Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies ICIS 2023

Data Analytics for Business and Societal Challenges

Dec 11th, 12:00 AM

Innovation Novelty and Firm Value: Deep Learning based Text Understanding

Wei HU
Tongji University, weihu24-c@my.cityu.edu.hk

Yuk Yee CHAN
City University of Hong Kong, yychan382-c@my.cityu.edu.hk

Jianming HUANG
City University of Hong Kong, jhuang339-c@my.cityu.edu.hk

Wanyue Zhou
City University of Hong Kong, wanyuzhou2-c@my.cityu.edu.hk

Xin Li
City University of Hong Kong, xin.li@cityu.edu.hk

Innovation Novelty and Firm Value: Deep Learning-Based Text Understanding

Short Paper

Yuk Yee Chan
City University of Hong Kong
Tat Chee Avenue, Hong Kong, China
yychan382-c@my.cityu.edu.hk

Wei Hu[1]
Tongji University
NO. 1239 Siping Rd, Shanghai, China
huwei72@tongji.edu.cn

Jianming Huang
City University of Hong Kong
Tat Chee Avenue, Hong Kong, China
jhuang339-c@my.cityu.edu.hk

Wanyue Zhou
City University of Hong Kong
Tat Chee Avenue, Hong Kong, China
wanyuzhou2-c@my.cityu.edu.hk

Xin Li
City University of Hong Kong
Tat Chee Avenue, Hong Kong, China
Xin.Li.PhD@gmail.com

Abstract

Innovation is widely acknowledged as a key driver of firm performance, with patents serving as unique indicators of a company's technological advancements. This study aims to investigate the impact of textual novelty within patents on firm performance, focusing specifically on biotechnology startups listed on the Nasdaq. Utilizing deep learning-based approaches, we construct measures for semantic originality in patent texts. Through panel vector autoregressive (VAR) analysis, our empirical findings demonstrate a positive correlation between textual novelty and abnormal stock returns. Further, impulse response function analysis indicates that the impact of textual novelty peaks approximately one week after patent issuance and gradually diminishes within a month. These insights offer valuable contributions to both the theoretical understanding and practical application of innovation management and strategic planning.

Keywords: Innovation, patent, text analysis, stock market, deep learning

Introduction

Innovation, especially pioneering breakthroughs, serves as a significant driver of organizational sustainability in the competitive and mercurial financial market. Meanwhile, the uncertainty of payoff for innovation expenditure poses a challenge in corporate decision-making (Schwartz, 2006). Such uncertainty has attracted researchers to analyze the impact of innovation on firm performance, which is crucial for entrepreneurs and investors.

Patents, which uncover technological details in exchange for legal protection, stand as a major representation of innovation. They have frequently been regarded as signals of an organization's innovation capacity, intellectual property, and research and development (R&D). Various metrics related to patents, such as the number of patents, the number of patent citations, and the number of patent claims, serve as

[1] Please send all correspondence to Wei HU (huwei72@tongji.edu.cn).

quantifiable measures of innovation (European et al., 2021; Lerner, 1994). However, these measures do not fully capture the relationship between patents and business success. Previous studies reported both positive (Useche, 2014) and negative (Teece, 1986) relationships between patent ownership and firm value.

While breakthrough innovations can bring potential to a business, firm value reflects the market's assessment and understanding of this potential. Thus, effective and timely identification of the novelty of patents would affect how fast the market recognizes the impact of innovation on a firm. From this perspective, we argue the textual content of a patent serves as an important channel to facilitate investors' understanding. Based on their reading of the textual content, investors need to recognize the invention's uniqueness from existing technologies and its novelty (Shi & Evans, 2023). Thus, in this study, we focus on textual patent novelty to examine innovation novelty's impact on firm value. To address the limitations of existing methods, this study introduces measures of textual patent novelty utilizing BERT (Bidirectional Encoder Representations from Transformers) and a deep learning VAE (Variational Autoencoder) model, thereby enabling more accurate and instantaneous evaluations of patent innovativeness.

