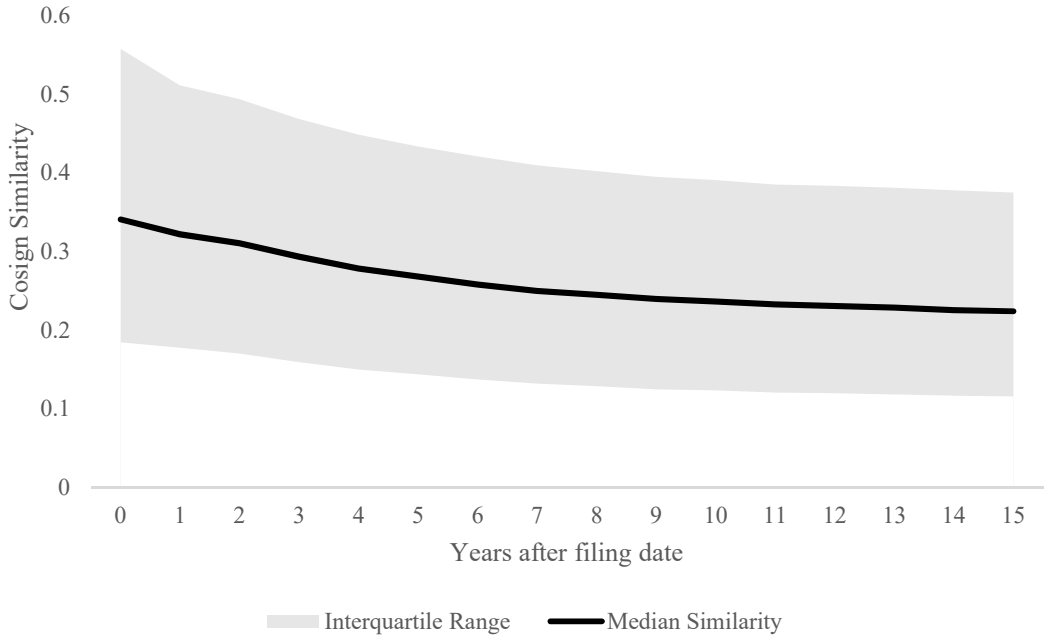


Figure 1-2 displays the median similarity score for cited and citing patents, along with bands representing the 75 and 25 percentiles. The interquartile range decreases overtime each year, while the median follows a similar path of the mean, starting around .4, while slowly reducing to a steady state around .25. The results indicate that over time, patents tend to drift away regarding similarity. This highlights an issue with the traditional patent analysis because technology changes over time, but technological classes remain the same, there may be uncaptured variation in the similarity of patents within a technology class.

Figure 1-2: Median Similarity for sample overtime (Backward Looking)



Both Figure 1-1 & 1-2 show the pairwise similarity shows meaningful results. Overtime patents lose their similarity to prior patents, supporting the need for eradicating stagnant patent classes. With pairwise correlations, we are able to investigate the similarity right away of patents over time. In section 1.4.3 we exploit the pairwise correlations to determine the likeness of a firm's patent portfolio.

1.4.2 Firm Ownership Information

Matching firms to their owners is no simple task. Several problems arise when trying to discern the ownership of any specific patent. The literature of Kogan et al. (2016) and Hall, Jaffe and Trajtenberg (2001) have made substantial upgrades to the link between firms and their patents. Both groups have faced the same complications when matching firm names to patent owners. For example, some firm names have several variations. International Business Machines may be listed

in numerous ways, such as, IBM, Inter Business Machine, among others. Over time, more data has been gathered that can aid in the process of ownership matching. A contribution of this paper is to extend the traditional datasets.

To connect patents to firms, we employ three well-established datasets as well as one additional source. First, we gather the developed firm-patent data sets provided by Kogan et al. (2016), and the NBER patent database of Hall, Jaffe, and Trajtenberg (2001). The Hall, Jaffe and Trajtenberg (2001) dataset links patents to firms for the years 1976-2006. The Kogan et al. (2016) dataset links patent ownership to firms from 1926-2010. By combining the two datasets, we take advantage of the prior firm matching. To further identify patent owners, we use the ownership identification of Lai (2013). This dataset uses Bayesian methods to match similar names/owners across patents over time.

The resulting dataset still has a maximum year of 2010. For researchers, extending the data beyond 2010 should be a priority. Therefore, we employ a novel approach. Each publicly traded firm must list their subsidiaries in their 10-K annual filings, Exhibit 21 on EDGAR. Using the information on firm subsidiaries, we are able to extend the link between firms and patents beyond the year 2010. Furthermore, the firm subsidiary information provides matches in years before 2010 that were not previously available.

We index patents by their application year, as is standard in the literature. Once matching with our corpus, our sample consists of 948,881 cited patents and 2,187,103 citing patents for a total of 3,135,984 patents. While we have several more patents available in our corpus, we limit the sample to those we can match to a cited patent assigned to a firm in Compustat.

1.4.3 New Variable Creation

As discussed previously, the text of a patent contains valuable information relevant to a patents similarity. In section 1.4.1 we were able to compute the pairwise similarity of two patents. With this information, we will create a new measure of patent originality that solves some of the problems identified in the prior literature. Jaffe and Henderson define originality as:

$$\text{Originality}_{\text{t}} = 1 - \sum_{j}^{n_i} s_{ij}^2 \tag{5}$$

Where s_{ij}^2 is the squared percentage of citations made by patent *i* that belong to patent class *j*, out of n_i asset classe. This measure is known as a backward-looking measure. However, this measure has many weakness that may hamper its use in empirical research. First, the measure assumes the distance between each class is equal. For example, if the original patent related to construction equipment the measure gives the same weight for a citation in technology as it does the jewelry category. Under certain circumstances, the difference between those categories can be drastically different. Next, the measure is undefined when the patent awards no citations outside of its patent class. This weakness prevents the measure from being used for single industry or inter-industry studies.

We propose a new measure based on the text of the patents. The measure is defined as follows:

(6) $$\text{Text Originality}_{\text{t}} = 1 - \sum_{j}^{n_j} p_j \sum_{i}^{n_i} s_{ij}^2$$

Where s_{ij}denotes the average cosign similarity for patent i's citations that belong to patent class j and, p_{ij} represents the proportion of citations for patent i's citations that belong to patent class j.

The new measure based on text solves both problems discussed earlier. First, classes are weighted in similarity by p_j. Therefore, no longer are the distances between classes treated as equal, but rather they are a function of the average similarity between the two classes. Next, the new variable is defined for patents that only cite their own class. This allows us to capture variation previously not studied.

The benefit of the new measure is depicted in Table 1-1.

Table 1-1: Originality Summary Statistics

	HJT	Text Based
Panel A.		
Mean (All)	0.52	0.31
Std (All)	0.35	0.41
Panel B.		
Mean (Single Cite)	X	0.60
Std (Single Cite)	X	0.39

Table 1-1 shows the summary statistics for both the HJT measure of originality as well as the text-based measure of originality. In Panel A, we compute the mean and standard deviation of the originality measures. We can see that the text-based measure has a lower mean, but higher standard

deviation. This text based originality measure may be weighting citations within the class differently than the traditional HJT measure. Panel B results are most impactful. In this panel, we compute the mean and standard deviation of the originality measure for only patents that cite a single class. The HJT measure is undefined, meaning we are unable to use this measure to investigate single industry studies. However, we are able to compute the inter-class similarity using the text-based measure. This is because the text-based measure is relying on pairwise similarities as well as class similarities. Therefore, our measure does not become undefined if the patent only cites one class. A similar measure can be created for generality.

After creating a similar measure for generality, we conduct a simple investigation of the properties of the new measure. Table 1-2 shows the underlying correlations between both the text-based originality measure and generality measure with essential financial variables. The inclusion of this correlation table is intended to give the reader an idea of the unconditional correlations between the new variable and standard variables used in an economic or financial analysis. Generality and firm age are positively correlated; we think of firms in their later years as having more general innovations and stable cash flows. Originality and firm age are also negatively correlated; this too reconciles with the literature which suggests that young firms create innovative products to capture attention and market share.

Table 1-2: Correlation Table

	Originality	Generality	Tobin's Q	Total Assets	Market Value	Book Value	Book/ Market	Firm Age
Originality		0.02*	0.08*	-0.03*	-0.05*	-0.03*	0.01*	0.16*
Generality	0.02*		0.05*	-0.05*	-0.03*	-0.07*	-0.02*	0.1*
Tobin's Q	0.08*	0.05*		-0.09*	0.17*	-0.02*	-0.09*	0.33*
Total Assets	-0.03*	-0.05*	-0.09*		0.59*	0.74*	-0.01*	-0.63*
Market Value	-0.05*	-0.03*	0.17*	0.59*		0.83*	-0.05*	-0.09*
Book Value	-0.03*	-0.07*	-0.02*	0.74*	0.83*		0.1*	-0.63*
Book/Market	0.01*	-0.02*	-0.09*	-0.01*	-0.05*	0.1*		0.02*
Firm Age	0.16*	0.1*	0.33*	-0.63*	-0.09*	-0.63*	0.02*	

Simple correlation between similarity measure and other variables. * represents .01 level of significance.

1.5 Inter-industry study using a new originality measure

To highlight the advantage of our new measure, to capture uniqueness in the same technology class, and to illustrate its effectiveness in capturing patent originality, we show its ability to predict abnormal returns. We gather FDA drug approval dates from FDA's orange book[3]. This data contains approval dates for FDA approved drugs from 1981 onward. Included in the data is specific patent numbers associated with the product under review.

We use a simple event study to show the value of the new measure. We highlight the application of our measure in pharmaceuticals because many drugs are categorized in the same category. First, we estimate the abnormal return and market reaction associated with FDA drug approval. For this estimation, we use CRSP value-weighted and equal weighted returns and the Scholes-Williams Market Model to estimate cumulative abnormal returns. We use [-60,20] for the estimation window and [-20,20] for the event window. Table 1-3 shows the result for equal weighted returns around several event windows, as well as Figure 1-3. The results show an overwhelmingly positive reaction following the announcement of a new drug, as to be expected.

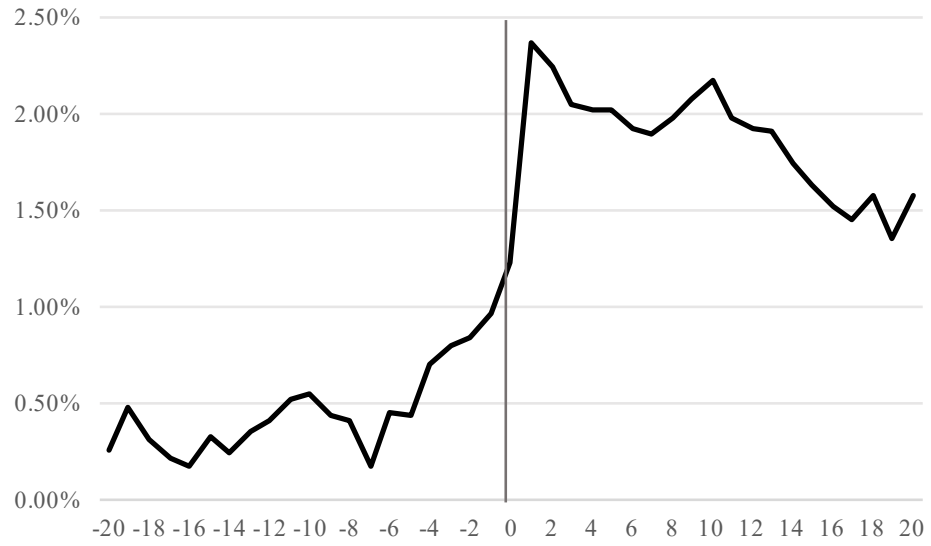
[3] Data and description can be found at http://www.fda.gov/Drugs/InformationOnDrugs/ucm129689.htm.

Table 1-3: CAARs using Scholes-Williams Market Model

	CAAR	T-Stat	P-value
[-1,1]	0.0141	4.85	0.00
[-2,2]	0.0115	3.06	0.00
[-3,3]	0.0104	2.34	0.02
[0,1]	0.0128	5.41	0.00
[0,2]	0.0106	3.66	0.00
[0,3]	0.0087	2.59	0.01

Results of Scholes-Williams CAAR regressions. Event times represent FDA approval of patented pharmaceutical products. Value weighted returns are used.

Figure 1-3: CARR after FDA Approval of Patented Product



We use the abnormal returns from Table 1-3 and regress them on our originality measure and several other control variables. We include firm and year-fixed effects, the results of which are shown in Table 1-4. A negative sign on the coefficient for originality would indicate that more original patents cause a larger market reaction.

Table 1-4: Results from Regressing CARR on Originality

	(1 Value Weighted) CARR[0,3]	(2 EQ Weight) CARR[0,3]
Originality	0.0485*	0.0285*
	(0.024)	*(0.044)*
Backwards Citations	0.0002	0.0030
	(0.922)	*(0.622)*
Log(at)	0.0119	0.0459
	(1.477)	*(0.987)*
B/M	-0.0711	-0.0312
	(-1.229)	*(-1.239)*
Firm Effects	*Yes*	*Yes*
Year Effects	*Yes*	*Yes*
R-squared:	0.815	0.843
Log-Likelihood:	487.63	497.23

Result from regressing CARRs as estimated in table 1-4 on *Originality,* the patent originality variable as computed from formula 6, *Backwards Citations, Log Total Assets, Book-to-Market.* Each regression includes Firm and Year effects. T-statistics are provided in parenthesis.

The results show that originality is associated with a larger market reaction, indicating that as a firm gets a more original product approved, it is associated with a larger market reaction. We do not propose that the full market reaction is explained by originality, but rather it is a demonstration of the capability of the originality measure.

1.6 Conclusion

Because patents are an important measure of innovation for firms, maintaining a functional patent dataset is important for future literature. Furthermore, we provide a new dataset that introduces a new measure of patent generality and originality that takes advantage of the information content

of the text contained in patents. Our new measure captures two features of the data not captured in Jaffe and Henderson (1997) - the distance between technological classes, and the similarity within classes. These two features make our measure more ideal for use, because it can detect heterogeneity of originality within industry as well as across it.

Additionally, we extend the patent-ownership data nearly one additional decade from prior datasets. The addition of this time frame will allow researchers to investigate important events of the last decade, such as, the financial crisis. Our dataset is publicly available for research.

To illustrate our originality measure's effectiveness, we demonstrate its ability to explain abnormal returns from FDA drug approvals. One of the benefits of our new measure is capturing the variation in originality within one patent class. Previous measures were unable to capture this variation because they relied upon USPTO classifications. The measure shows how more original FDA drugs are associated with larger abnormal returns.

References

Acharya, V., & Xu, Z. (2016). Financial dependence and innovation: The case of public versus private firms. The Journal of Finance.

Alcacer, J., & Gittelman, M. (2006). Patnet citations as a measure of knowledge flows: The influence of examiner citations. The Review of Economics and Statistics, 774-779.

Amore, C., & Zaldokas. (2013). Credit supply and corporate innovation. Journal of Financial Economics, 835-855.

Antweiler, W., & Frank, M. (2004). Is all that talk just Noise? The Information Content of Internet Stock Message Boards. Journal of Finance, 1259-1294.

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

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. Management Science, 1375-1388.

Frazier, K. B., Ingram, R. W., & Tennyson, B. M. (1984). A Methodology for the Analysis of Narrative in Accounting Disclosures. Journal of Accounting Research, 318-331.

Griliches, Z. (1984). R&D, Patents, and Productivity, NBER Conference Proceedings. University of Chicago Press.

Griliches, Z., Hall, B., & Pakes, A. (1987). The value of patents as indicators of inventive activity. In P. Dasgupta, & P. Stoneman, Economic Plicy and Technological Preformance (pp. 97-124). Cambridge, England: Cambridge University Press.

Hall, B., Jaffe, A., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools. National Bureau of Economic Research.

Hall, B., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. Rand Journal of Economics, 16-38.