To evaluate the efficacy of our proposed measure of textual patent novelty, we conduct an empirical investigation focusing on the impact of textual patent novelty on firm value for biotechnology startup companies—a field highly reliant on patents and innovation for success. After integrating the data from the Nasdaq exchange and the United States Patent and Trademark Office (USPTO), we built a panel dataset comprising 111 biotechnology startups and spanning the years 2008 to 2019. We conduct panel VAR analysis on the dataset and control the well-established confounding factors (e.g., market status, patent amount, and existing patent metrics). We find that textual patent novelty has a positive and significant impact on firms' abnormal returns. This impact is larger than that of traditional novelty attributes. Further, the impulse response functions (IRFs) analysis reveals that the impact of textual patent novelty on firm performance peaks in approximately one week and then diminishes within one month. Our findings have significant implications for the field of innovation management and strategic planning for entrepreneurs.

Literature Review

Innovation and Firm Value

Patents, which disclose details of a company's technology in exchange for legal protection, serve as a crucial indicator of innovation. While the impact of innovation on firm value is well-studied, the magnitude and direction of this impact remain a subject of ongoing debate. On the one hand, technological breakthroughs may lead to substantial economic returns. Studies such as Lerner (1994) demonstrated that the breadth of patent protection significantly affects valuations. Useche (2014) found a significant and robust positive correlation between patent applications and IPO performance, which is moderated by regional variations in patent obtainability. On the other hand, Teece (1986) argued that innovators often fail to benefit from their innovations. Fitzgerald (2007) pointed out that the fuzzy and ambiguous nature of software inventions renders software patents less impactful on a company's earnings. Many practitioners (e.g., venture capitalists) have been skeptical about patent values (Graham & Mowery, 2006). Due to this debate, proper measurement of patents and innovation becomes important in innovation management.

First, the volume of patents published, often referred to as patent amount, serves as a common measure of innovation capability. Previous studies showed that patent amount has a direct effect on the financial market for both the firm and its rivals (McGahan & Silverman, 2006). Second, citations of patents are often used to measure innovation impact. Deng et al. (1999) revealed that citation is positively associated with the ratio of market value to book value. Hall et al. (2005) provided evidence that the citations of companies' patents are contemporaneously associated with their market value. Third, the number of patent claims or patent classes often represents the technological diversity and scope of patents. Tong and Frame (1994) showed that the number of claims outperforms the number of patents when measuring the technological capacity of a company. Lerner (1994) showed that patent scope significantly affects the firm value in venture financing. Lastly, the number of patent offices in which a patent is registered can signal the firm's valuation of the patent. Filing in multiple patent offices entails higher costs, suggesting that firms reserve this approach for particularly valuable patents (European et al., 2021).

Innovation Novelty

Going beyond the above measures, patent novelty represents a more substantial aspect of innovation. Breakthrough innovations have the potential to push the enterprise to a new level. A pioneering patent could not only create a new technology direction (or even a technical field) but also exert a considerable impact on the stock market (Srinivasan et al., 2009). Li et al. (2020) showed that introducing new methods or improvements on a process, machine, manufacture, or composition of matter has a significant influence on firm value. In the empirical studies examining the impact of patent novelty on firm value, three primary types of measures have been employed.

- **Category-based measures**: In literature, novelty is often considered as the new recombination of knowledge from different domains. Hirshleifer et al. (2018) proxied innovative originality using the average number of unique technological classes cited in the company's patents. They found that innovative originality is associated with less volatile profitability and higher abnormal stock returns.
- **Citation-based measures**: Citation-based measures focus on the citations a patent receives, and novel patents should receive more citations. Trajtenberg et al. (1997) showed that the citation count can be a viable indicator of patent novelty.
- **Content-based measures**: Content-based novelty measures focus on the words and sentences of the patent description and claim section. The purpose is to better understand the semantics of patents to infer their novelty. In this area, the research is limited. Hogenboom et al. (2021) evaluated the merits of word sense disambiguation in event-based stock price prediction and found that word sense disambiguation leads to 70% higher excess returns in the accuracy of the buy and sell signals.