Hoberg, G., & Phillips, G. (2010). Product market synergies and competition in mergers and acquisitions: A text-based analysis. Review of Financial Studies, 3773-3811.

Hou, K., & Robinson, D. (2006). Industry concentration and average stock returns'. The Journal of Finance, 1927-1956.

Jones, M. J., & Shoemaker, P. A. (1994). Accounting Narratives: A Review of Empirical Studies of Content and Readability. Journal of Accounting Literature, 142-184.

Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2016). Technological innovation, resource allocation, and growth. NBER.

Lerner, J., & Seru, A. (2015). The use and misuse of patent data: Issues for corporate finance and beyond. Booth/Harvard Business School Working Paper.

Li, F. (2008). Annual Report Readability, Current Earnings, and Earnings Presistence. Journal of Accounting and Economics, 221-247.

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. Retrieved from SSRN: http://dx.doi.org/10.2139/ssrn.2504147

Packalen, M., & Bhattacharya, J. (2012). Words in patents: Research inputs and the value of innovativeness in invention. NBER.

Scherer, F. (1982). Inter-Industry Technology Flows and Productivity Growth. Review of Economics and Statistics.

Schmookler, J. (n.d.). Invention and Economic Growth. 1966.

Tetlock, P. C., Saar-Tsechansky, M., & MacSkassy, S. (2008). More than Words: Quantifying Language to Measure Firms' Fundamentals. Journal of Finance, 1437-1467.

Trajenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. Economics of Innovation and new technology, 19-50.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. The Rand Journal of Economics, 172-187.

Abstract

Innovation is a primary driver of growth in both macroeconomics and microeconomics. For a firm, innovation is vital to the future growth prospects of the firm. The prior literature has assumed more innovation is better, meaning the more innovation a business creates; the better off it is in the long-run. However, not all innovations are created equal. We contribute to the literature by investigating how institutional investors change future innovation, not in quantity, but diversity. Using several unique measures of technological diversification created from firm-level patent data, we show that institutional investors increase the focus on a firm's future innovation. Our results are robust to the classification scheme. Ultimately, our results indicate institutional investors create value by encouraging firms to build on prior knowledge.

2.1 Introduction

Innovation is a primary driver of growth in both macroeconomics and microeconomics. For a firm, innovation is vital to the future growth prospects of the firm. The prior literature has assumed more innovation is better, meaning the more innovation a business creates; the better off it is in the long-run. As a result, a robust line of literature is dedicated to investigating the motivation of managers to increase innovation output (sources). However, not all innovations are created equal. Therefore, we contribute to the literature by investigating how the nature of innovation affects firm value. Using institutional ownership, our goal is to show that external shocks, such as intuitional ownership, not only increases the quantity of a firm innovation but also how it changes the technology portfolio of the firm's overall portfolio.

Several features have been found to increase the future innovation output of a firm over time. For example, Institutional ownership is one feature shown to increase the future innovation production of business. Aghion et al. (2013) ask if institutional investors increase short-termism or if they allow managers to "swing for the fences?" They ultimately conclude managers are incentivized to produce more innovation. We ask the question, as a firm increases its innovation following investment by an institutional owner, do they also change the types of innovation they create. Just as firms can be segmented into sectors or industries, innovation can be related to specific industries.

This paper investigates what types of innovation an institutional investor is incentivizing. For example, do the incentives offered by principles incentivize managers to innovate in a more focused manner, ultimately increasing firm value? To answer this question we use a rich and unique dataset of firm patent data collected from the USPTO. With this data, we provide the most

detailed investigation into the outcome of incentives offered to managers to innovate. We look not at the quantity or quality of the patents produced by the firm, but rather their ability to create focused, value-enhancing patents over time.

First, we show there is a robust relationship between institutional ownership and future innovation in that institutional ownership increases the focus of a firm's innovation over time. We create several unique measures of unrelated innovation defined in general as a patent that cannot be directly linked to a firm's prior innovation (either by class or citation). After controlling for industry and year fixed effects, we find a significant negative relationship between institutional ownership and unrelated patents. Our results hold both with levels of institutional ownership as well as changes in institutional ownership. Next, we further show the robustness of our results by employing Instrumental Variables regressions as well as alternative specifications of our primary findings. The results are not sensitive to the specification of the regression or the classification scheme used. Our evidence supports the conclusion that institutional owners motivate managers to produce higher quality and more impactful innovations over time.

A firm's technological diversification, as defined by the diversification of firm innovation has implications for firm value (Silverman (1999); Garcia-Vega (2006); Leten et al. (2007)). The literature finds diversification can be both bad (Leten et al. (2007)) and good (Silverman (1999); Leten et al. (2007)) for firms. We argue that the institutional investors, through increased monitoring and proper incentivizing, focus firm innovation efforts and subsequently increase firm value. Furthermore, by focusing on firm innovations, institutional investors increase firm value. Our research indicates this is a unique method by which institutional investors increase the value of the firm.

Our results provide both a unique perspective as well as support for several theories regarding institutional ownership. For example, our results are consistent with Aghion et al. (2013) in that institutional investors have a positive, measurable effect on firm innovation. Furthermore, our results seem to support the career concern hypothesis model (Holmstrom (1982); Aghion et al.(2013)). Managers dislike the inherent risk involved in innovation. In the case of misguided or failing innovation, managers risk losing their job. Therefore, when monitored closely (by institutional investors) they choose to not only innovate more, but they choose to innovate more efficiently.

The paper is organized as follows; in Section 2, we survey the literature surrounding how ownership and management affect innovation. In Section 3, we provide a detailed description of the data and methodology of the study. We describe the variables used in this study in Section 4. Section 5 shows robust results indicating institutional ownership increases not only the quantity of innovation, but also the focus of future innovation. In Section 6, we conclude our study by discussing the implications of our results as well as the avenues for future research.

2.2 Literature Review

Innovation is a primary driver of growth in a firm. This fact has led to the popularity of researching how companies innovate and what factors that contributes to the firm's ability to innovate. The focus of this study is the impact of agency problems on future innovation.

Researchers have proposed several theories regarding the relationship between ownership and future innovation outcomes. Most have focused on the classic agency issue, or separation of interests between management and future innovation, introduced by Jensen & Meckling (1976).

Theoretical studies have modeled the potential agency relationship between owners and innovation.

Owner-manager relationships can influence innovation and innovation growth in several ways. Some of the earliest principal-agent models begin with the work of Harris and Raviv (1978) and Holstrom (1979). These early models show the importance of proper incentives to motivate agents to act in the best interest of the principal. More recently, Manso (2011) models optimal incentives that motivate innovation, again highlighting the importance of proper incentives, monitoring, feedback, and a long-term perspective. Institutional investors are uniquely positioned to incentivize agents to promote innovation.

Institutional investors have both the "carrot" and a "stick" in a sense they may both reward well-preforming managers and punish poor performing managers. Career pressures, such as the demand for strong quarterly reports or the risk of termination cause managers to focus on the short term, institutional holders may have superior monitoring abilities to realign the manager's focus with the longer horizon. Because institutional owners have superior supervision and oversight over managerial actions, they are positioned to encourage managers to innovate in a value-maximizing manor. This view is similar to Aghion et al. (2013) who propose and test a career concerns model. The authors show support for managerial innovation due to career concerns associated with institutional ownership. Furthermore, after modeling the relationship between ownership and management, the authors can distinguish between the actions of a "career concerned" manager and a "lazy" manager.

The empirical literature supports a robust theoretical and empirical relationship between managers, ownership, and innovation such as CEO overconfidence, (Hirshleifer et al. (2012); Galasso and Simcoe (2011)), executive hubris (Tang (2015)), or CEO-connections (Faleye (2014)). The

literature shows that adequately motivating a manger to innovation can lead to better innovation outcomes. For example, not only stimulating a manager to innovate, but also motivating a manager for a long-term growth perspective support positive innovation outcome: Manso (2011) and Ederer & Manso (2013). External factors, such as financial development, Hsu et al. (2014), or analysis coverage, He & Tian (2013), have been found to affect innovation output of firms also.