Research Gap

While there are extensive studies on patents, studies specifically focusing on the novelty expressed in patent text are relatively limited. Meanwhile, textual content in patent documents contains rich information related to the technical details of innovation. To investors, text conveys important cues for understanding the invention's uniqueness from existing technologies (Shi & Evans, 2023) and project its potential value. Therefore, the development of more advanced measures for accessing innovation novelty based on patent text is essential for evaluating its impact on firm value, and we aim to address this gap in this study.

Textual Innovation Novelty Measurement Development

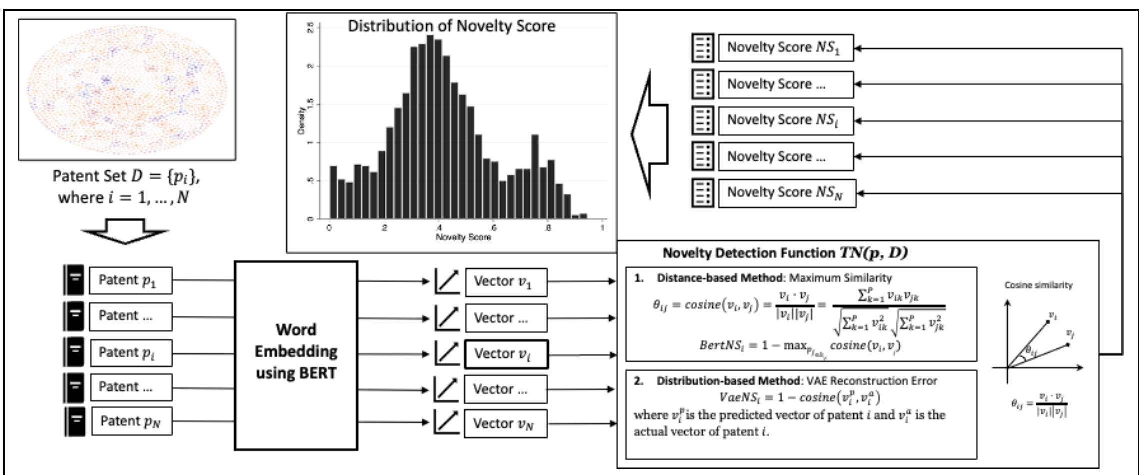

Figure 1. Diagram of Textual Patent Novelty Measure Generation Process

The textual content of a patent represents the patent's technical innovation through the semantics of descriptions and may also introduce new terminologies/concepts, which signals technology advances. Such descriptions, as compared with existing patents, could provide investors clues to the importance of the invention. In this study, we construct measures for textual patent novelty by examining the textual difference between a given patent and its antecedent patents. A higher degree of textual difference in semantic expressions and terminologies suggests that the patent is more likely engaged in exploring some emerging technology.

Prior studies take an outlier view of novel patents (Wang & Chen, 2019). Given a new patent document p and a set of existing documents $D = [p_i]$, where $i = 1, \dots, N$, the objective of textual novelty detection is to define a function *TN(p, D)* that quantifies the novelty of a patent p given the existence of D. In the literature, the distance-based method (Hautamaki et al., 2004) and distribution-based methods (Mei et al., 2018) are the state-of-the-art novelty detection methods to check whether new data is farther away from the existing data. The crucial step involves transforming the text documents into vectors through a technique known as word embedding.