Encouraging innovation is vital to the growth prospects of a firm. Fagerberg (2006) draws a distinction between invention as innovation "first occurrence of an idea for a new product or process, while innovation is the first attempt to carry it out into practice." For firms and society, innovation represents growth and advances. Whether growth is in technology, services, or a new process; firms benefit. Schumpeter (1942) is attributed to acknowledging the importance of innovation to businesses. However, Schumpeter is not alone in his conclusion based on the importance for firms to innovate, Baumol (2002). The primary reason firms' as a whole peruse innovation is a method by which firms can increase their opportunities and grow their firms. For some, innovation can increase the likelihood of survival, Cefis & Marsili (2006), and for others, innovation is the most efficient way to drive growth, Coad & Rao (2008).

Aghion et al. (2013) show that institutional investors increase the innovation output of firms. However, no prior study has investigated the knowledge-relatedness of subsequent innovation following the investment by institutional investors. Using a similar methodology to Aghion et al. (2013) we study how firm technological diversification changes after investment by institutional investors. Encouraging a firm to focus on value-enhancing innovation should lead to increased firm value in the long-run. The findings of Makri et al. (2006) highlight the importance of our study. The authors find that as technological intensity increases, effective incentives are focused more on increasing innovation resonance rather than only innovation quantity. As a manager,

increasing innovation may lead to an increased firm value; however, increasing both the amount and the focus on innovation may increase the firm value even higher. We believe our study is similar to that of Berger & Ofek (1995), who show that firms that overinvest and cross-subsidize innovation incur a value loss.

2.3 Data

We use firm-level data on institutional ownership and innovation from a variety of sources (See Appendix for more Information). For patent data, we use a novel new dataset that combines the known patent ownership information, as well as new firm-patent pairings. The data set is created by first combining three impactful patent datasets. The first two datasets combined are the NBER Patent data files match more than three million patents and associated assignees to U.S. firms for the period 1976 – 2006, the Kogan et al (2016) 1926-2010. By combining the two datasets we take advantage of the prior firm matching. To further identify patent owners, we use the ownership identification of Lai (2013). This new dataset uses Bayesian methods to match similar names/owners across patents over time.

After matching all three prior datasets, we extend the data by matching patent names to firm subsidiary information available on EDGAR. Each publicly traded firm must list their subsidiaries in their 10-K annual filings; Exhibit 21 on EDGAR. Using the information on firm subsidiaries, we are able to extend the link between firms and patents beyond the year 2010. The firm subsidiary information provides matches in years prior to 2010 that were not previously available.

The resulting data file allows us to match firm-level patent information to firm fundamental data from Compustat and Institutional ownership data from S&P Capital IQ. Compustat contains firm

fundamental data for all U.S. publicly listed firms. Compustat information relevant to this study includes; Total Assets, R&D, Sales, Capital Expenditure, Book Leverage, and the Number of Employees.

The new dataset allows us to create new measures of technological change that will be discussed in Section 2.4. Our primary analysis is based on a text-based measure of patent similarity for a firm.

For institutional ownership, we use S&P Capital IQ, which contains ownership data including the number of institutional investors, the number of shares issued, as well as the number of shares held by institutions. After combining all files, our sample includes 1,015 firms and over 11,000 firm years. Descriptive statistics for the baseline sample are documented in Table 2-1. Firms used in this study have an average institutional ownership of 36.8%. The average number of patents produced per firm-year is forty-eight; however, this is highly skewed, as the median is six. Table 2-1 also includes summary statistics for several control variables employed in this study.

Table 2-1: Summary Statistics

	Mean	Sd	Min	Max
count_f	0.3015	1.1017	0.0000	31.0000
inst_own	0.3680	0.2378	0.0000	1.0000
rnd	0.0376	0.0763	0.0000	2.8443
lat	6.1257	2.1983	-0.6218	13.5921
lsale	6.1580	2.1191	-1.0217	12.4103
capx	223.3620	902.4935	0.0000	29657.0000
leverage	0.3175	0.3104	-14.8515	11.5709
lemp	1.2166	1.9767	-6.9078	7.2442

Table 2-1 displays descriptive statistics for the sample used in this study.

The control variables selected for this study are Total Assets, R&D, Sales, Capital Expenditure, Book Leverage, and the Number of Employees. In our results, we measure Total Assets as the log transformation of total assets (lat), R&D as the percentage of R&D expenditures to total sales (R&D), Sales as the log transformation of sales (lsale), and the number of employees as the log transformation (lemp). We compute leverage as the firm's book leverage, short-term debt to total debt (leverage).

The control variables selected are based on the findings of the prior literature. For example, Total Assets, a proxy for firm size, is associated with increased innovation. We expect Total Assets, R&D, Capital Expenditures, Book Leverage, the Number of Employees to have a positive impact on the change of future innovation and sales to have a negative impact on changing innovation.

2.4 Technological Diversity

2.4.1 Measuring Technological Diversity

Many studies investigate the impact of institutional ownership on the future patent outcomes by measuring the simple count or citation weighted future patent outcomes. However, our study

examines how a firm's innovation changes, not in quantity but focus, as an investment by institutional investors. To measure the firms changing patent portfolio over time, we create three unique measures of patent portfolio diversification. Two measures are based on patent citations and classifications. The third is a novel approach is employing the text of patents to determine their similarity. Prior studies pass over the information contained in the text, in part because of its size; however, employing text as a measure of similarity has advantages over traditional measures.

Each patent is issued a patent classification that measures the technical content of individual patent. There are several different classification schemes. In the United States, the two primary patent classifications are the International Patent Classification (IPC) and the United States Patent Classification (USPC). The International Patent Classification (IPC) is an internationally agreed upon standard used by over 100 countries. The United States Patent Classification (USPC) is determined and maintained by the United States Patent and Trademark Office (USPTO). Each system classifies technology in a similar, but unique manner. Each measure classifies patents by assigning them a unique letter/number combination and many patents are assigned both a primary and secondary classification. An example of each classification method is included in Appendix C.

Prior researchers have measures technology diversification using patent classification methods. One approach taken by several researchers is the approach of Jaffe (1986, 1989) who measures technological diversity by observing the distribution of patents over technology classes. Others have defined technological diversification in a different manner. Engelsman & Van Raan (1991, 1992) and Verspagen (1997) take advantage of the difference between each primary and secondary technology class. Breschi et al. (2003) conduct a similar study using several unique variables that measure technological diversity.

We create two variables that capture the firms changing innovation focus over time. The first variable, Category_(i,t±3) measures the unique number categories a firm creates a patent in given it previous three years of patent history. For example, if a firm in the past three years has patented in the categories of food, and the firm then creates a patent in technology, this would increase the variable's count by one. We generate this variable by creating a moving tally of patent classes a firm establishes a patent in for the past three years. Similarly, we create a score of all the firm's courses in the leading three years. After taking the inverse intersection of the two lists, we count the number of unique classes and define this as Category_(i,t±3). A more significant number indicates a firm that creates more "new" or unique innovation over time. We use both the USPTO's United States Patent Classification (USPC) as well as the International Patent Classification (IPC) to show our results are not dependent upon the classification method.

$Category_{i,t\pm3}$ is defined as:

$$Category_{i,t\pm3} = |B \backslash A|$$

Where A = Set of all patent classifications in year [t-3,t) and,

B = Set of all patent classifications in years [t, t+3)

The next variable we use to capture a firm's future patent outcome is New〖Technology〗_(t+1). This variable is defined as the number of patents applied for those that cite no previous firm patent. Most firms build on technology created previously. Even if a patent represents "new technology" it still is likely to cite or be the product of prior work. A patent that does not cite any prior work by a firm represents a systematic change in innovation for a firm. Therefore, a more significant number indicates that a firm creates patents in more technological classes any given year.

Both of these measures indicate how “diverse” or “focused” a firm’s innovation output is. Whereas most studies consider the effect of institutional ownership on innovation output, there has been no in-depth study on the firm's innovation focus. A more detailed description of the variables used in this study is included in Appendix A.

2.4.2 Hypothesis Development

With several measures for a firm’s patent portfolio, we can investigate the impact institutional investors on firm diversification, and subsequently diversification on firm value. The first hypothesis we propose is,

H1: Greater institutional investor ownership increases the focus of a firm’s innovations.