For the distance-based method, we utilize the concept of maximum similarity to measure the patent novelty (Luo et al., 2022). We employ BERT, a pre-trained model released by Google, to vectorize the patent documents. BERT, trained using Wikipedia and book corpus data (Devlin et al., 2018), offers an advantage over other contextual embedding models by dynamically generating word representations based on surrounding words. Specifically, we use Sentence-BERT (Reimers & Gurevych, 2019), a modified version of BERT, to map sentences and paragraphs to a fixed 384-dimensional dense vector space. The similarity is quantified using the cosine similarity of document-based vectors. We calculate the BERT-based novelty score using maximum similarity between patent document p_i and every other patent document p_j before the date of p_i in the document set $D_{\backslash i}$ (document set excluding p_i). For the distribution-based similarity, we measure patent novelty using reconstruction errors. We employ a deep learning Variational Autoencoder (VAE) model (Kingma & Welling, 2013) to encode the distribution of training data and extract low-dimensional representations of new patent documents. Following the methodology of Mei et al. (2018), we construct a 7-layer VAE model to map the 384-dimension BERT vectors to a normal distribution. The reconstruction error is calculated as the cosine similarity of the vectors predicted by VAE and actual vectors. We normalize the novelty score without changing its magnitude. As shown in Figure 1, the patent novelty score NS_i falls within the range [0,1], with larger values indicating greater innovation. Both BERT and VAE effectively capture the semantics of textual content, allowing us to extract the essence of the patents.

Novelty Measures	Pearson	Kolmogorov-Smirnov
TF-IDF (Baseline)	0.679	0.698
VAE	0.782	0.865
BERT	0.808	0.802

Table 1. Novelty Measure Performance Comparison

Before applying our textual patent novelty measures to study their impact on firm performance, we need to evaluate their validity and reliability. As a baseline, we employ the maximum similarity measure based on TF-IDF representations. To access the consistency and robustness of our measures, we utilize the *20 Newsgroups Dataset*, a benchmark dataset comprising 18,828 text documents across 20 categories. We arbitrarily designate three classes—*alt.atheism, comp.graphics,* and *comp.os.ms-windows.misc*—as normality classes and one class, *rec.motorcycles*, as a novelty class, following the common practice (Bhattarai et al., 2020). We calculate the novelty scores for both normal and novel documents using TF-IDF, BERT, and VAE-based measures. Following Shibayama et al. (2021), we evaluate these three novelty scores using various statistical tests, including the Pearson correlation coefficient and Kolmogorov-Smirnov statistics. As shown in Table 1, both the BERT-based and VAE-based novelty scores outperform the baseline measure. Consequently, we employ both BERT-based and VAE-based novelty scores in our subsequent empirical analysis.

Econometric Model

Dependent Variable: Firm Performance

In this study, we adopt the abnormal return as a measure for accessing firm equity value in alignment with Luo et al. (2013). An abnormal return is defined as the difference between the actual return and the expected return. To measure it, we first estimate the abnormal return using the Fama-French three-factor model (Fama & French, 1992):

$$AR_{i,t} = R_{i,t} - R_{f,t} - (\alpha_i + \beta_{1i}(R_{m,t} - R_{f,t}) + \beta_{2i}SMB_t + \beta_{3i}HML_t) \quad (1)$$

where $R_{m,t}$ is the average market return; $R_{f,t}$ is the risk-free rate of return; SMB_t is the effect of market size; HML_t is the effect of market value; α_i is the intercept; $R_{i,t}$ is the stock return for firm i at time t, where P_t is the stock price at time t; and Δt is a unit interval of time. We apply log transformations here.

Panel VAR Model

We use the panel vector autoregressive (VAR) model to examine the dynamic interaction between a firm's textual patent novelty and firm performance. Using the panel VAR model, we can account for biases of endogeneity and autocorrelations and quantify the short-/long-term impacts of textual patent novelty in predicting firm value. We specify the model as follows:

$$V_{i,t} = \sum_{k=1}^{K} \Phi_i^k \cdot V_{i,t-k} + \delta_t + f_i + \varepsilon_{i,t} \quad (2)$$

where $V_{i,t} = [AR_{i,t}, X_{i,t}, Ctrl_{i,t}]$ is the vector for firm i at time t. $\mathrm{X}_{i,t}$ is textual patent novelty for firm i at time t (see Figure 1); ϕ_i^k is the coefficient; K is the number of lags; $Ctrl_{i,t-1}$ is the vector of control variables; δ_t is a vector of time-fixed effects; f_i is a vector of unobserved individual fixed effects, characterizing firms' time-invariant attributes; and $\varepsilon_{i,t}$ is a vector of errors. To deal with the biased within-group estimator for the fixed effects model, we estimate the proposed panel VAR model using the standard GMM estimator following Binder et al. (2005). To address the concern that too many instruments may be used, we use Hansen's J statistic of overidentifying restriction to test the validity of instruments for a chosen lag. Our econometric analysis follows prior literature (Luo et al., 2013; Tirunillai & Tellis, 2012).

Control Variables

To demonstrate the efficacy of our developed textual patent novelty measure, we control for several existing novelty attributes, novelty measures, and market activity indicators as follows: (1) *RefCnt* represents the sum of the reference counts for patents within a given time range, with missing values filled in as zero. Log transformations are applied here. (2) *NoPatent* serves as a dummy variable indicating whether there is a patent published during a specific period; it takes the value of 1 when no patent is published and zero otherwise. (3) *PatentCnt* represents the number of patents published within a specific time range, with missing values filled in as zero. (4) *Overlap* is calculated following the methodology of Dahlin and Behrens (2005), representing 1 minus the ratio of co-occurrence of a reference in a patent and its preceding patents; missing values are filled in as zero. (5) *Volume* represents the number of shares traded in a single day. We apply log transformation here. (6) *Turnover* is calculated as the standard deviation of residuals in Equation (1), based on a rolling window of 30 trading days before the target day (Luo et al. 2013).

Empirical Study

Dataset

In this study, we focus on high-tech companies in their early stages to examine the impact of textual patent novelty on firm value, with a particular focus on the biotechnology industry due to its heightened dependency on innovation. According to internal statistics, 82% of U.S. biotechnology companies are listed on the Nasdaq stock market. We obtain the list of biotechnology firms from the U.S. Securities and Exchange Commission (SEC) and collect stock market data for these companies from the Nasdaq. Companies are classified based on their Standard Industrial Classification (SIC) Codes, specifically *SIC:*

2836—Biological Products, Except Diagnostic Substances, as indicated in the EDGAR filings. Finally, we collect daily stock prices and trading volumes for 111 biotechnology companies from December 31, 2007, to December 31, 2019. To avoid the influence of the Covid-19 pandemic, we restrict our analysis to data up to the end of 2019. Patent data is obtained from the United States Patent and Trademark Office (USPTO). We identify 4,878 patents assigned to these biotechnology companies from 2008 to 2019. Additionally, we compile a comprehensive dataset of USPTO patents from January 1, 2005, to October 18, 2022, to include patent citation and reference data. During our period of interest (2008-2019), we observe a noticeable upward trend in the number of patents.

Summary Statistics

We have aggregated the dataset into weekly panel data. Table 2 displays the descriptive statistics for the variables employed in our analysis throughout the study period. It should be noted that certain observations for the abnormal returns, volume, and turnover are missing for various reasons, such as a company not being listed for part of the study period. The dataset covers 310 weeks. We find that the average weekly abnormal return (AR) is -1% ($e^{-0.01} - 1$). Most of the abnormal returns (AR) are negative, indicating that the actual returns are frequently lower than the expected returns in the biotechnology industry. The average weekly BERT-based (VAE-based) novelty score $BertNS$ ($VaeNs$) is 0.003 (0.006). $BertNS$ ($VaeNs$) ranges from 0 to 0.896 (0 to 0.945). Most of the observations have small values. Remarkably, no patents were published for 98.6% of the weekly observations, showing that there is no innovation impact on the market most of the time. The average weekly log-transformed patent count ($PatentCnt$) is 0.024, indicating that a patent is issued almost every 41.67 weeks (1/0.024), which is a long innovation cycle of almost 10 months.