We investigate this hypothesis by regressing one of are alternative patent portfolio diversification measures on the percent institutional ownership as well as control variables. For this study, we follow the prior literature and include Sales, R&D, Total Assets, Capital Expenditures, and the Total Number of Employees as control variables. We argue the mechanism by which institutional investors’ decrease the diversification of a firm’s patent activity is through manager monitoring, similar to the model of Aghion et al. (2013). We show our results are robust to endogeneity concerns by employing instrumental variables, as well as regressions on subdivided populations.

Next, we investigate the impact of reduced firm patent diversification on firm value with the following hypotheses,

H2: Lower patent diversification increases firm value.

We study this hypothesis by regressing firm value (as measured by Tobin's Q) on our unique patent diversification measure as well as control variables. Under the argument of Makri et al. (2006), a firm should consider the quality of innovation as well as their quantity to increase firm value.

2.5 Results

2.5.1 Intuitional Ownership and Firm Innovation

We expect the investment by institutional investors to increase a firm innovation focus. Because institutional investors have superior monitoring abilities and can better align manager incentives in the interest of shareholders, we believe as institutional holdings increase firms become more focused with their future innovations.

The first investigation of institutional ownership on a firm's patent focus is presented in Table 2-2. Table two regresses the variable $Catagory_{i,t\pm3}$ on institutional ownership and other control variables. To show our results are not subject to model specification, we attempt several model specifications. From left to right, our regressions include Poisson, negative binomial, OLS, and logit. Traditionally, citations and patent counts are assumed to be Poisson or negative binomial distributed. A negative and significant coefficient for $Catagory_{i,t\pm3}$ would indicate firms focus their innovations as the percentage of institutional ownership increases.

Table 2-2: Institutional Ownership and Innovation

	(1)	(2)	(3)	(4)	(5)
	Poisson	Negative Binomial	OLS	OLS	OLS

	$Category_{t+3}$	$Category_{t+3}$	$Log(Category)$	$Category_{t+3}$	$Text\ Origanality$
%inst_own	-1.0922***	-0.9429***	-0.1859***	-0.4866***	-1.1508***
	(0.1992)	(0.1755)	(0.0336)	(0.1032)	(0.1966)
rnd	1.2528**	1.6544***	0.1807*	0.5134*	1.1890**
	(0.3980)	(0.4939)	(0.0711)	(0.2527)	(0.3948)
lat	0.3368***	0.2486**	0.03803***	0.1243***	0.2297**
	(0.0758)	(0.0808)	(0.0099)	(0.0298)	(0.0698)
lsale	-0.2568**	-0.1264	-0.01258	-0.06819*	-0.06633
	(0.0880)	(0.0957)	(0.0131)	(0.0338)	(0.0928)
capx	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
leverage	0.1562*	0.2578	0.02465	0.05799	0.1544
	(0.0635)	(0.1464)	(0.0135)	(0.0389)	(0.0810)
lemp	0.1764*	0.09626	0.01188	0.04101	0.08621
	(0.0814)	(0.0799)	(0.0113)	(0.0262)	(0.0852)
lnalpha		1.7198***			
		(0.0656)			
Year/Industry Controls	YES	YES	YES	YES	YES
N	11542	11542	11542	11542	11530

The results of Table 2-2 are our baseline results; they document a negative and significant effect. Standard errors are reported in parentheses and clustered by firm. The results show that as institutional ownership increases, the firm's innovation becomes more focused. Our findings are not sensitive to the specification. Additionally, parameter estimates remain relatively stable across all specifications. Our baseline (Poisson) specification shows that an increase in institutional ownership by five percent decreases the rate of new patent classifications in the preceding years

by approximatively 5%. In unreported results, changing the windows of $Catagory_{i,t\pm3}$ by plus-one or minus-one years does not change the parameter estimates in a meaningful way.

It is possible our results are sensitive to the classification method used to create the variable, $Catagory_{i,t\pm3}$. To alleviate this concern, we investigate our baseline results using an additional classification method. Our dependent variable $Catagory_{i,t\pm3}$ is calculated in the same method as the previous results. However, to show our results hold we use $Catagory_{i,t\pm3}$ calculated using the International Patent Classification (IPC) categories to determine the dependent variable. Compared to the USPC method the IPC method has approximately had a different method of subdivision. If our results are not sensitive to the patent classification method, we should find our results are similar to those from Table 2-2.

Table 2-3: Institutional Ownership and Innovation (Text Based Measure)

	Poisson $Category_{t+3}$	Negative Binomial $Category_{t+3}$	OLS $Log(Category)$	OLS $Category_{t+3}$	OLS $Text\ Origanality$
inst_own	-1.0358***	-0.9097***	-0.1878***	-0.4657***	-1.1508***
	(0.2067)	(0.1945)	(0.0375)	(0.1001)	(0.1966)
rnd	1.4047**	1.4866**	0.2640*	0.9316*	1.1890**
	(0.4314)	(0.5269)	(0.1126)	(0.4727)	(0.3948)
lat	0.2686**	0.1951*	0.02964**	0.08895***	0.2297**
	(0.0863)	(0.0816)	(0.0101)	(0.0268)	(0.0698)
lsale	-0.1038	-0.05017	0.004339	-0.005169	-0.06633
	(0.0934)	(0.0936)	(0.0122)	(0.0291)	(0.0928)
capx	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
leverage	-0.01583	0.1100	-0.01353	-0.1265	0.1544
	(0.0743)	(0.1022)	(0.0271)	(0.1218)	(0.0810)
lemp	0.1139	0.1017	0.006529	0.01885	0.08621
	(0.0856)	(0.0813)	(0.0111)	(0.0251)	(0.0852)
lnalpha		1.7622***			
		(0.0720)			
Year/Industry Controls	YES	YES	YES	YES	YES
N	10537	10537	10537	10537	11530

The results of Table 2-3 are very similar to those of Table 2-2. We find the same negative significant effect. Again, the results show that as institutional ownership increases, the firm's innovation becomes more focused. The parameter estimates are not only stable across specifications but also similar in magnitude to those in Table 2-2.

The results of Table 2-2 & 2-3 indicate that institutional ownership has a significant negative effect on a firm's future technological diversification. As the level of institutional ownership increases in a firm, the firms tend to focus their innovation in the preceding years. Next, we investigate if institutional ownership affects the firm's creation of new technology, defined as technology not referenced by the firm prior.

Our first measure of changing innovation could be dependent on the frequency of a firm's patents. Furthermore, it may be dependent on the window we use to determine changing innovation. To alleviate these concerns, we employ a separate variable New〖Technology〗_(t+1) as the number of patents a firm creates absent of a citation of any of the firm's previous patents. A larger number would indicate a firm that is creating patents that are unrelated to its previous patents.

Using the dependent variable New〖Technology〗_(t+1) again look the impact of institutional ownership on firm innovation focus. The results are documented in Table 2-4.

Table 2-4: Institutional Ownership and New Patents

	Probit _new	Negative Binomial _new	OLS ratio	Logit dummy
inst_own	-0.6008*	-0.1065	-0.1439***	-1.1422
	(0.2391)	(0.1596)	(0.0305)	(0.6768)
rnd	3.4061***	4.6040***	-0.4017***	-3.6045*
	(0.5425)	(0.7010)	(0.1050)	(1.5231)
lat	0.2076	0.6701***	-0.02821	-0.1023
	(0.1305)	(0.0790)	(0.0144)	(0.2203)
lsale	0.7874***	0.1322	0.002729	0.2556
	(0.1823)	(0.0965)	(0.0188)	(0.3630)
capx	-0.0001***	-0.0001	-0.0001	0.0026**
	(0.0000)	(0.0001)	(0.0000)	(0.0009)
leverage	0.2694*	-0.3076**	-0.0003949	-0.08327
	(0.1369)	(0.1184)	(0.0229)	(0.3477)
lemp	-0.08476	0.02211	0.02054	0.3879
	(0.1295)	(0.0879)	(0.0167)	(0.3512)
Year/Industry Controls	YES	YES	YES	YES
N	5210	5210	5210	3459

The results of Table 2-4 are consistent with our prior results, and they highlight a new level. The results indicate higher institutional ownership is associated with more focused innovation. All standard errors are clustered at the firm level.