Variable	Interval	N	Mean	Std. Dev.	Min.	Median	Max.
AR	Week	34,030	-0.010	0.131	-2.297	-0.008	3.817
BertNs	Week	122,070	0.003	0.043	0.000	0.000	0.896
VaeNs	Week	122,070	0.006	0.055	0.000	0.000	0.945
NoPatent	Week	122,070	0.986	0.119	0.000	1.000	1.000
RefCnt	Week	122,070	0.019	0.198	0.000	0.000	5.357
PatentCnt	Week	122,070	0.024	0.238	0.000	0.000	9.000
Overlap	Week	122070	0.006	0.065	0.000	0.000	1.000
Volume	Week	34,171	10.195	3.767	0.000	11.175	17.945
Turnover	Week	33,923	0.099	0.283	0.0004	0.070	22.332

Table 2. Descriptive Statistics

Panel VAR Results

For the panel VAR model specification, we carry out the lag selection procedures following Andrews and Lu (2001), and the first-order panel VAR is selected because it has the smallest MBIC, MAIC, and MQIC. We use weekly datasets in panel VAR to investigate the short-/long-term effect of textual patent novelty on the firm value. We checked the stability condition of panel VAR estimates by calculating the modulus of each eigenvalue of the fitted model and confirmed that the estimated panel VAR model satisfies the stability condition. The results of Granger causality tests suggest that the causal mechanism is established from innovation novelty to firm value rather than from firm value to innovation novelty. The test of the overidentifying restrictions fails to reject at the 10% level, showing that all instruments are significant at the 10% level or better.

Table 3 reports the parameter estimates of the panel VAR model. Our primary goal is to ascertain the impact of textual patent novelty on firm value while controlling for other factors. The coefficient on one-period lagged dependent variables offers insights into the short-term effects. As shown in column (1) in Table 3, the coefficient of $VaeNs$ at lag 1 is 0.250, which is positive and significant at the 1% level, suggesting that one standard deviation increase in VAE-based textual novelty (0.055) would lead to a 1.384% ($e^{0.055\times0.250}$ −

1) increase in abnormal return in the subsequent week. Furthermore, column (2) corroborates these findings for *BertNs*. According to prior literature (Heeley et al., 2007), textual patent novelty influences financial markets through investors' understanding. Aligning with the elaboration likelihood model (Cacioppo & Petty, 1984), the effect of textual patent novelty—or semantic originality—on firm value depends on investors' text-processing ability. Our findings imply that when pioneering patents are issued in the biotechnology sector, investors are capable of quickly digesting the patent content and responding promptly, often within one week.

	(1)	(2)	(3)
	$AR_{i,t}$	$AR_{i,t}$	$AR_{i,t}$
$AR_{i,t-1}$	-0.018	-0.021*	-0.018
	(0.011)	(0.011)	(0.011)
$VaeNS_{i,t-1}$	0.250***		
	(0.019)		
$BertNS_{i,t-1}$		0.159***	
		(0.016)	
$TfidfNS_{i,t-1}$			0.127***
			(0.015)
$NoPatent_{i,t-1}$	0.281***	0.280***	0.282***
	(0.019)	(0.017)	(0.020)
$RefCnt_{i,t-1}$	0.043***	0.034***	0.022***
	(0.007)	(0.005)	(0.006)
$PatentCnt_{i,t-1}$	0.043***	0.034***	0.047***
	(0.007)	(0.005)	(0.008)
$Overlap_{i,t-1}$	0.122***	0.079***	0.124***
	(0.016)	(0.012)	(0.016)
$Volume_{i,t-1}$	-0.005***	-0.005***	-0.005***
	(0.0005)	(0.000513)	(0.0005)
$Turnover_{i,t-1}$	-0.019***	-0.017***	-0.020***
	(0.005)	(0.005)	(0.005)
# of Obs.	33,709	33,709	33,709
# of Firms	109	109	109