The results show the benefit institutional investors bring. Our results support the theory of prior literature. For example, Aghion et al. (2013) show institutional investors provide proper incentives to managers. Our results indicate that after the investment of institutional investors, firms innovation more in-line with their prior innovation. In the next section, we address the robustness of our results with regards to causality and selection bias.

2.5.2 Institutional Ownership and Causality

One issue with our prior analysis is the issue of causality. We find a significant relationship between institutional ownership and firm innovation. However, just as plagued other similar studies, we have said very little about the causal relationship between institutional ownership and firm innovation. In the following section, we include new test created to investigate the causal relationship between institutional ownership and innovation. The first set of tests includes regressing the measure of firm innovation on the change in institutional ownership. The second test uses a method similar to Aghion (2013). We use instrumental variables where the first stage regresses institutional ownership on membership in the S&P 500.

In our first robustness test, we regress the change in technological areas on the change in institutional ownership. We compute the change in institutional ownership as $(InstOwn_t - InstOwn_{t-1}) * I_{(InstOwn_t - InstOwn_{t-1})>0}$ where the indicator variable is equal to one when the change in institutional ownership is positive. We expect that if institutional owners are encouraging firms to become more focused on their innovation, the coefficient to be significant and negative. The results of this analysis are reported in Table 2-5.

Table 2-5: Change in Institutional Ownership

	Negative Binomial *count_f*	OLS *l_count_f*	Poisson _new	OLS ratio
Δ inst_own	-0.0002***	-0.5757***	-0.0002**	-0.0704***
	(0.0000)	(0.0916)	(0.0000)	(0.0503)
Controls	YES	YES	YES	YES
Year/Industry Effects	YES	YES	YES	YES
N	11,542	11,542	5,210	5,210

The results show that a significant negative relationship between technological diversification and institutional ownership. As firms experience greater institutional ownership, they decrease their future technological diversification.

The prior analysis may be subject to endogeneity problems. In our analysis, institutional investors may choose more focused firms or firms with greater firm value. If this is the case, then naturally a higher level of institutional investors will look like it increases firm value, when in reality, institutional investors are following value. To alleviate this bias, we employ instrumental variables regression. For our instrument, we use membership in the S&P 500. The use of this instrument is similar to the analysis of Aghion et al. (2013). The inclusion of a firm in the S&P 500 may increase the likelihood a firm is owned by an institutional investor for several reasons. 1) The Employee Retirement Income Security Act and other fiduciary duty measures have been shown to influence portfolio section through an implied endorsement of broad indexing. 2) Additionally, many funds are benchmarked or indexed to the S&P 500 which either requires or encourages them to invest in the fund. 3) Furthermore, inclusion in the S&P 500 is not influenced by the company and inclusion is based on sector restrictions, not firm preference. Because firms have no choice as to their inclusion or exclusion from the index, the variable is semi-random.

In the first stage, we regress institutional ownership on the membership in the S&P 500. Next, we used our technological diversification measures on predicted ownership. The results of this analysis are presented in Table 2-6.

Table 2-6: IV Regression using S&P 500

	First Stage Institutional Ownership	Second Stage Unique Class	First Stage Institutional Ownership	Second Stage Unique Patent
inst_own		-0.5757***		-0.0708**
		(0.0916)		(0.0503)
S&P 500	0.1373***		0.2022***	
	(0.0053)		(0.0073)	
Controls	YES	YES	YES	YES
Year/Industry Effects	YES	YES	YES	YES
N	11,542	11,542	5,210	5,210

Column (1) and Column (3) represent the regression of institutional ownership on membership in the S&P 500. As expected each shows a positive relationship, indicating firms in the S&P 500 are more likely to have higher institutional ownership. In both regressions the magnitude of this effect is similar. In the second stage, we regress new technological diversification on the predicted institutional ownership from the first stage. This reduces the bias introduced by endogeneity problems. The relationship between predicted institutional ownership and patent uniqueness is negative across both our control variable and our new measure of diversification. The results support the relationship between institutional ownership and new technological diversification.

2.5.3 Additional Results

Prior results confirm that institutional investors increase the focus of a firm's innovation. We show our results are robust to endogeneity issues as well as the specification of firm innovation. In the next section, we will demonstrate that a firm's technological diversification affects its firm value. We draw parallels to the infamous "Diversification Discount."

Early research by Lang and Stulz (1994), Berger and Ofek (1995), and Servaes (1996) show that diversified firms trade at a discount relative to their less diversified peers. Later studies, such as Villalonga (2004), suggest that the diversification discount is a symptom of measurement specification. Our research is similar to previous studies; however, as an alternative to the diversification of firm production as measured by the previously cited studies, we look at the diversification of firm innovation activities.

If a firm produces overly diverse new technology, they may be inefficiently allocating capital amounts their R&D divisions. In this case, a firm that creates patents that are of little value to the firm is inefficiently allocating capital. Whereas, a firm producing meaningful innovation, highly related to the firm's prior products is assumed to be more efficient.

To investigate this relationship, we employ both the measures of firm diversification from the previous section as well as firm value, as measured by Tobin's Q. We compute Tobin's Q as the total market value of the firm divided by the firm's total assets. When a firm becomes more focused in their innovation, they increase their firm value over time, thereby increasing the firms Tobin's Q.

In Table 2-7 we investigate the impact of a firm changing patent portfolio on firm value. To ensure that firms that have large diversified segments do not drive the results, we subdivide the sample

into three groups. High, medium, and low patent activity. The High patent activity group is defined as those firms that produce patents in the upper 30th percentile of all firms. Similarly, the low patent activity group is defined as those in the lower 30th percentile of all firms. All remaining firms are placed in the medium patent activity group. By subdividing our sample, we can determine if the effect of focusing a firm's innovation is consistent across firms with different levels of patent activity. Based on our hypothesis, we would expect a negative relationship between our text-based measure of patent originality, and firm value. The results of this analysis are presented in Table 2-7.

Table 2-7: Impact of Innovation on Different Firm Sizes

	Text Diversity	*Tobin's Q*
Panel A: Sorted by Patent Activity		
High Activity (> 70th Percentile, n = 2,893)		
inst_own	-0.2729*	
	(0.1070)	
Text Diversity		-0.0024*
		(0.0017)
Mid Activity (30th - 70th Percentile, n = 3,858)		
inst_own	-1.2398***	
	(0.3070)	
Text Diversity		-0.0024**
		(0.0013)
Low Activity (< 30th Percentile, n = 2,640)		
inst_own	-0.7831***	
	(0.4123)	
Text Diversity		-1.2398***
		(.1212)

Table 2-7 divides the sample firms into three groups of high, medium, and low patent activity. Across all groups, institutional ownership seems to focus firms patenting activities. This effect is significant at the ten percent level across all groups. For the medium and low patent activity groups, the impact is significant at the one percent level.

The effect seems to be the greatest for the medium and low patent activity groups, where the high activity group sees the smallest impact. This may be due to monitoring costs associated with large firms. As described in the prior section, the primary method by which institutional investors affect innovation is through monitoring. In a large or highly diversified corporation, supervision by an institutional investor may become more costly and expensive. Ultimately, our results show that institutional ownership reduces a firm's diversity making them more focused over time.

In the second column of Table 2-7 we investigate the impact of a firm's technological diversification, using our unique text-based measure of patent originality, on firm value, as measured by Tobin's Q. Again, we look at the impact of changing technological diversity using three groups of patent activity. The results indicate a negative relationship between patent originality and firm value. As patent originality increase, the firm value decreases. The stated effect remains consistent across all patent activity groups. More notably, the results are highly significant for the low patent activity group. This may indicate that our results are not being driven by large firms with large patent portfolios. Firms that produce few patents also receive a benefit from focusing their patent portfolio.

These results are consistent with the results of Makri et al. (2006). Increasing firm value using patenting activity should consist of focusing on more than the number of patents produced. In fact,

our results show that it is the types of patents that are created that also lead to changing firm value. Quality and quantity lead to higher firm value over time. It seems as though institutional investors increase both patent quantity and quality.

The final analysis considers both Hypothesis 1 & 2. Does institutional ownership lead to a firm focusing their patent portfolio, which ultimately increases firm value? The results have shown each independent component of this, however, to control for endogeneity, we use three-stage least squares (3SLS). In the first equation, we estimate the predicted institutional ownership given inclusion in the S&P 500 index. This is similar to the IV regression in the previous section. The second equation uses the predicted institutional ownership to find the predicted patent portfolio diversity. Lastly, using the predicted patent diversification, we investigate the impact on firm value.

We use both 3SLS and seemingly unrelated regression or SUR. First, to ensure that the S&P is not correlated with the firm value in our sample, we test the means of both a group of firms included in the S&P and a group excluded from the S&P 500. We find the result insignificant indicating that our sample is balanced and Tobin's Q and S&P members are not strongly correlated. The results are reported in Table 2-8.

The results in Table 2-8 are in line with expectations. As a firm becomes included in the S&P 500, they are more likely to be held by institutional investors. Furthermore, the increase in institutional investors focuses firm innovation and that focus in innovation is associated with an increase in firm value. The results hold for both estimation methods.

Table 2-8: Three Stage Least Squares

Panel A: Sample Balance

	Not Included in S&P	Included in S&P	Two Tail T-test (P-value)
Mean Tobin's Q	1.8546	1.8953	(0.1344)
N	19,182	10,737	

Panel B: Three Stage Least Squares

	Estimation Method: 3SLS			Estimation Method: SURE		
	First Stage	Second Stage	Third Stage	First Stage	Second Stage	Third Stage
Dependent Variable:	Institutional Ownership	Text Originality	Tobin's Q	Institutional Ownership	Text Originality	Tobin's Q
Patent_Portfolio			0.7605***			0.1593**
			(0.0627)			(0.0700)
inst_own		-2.4446***			-1.1931***	
		(0.1285)			(0.0441)	
S&P 500	0.1979***			0.1963***		
	(0.0031)			(0.0031)		

2.6 Conclusion

We investigate the relationship between technological diversification and institutional ownership. We measure technological diversification using unique methods determined by patent classifications and by using the text of patents. Using the text of patents allows a researcher to capture variation within patent classes, and provides greater flexibility over tradition measures. Our results show strong statistical support indicating that increases in institutional ownership lead to more focused, less diversified innovation.

As majority owners, institutional investors have the ability to increase monitoring and pressure managers to increase the value of a firm. Because innovation is a value-enhancing activity, it can be used as a value-enhancing signal. However, prior studies only consider the quantity of innovation, not the type of innovation. Our research shows the impact of institutional investors on the technological diversification of a firm's innovation.

If managers can incentivize firms to increase the value, or in the terms of Makri et al. (2006), the resonance, or quality, of their innovation, future innovation will be more impactful. The results of our study show that institutional investors encourage firms to increase the resonance of their innovation. Our results, as well as the results of prior studies, support the view of institutional investors being value enhancers.

Our results have implications for ownership structure as we show that institutional investors have an impact on the future innovation of a firm. Additionally, our results reveal a mechanism by which institutional investors add value to a firm by focusing their innovation. The conclusions of our study support value the enhancing the role of institutional investors documented in Cornett (2007),

Maury & Pajuste (2005), and Elyasiani & Jia (2010), among others. Our adds to the growing collection of the impact institutional investors have on firm value.

References

Aghion, P., Van Reenen, J., & Zingales, L. (2013). Innovation and institutional ownership. *American Economic Review, 103*, 277-304.

Baumol, W. J. (2002). *The free-market innovation machine: Analyzing the growth miracle of capitalism.* Princeton university press.

Berger, P. G., & Ofek, E. (1995). Diversifications effect on firm value. *Journal of Financial Economics, 37*, 39-65.

Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research policy, 32*, 69-87.

Cefis, E., & Marsili, O. (2006). Survivor: The role of innovation in firms' survival. *Research Policy, 35*, 626-641.

Coad, A., & Rao, R. (2008). Innovation and firm growth in high-tech sectors: A quantile regression approach. *Research policy, 37*, 633-648.

Ederer, F., & Manso, G. (2013). Is pay for performance detrimental to innovation? *Management Science, 59*, 1496-1513.

Engelsman, E. C., & Van Raan, A. F. (1991). Mapping of technology. *A First Exploration of Knowledge Diffusion Amongst Fields of Technology. The Hague: Ministry of Economic Affairs*.

Fagerberg, J., & others. (2006). Innovation, technology and the global knowledge economy: Challenges for future growth. *Green Roads to Growth" Project and Conference, Copenhagen.*

Faleye, O., Kovacs, T., & Venkateswaran, A. (2014). Do better-connected CEOs innovate more? *Journal of Financial and Quantitative Analysis, 49*, 1201-1225.

Galasso, A., & Simcoe, T. S. (2011). CEO overconfidence and innovation. *Management Science, 57*, 1469-1484.

Garcia-Vega, M. (2006). Does technological diversification promote innovation?: An empirical analysis for European firms. *Research Policy, 35*, 230-246.

Harris, M., & Raviv, A. (1978). Some results on incentive contracts with applications to education and employment, health insurance, and law enforcement. *The American Economic Review, 68*, 20-30.

He, J. J., & Tian, X. (2013). The dark side of analyst coverage: The case of innovation. *Journal of Financial Economics, 109*, 856-878.

Hirshleifer, D., Low, A., & Teoh, S. H. (2012). Are overconfident CEOs better innovators? *The Journal of Finance, 67*, 1457-1498.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 324-340.

Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: evidence from firms´patents, profits and market value. National bureau of economic research Cambridge, Mass., USA.

Jaffe, A. B. (1989). Real effects of academic research. *The American economic review*, 957-970.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*, 305-360.

Leten, B., Belderbos, R., & Van Looy, B. (2007). Technological diversification, coherence, and performance of firms. *Journal of Product Innovation Management, 24*, 567-579.

Manso, G. (2011). Motivating innovation. *The Journal of Finance, 66*, 1823-1860.

Schumpeter, J. A. (1942). *Socialism, capitalism and democracy.* Harper and Brothers.

Silverman, B. S. (1999). Technological resources and the direction of corporate diversification: Toward an integration of the resource-based view and transaction cost economics. *Management Science, 45*, 1109-1124.

Tang, Y., Qian, C., Chen, G., & Shen, R. (2015). How CEO hubris affects corporate social (ir) responsibility. *Strategic Management Journal, 36*, 1338-1357.

Van Raan, A. F., Noyons, E. C., & Engelsman, E. C. (1992). Unravelling the texture of science and technology by bibliometric cartography. *Journal of AGSI, 78*, 77-88.

Verspagen, B. (1997). Estimating international technology spillovers using technology flow matrices. *Review of World Economics, 133*, 226-248.

Appendix

Appendix A: Patent Data

Patent data can be grouped into two primary generations. The first lead by the work of Schmookler (1966), Scherer (1982), and Griliches (1984), hand collected and hand matched patents to their owners. Furthermore, they relied on a simple patent count to measure a firms innovation activity. The second and most recent generation is led by the work of Trajtenberg, Jaffe, and Henderson (1997) and Kogan et al. (2016). These researchers used publicly available data to match patents to their corresponding firms using technology. When combined, the previous data spans from nearly seventy years, 1926-2010, and has made a meaningful impact on innovation research.

In this paper, we propose a third generation of patent data. The third generation of patent data has two distinct contributions. First, it extends patent-firm ownership information beyond 2010 to 2016. The new dataset used the already discovered connections between prior datasets and additionally data on firm names. The second contribution of this dataset is the available to use the text information of patents. Previous studies have relied on patent counts, citations, or patent classes to define the nature and value of a patent. However, few studies have unlocked the data available in a patents text. The new dataset allows for the use of the text data. We also provide an example variable and test study.

The appendix provides more detail on the collection and transformation of the new dataset. This information is technical and may provide the reader with a clear picture of how the data is manipulated.