Table 3. Results of the Panel VAR Model

From column (1) in Table 3, the coefficient of *Overlap*, which is positive (0.122) and significant, suggests that a one standard deviation increase in *Overlap* (0.065) would lead to an 0.796% ($e^{0.122\times0.065} - 1$) increase in abnormal return in the next week. This reveals that both the technology uniqueness (*Overlap*) and semantic originality (*VaeNs*) of patents have positive impacts on firm value. However, semantic originality appears to exert a greater impact. The possible reason is that for biotechnology innovations, investors pay more attention to details of interpretation on novelty rather than simple references and categories. The results of the Granger causality tests suggest that investment activities in the biotechnology industry will be affected by the content of patent documents. Importantly, we do not find evidence that the patent applicants intentionally manipulated the semantical originality of the patent documents. To quantify the effect of the change in dependent variables lagged more than one period, impulse response functions (IRFs) are often used to visually interpret the coefficient estimates generated for panel VAR models by simulating the fitted panel VAR model through a Monte Carlo simulation with 1,000 runs. As shown in Figure 2, the IRF results reveal that the effect of the textual patent novelty score on firm value lasts for about one week and gradually decreases to zero as the effect eventually dies out.

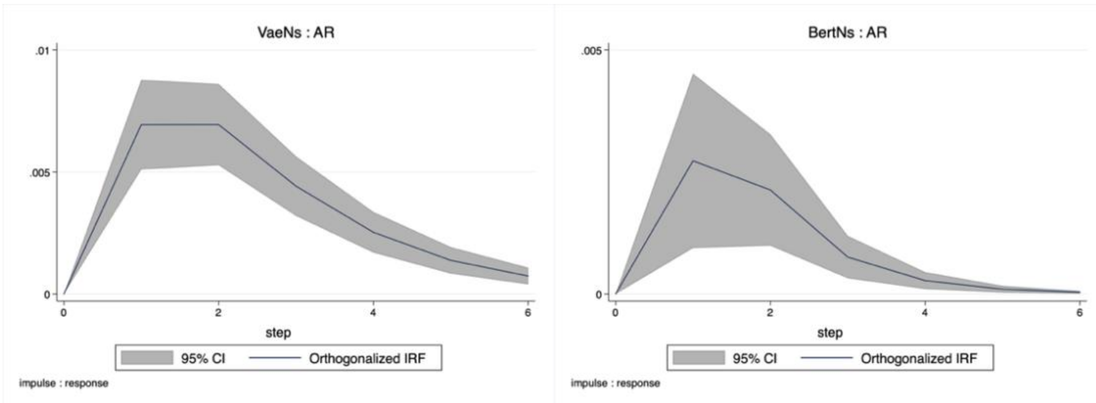

Figure 2. IRF of AR Due to a Shock to Textual Patent Novelty (a) VAE; (b) BERT

Discussion and Conclusions

This study introduces deep learning-based approaches, utilizing both VAE and BERT algorithms, to measure textual patent novelty and examine its impact on firm value. Employing the panel VAR model, we find that textual patent novelty positively influences firms' abnormal returns. The impulse response functions (IRFs) results reveal that the impact of textual patent novelty on firm performance peaks within one week and then diminishes within one month. Our findings underscore the significance of innovation and enrich our understanding of patent text by providing a state-of-the-art measure for innovation.

Our study also has its limitations. First, our analysis is confined to the early stages of the biotechnology industry. Explorations in other industries and of firms at other stages will be beneficial. Second, this study focuses exclusively on U.S. patents. Since some companies can be multinational, it would be useful to explore patents filed in other countries. Third, future studies can further improve the endogeneity in identification, such as by including additional control variables or employing alternative econometric models. Lastly, additional metrics may exist that can further elucidate the relationship between innovation novelty and firm value. Future research will aim to address these limitations to better understand innovation's impact on firms.

Acknowledgements

This work was partially supported by Research Grants Council of the Hong Kong Special Administrative Region, China [GRF 11500519], and City University of Hong Kong [SRG 7005767].

References

Andrews, D. W., & Lu, B. (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of econometrics, 101*(1), 123-164.

Bhattarai, B., Granmo, O.-C., & Jiao, L. (2020). Measuring the novelty of natural language text using the conjunctive clauses of a tsetlin machine text classifier. *arXiv preprint arXiv:2011.08755*.

Binder, M., Hsiao, C., & Pesaran, M. H. (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory, 21*(4), 795-837.

Cacioppo, J. T., & Petty, R. E. (1984). The Elaboration Likelihood Model of Persuasion. *Advances in Consumer Research, 11*, 673-675. <Go to ISI>://WOS:A1984TK41700135

Dahlin, K. B., & Behrens, D. M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. *Research Policy, 34*(5), 717-737.

Deng, Z., Lev, B., & Narin, F. (1999). Science and technology as predictors of stock performance. *Financial Analysts Journal, 55*(3), 20-32.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

European, C., Directorate-General for, R., Innovation, Hoogland, O., Torres, P., Janzow, N., & Kralli, A. (2021). *Patents as a measure of innovation performance : selection and assessment of patent indicators : provision of technical assistance and study to support the development of a composite indicator to track clean-energy innovation performance of EU members*. Publications Office. https://doi.org/doi/10.2777/77424

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance, 47*(2), 427-465.

Fitzgerald, M. (2007). 'A Patent Is Worth Having, Right? Well, Maybe Not. *New York Times*(July 15).

Graham, S. J., & Mowery, D. C. (2006). The use of intellectual property in software: implications for open innovation. *Open Innovation: Researching a New Paradigm*, 184-204.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of economics*, 16-38.

Hautamaki, V., Karkkainen, I., & Franti, P. (2004). Outlier detection using k-nearest neighbour graph. Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.,

Heeley, M. B., Matusik, S. F., & Jain, N. (2007). Innovation, appropriability, and the underpricing of initial public offerings. *Academy of Management Journal, 50*(1), 209-225.

Hirshleifer, D., Hsu, P.-H., & Li, D. (2018). Innovative originality, profitability, and stock returns. *The Review of Financial Studies, 31*(7), 2553-2605.

Hogenboom, A., Brojba-Micu, A., & Frasincar, F. (2021). The impact of word sense disambiguation on stock price prediction. *Expert systems with applications, 184*, 115568.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Lerner, J. (1994). The importance of patent scope: an empirical analysis. *The RAND Journal of Economics*, 319-333.

Li, Z., Liao, G., & Albitar, K. (2020). Does corporate environmental responsibility engagement affect firm value? The mediating role of corporate innovation. *Business Strategy and the Environment, 29*(3), 1045-1055.

Luo, X., Zhang, J., & Duan, W. (2013). Social media and firm equity value. *Information Systems Research, 24*(1), 146-163.

Luo, Z., Lu, W., He, J., & Wang, Y. (2022). Combination of research questions and methods: A new measurement of scientific novelty. *Journal of Informetrics, 16*(2), 101282.

McGahan, A. M., & Silverman, B. S. (2006). Profiting from technological innovation by others: The effect of competitor patenting on firm value. *Research Policy, 35*(8), 1222-1242.

Mei, M., Guo, X., Williams, B. C., Doboli, S., Kenworthy, J. B., Paulus, P. B., & Minai, A. A. (2018). Using semantic clustering and autoencoders for detecting novelty in corpora of short texts. 2018 International Joint Conference on Neural Networks (IJCNN),

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Schwartz, J. (2006). The five founding principles That drive innovation. *The Financial Times*.

Shi, F., & Evans, J. (2023). Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications, 14*(1), 1641.

Shibayama, S., Yin, D., & Matsumoto, K. (2021). Measuring novelty in science with word embedding. *PloS one, 16*(7), e0254034.

Srinivasan, S., Pauwels, K., Silva-Risso, J., & Hanssens, D. M. (2009). Product innovations, advertising, and stock returns. *Journal of Marketing, 73*(1), 24-43.

Teece, D. J. (1986). Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy, 15*(6), 285-305.

Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science, 31*(2), 198-215.

Tong, X., & Frame, J. D. (1994). Measuring national technological performance with patent claims data. *Research Policy, 23*(2), 133-141.

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. *Economics of Innovation and new technology, 5*(1), 19-50.

Useche, D. (2014). Are patents signals for the IPO market? An EU–US comparison for the software industry. *Research Policy, 43*(8), 1299-1311.

Wang, J., & Chen, Y.-J. (2019). A novelty detection patent mining approach for analyzing technological opportunities. *Advanced Engineering Informatics, 42*, 100941.