1. Data Collection

1.1 Patent Data

We gather the full text of all U.S. utility patents granted from 1926 to 2010 from the U.S. Patent and Trademark Office (USPTO). All patent information collected is publicly available information and contained in the USPTO's bulk data files[4]. The bulk data files are divided into several different file formats, with data before 1975 in PDF format. While data is available from 1975 or earlier, the text is scanned using an Optical character recognition process and is extremely unreliable and unstructured. Previous researchers have used data pre-1975, such as Packalen and Bhattacharya (2012). However, we find the reliability of the character recognition to be substandard. For patents granted after 1975, we write a python script to filter and sort the data into a malleable database. We collect the relevant sections of this study. They include: The Patent Number, Application Date, Award Date, Inventor, Assignee, References Cited, Abstract, Claims, and Description. The final sample comprises 4,131,597 patents.

1.2. Firm Ownership Information

Matching firms to their owners is no simple task. Several problems arise when trying to discern the ownership of any specific patent. For example, some firm names have several variations. International Business Machines may be listed in numerous ways, such as IBM, Inter Business Machine, among others. Over time, more data has been gathered that can aid in the process of ownership matching. The literature of Kogan et al. (2016) and Hall, Jaffe and Trajtenberg (2001) have made substantial upgrades to the link between firms and their patents. This dataset uses the knowledge of the previous datasets.

[4] https://www.uspto.gov/learning-and-resources/bulk-data-products

To extend the data beyond the previous datasets, we employ a novel approach, matching firms to their subsidiaries. Firms can assign patents to their parent firm, or any one of their listed subsidiaries. Each publicly traded firm must list their subsidiaries in their 10-K annual filings, Exhibit 21 on EDGAR. Using the information on firm subsidiaries, we are able to extend the link between firms and patents beyond the year 2010. Furthermore, the firm subsidiary information provides matches in years before 2010 that were not previously available.[5]

First, we gather the developed firm-patent data sets provided by Kogan et al. (2016), and the NBER patent database of Hall, Jaffe, and Trajtenberg (2001). The Hall, Jaffe and Trajtenberg (2001) dataset links patents to firms for the years 1976-2006. The Kogan et al. (2016) dataset links patent ownership to firms from 1926-2010. By combining the two datasets, we take advantage of the prior firm matching. To further identify patent owners, we use the ownership identification of Lai (2013). This dataset using Bayesian methods to match similar names/owners across patents over time. Each dataset contains matches that are unique to each dataset.

Now we combine the resulting dataset with the new matches from the Exhibit 21 files. Names from the Exhibit 21 are cleaned and standardized according to the NBER name standardization routine. In summary, the process involves normalizing the text, including the company suffixes, and removing all punctuation. The routine then tries to find all perfect matches between the patent datafiles and the EDGAR files. After an initial match, a secondary match is performed after further cleaning of the names is performed.[6]

[5] More information on Exhibit 21 can be found here: https://www.sec.gov/oiea/Article/edgarguide.html
[6] A Stata version of the name standardization routine can be found here: https://sites.google.com/site/patentdataproject/Home/posts/namestandardizationroutinesuploaded

2. Data Processing and Cosign Similarity

Each data source is gathered, the data need to be processed and converted into qualitative data. The processing of data happens in five general steps:

1. Collect - Gather the data from the USPTO.
2. Pre-process - Process each file to gather macro information (Name, Application Year, etc.) as well as patent Description, Abstract, and Claims.
3. Process - Process the text data including cleaning "stop" words and other text that does not contain information.
4. Ownership Identification - Matching patents to firm ownership.
5. Variable Creation - Using the processed text to create variables of interest.

The pre-processing stage involves extracting information from the USPTO bulk patent data files. Variables of interest are captured as well as text from the various sections of the patents During the Process stage; the text is cleaned of punctuation and only phrases made up of only uncapitalized English nouns. We also remove all words lacking information content, stop words. Stop words are devoid of any informational material (such as: and, it, or) and therefore contribute very little to the understanding of similarities between two documents. The cleaning process is similar to prior literature.

During the process state, the remaining text (referred to as tokenized) is converted into count vectors of n length (where n refers to the number of unique words). Each input in the vector

corresponds to the count of a unique word. These vectors are referred to as term vectors. Each vector is then normalized to one.

Given any two vectors (or documents) one can compute the similarity between them using cosign similarity. The cosign similarity or cosign distance refers to the distance between two vectors. Given two non-zero vectors, one can compute the cosign similarity using the dot product of each term vector.

$$Cosign\ Similarity = \frac{A \cdot B}{||A||||B||}$$

Where A and B refer to the non-zero term frequency vector.

Appendix B: Definition of Variables

Variables	Definition	Source
Panel A. Institutional Ownership		
Inst. Own%	Total institutional ownership as a fraction of shares outstanding	
Max. Inst. Own%	Percentage of shares outstanding held by the firm's largest institution	
Top 5 Own	Percentage of shares outstanding held by a firm's top 5 institutional investors	
Top 10 Own	Percentage of shares outstanding held by a firm's top 10 institutional investors	
Top Block Hold.	Percentage of shares outstanding held by a firm's institutional investors whose holdings are greater than 5%	
Panel B. Patent Variables		
$Catagory_{i,t\pm3}$	The number of unique categories a firm creates a patent in given it previous three years of patent history	NBER Patent Data
$NewTechnology_{t+1}$	The number of patents a firm creates absent of a citation of any of the firm's previous patents	NBER Patent Data
Panel C. Control Variables		
Book Leverage	Long-term debt divided by book value of assets	COMPUSTAT
Log(TotalAsset)	Log transformation of total assets	COMPUSTAT
R&D	Income before extraordinary items plus depreciation and amortization divided by book value of assets	COMPUSTAT
Log(Sales)	Log transformation of firm i's sales.	COMPUSTAT
Tobin's Q	Market value of assets divided by book value of assets	COMPUSTAT
Log(FirmAge)	Log transformation of firm age	COMPUSTAT
HHI Index	Herfindahl index based on the firm's sales in a given 4-digit SIC industry.	COMPUSTAT
CAPX	Capital expenditure of firm i in year t.	COMPUSTAT
Log(#Emp)	Log(1 + # of employees) for firm i in year t.	COMPUSTAT
S&P500	Binary variable if the firm is in S&P 500 index, and zero otherwise	S&PCapital IQ Database

Appendix C: Patent Classification Methods

USPC Classification Example (002-030)		IPC Classification Example	
Class Number	Class Title	Class Identifier	Class Title
002	Apparel	A01	AGRICULTURE; FORESTRY; ANIMAL HUSBANDRY; HUNTING; TRAPPING; FISHING
004	Baths, closets, sinks, and spittoons	A21	BAKING; EQUIPMENT FOR MAKING OR PROCESSING DOUGHS; DOUGHS FOR BAKING
005	Beds	A22	BUTCHERING; MEAT TREATMENT; PROCESSING POULTRY OR FISH
007	Compound tools	A23	FOODS OR FOODSTUFFS; THEIR TREATMENT, NOT COVERED BY OTHER CLASSES
008	Bleaching and dyeing	A24	TOBACCO; CIGARS; CIGARETTES; SMOKERS' REQUISITES
012	Boot and shoe making		PERSONAL OR DOMESTIC ARTICLES
014	Bridges	A41	WEARING APPAREL
015	Brushing, scrubbing, and general cleaning	A42	HEADWEAR
016	Miscellaneous hardware	A43	FOOTWEAR
019	Textiles: fiber preparation	A44	HABERDASHERY; JEWELLERY
023	Chemistry: physical processes	A45	HAND OR TRAVELLING ARTICLES
024	Buckles, buttons, clasps, etc.	A46	BRUSHWARE
026	Textiles: cloth finishing	A61	MEDICAL OR VETERINARY SCIENCE; HYGIENE
027	Undertaking	A62	LIFE-SAVING; FIRE-FIGHTING
028	Textiles: manufacturing	A63	SPORTS; GAMES; AMUSEMENTS
029	Metal working		
030	Cutlery		

Vita

Blake Rayfield was born in Ocala, Florida. He obtained his Bachelor's degree in Business Administration with a concentration in Finance at Florida International University in 2014. He joined the UNO Financial Economics Ph.D. in August 2014. He was awarded a Masters of Science, Financial Economics in 2016. He was ultimately awarded his Ph.D. in Financial Economics May 2018.