Technovation 114 (2022) 102434

Contents lists available at ScienceDirect

Technovation

journal homepage: www.elsevier.com/locate/technovation

Measuring open innovation practices through topic modelling: Revisiting their impact on firm financial performance

Qinli Lu [a,*], Henry Chesbrough [b,c]

[a] Institute of Management & Strategy, University of St. Gallen, Switzerland
[b] Institute for Business Innovation, University of California, Berkeley, United States
[c] Maire Tecnimont Professor of Open Innovation and Sustainability, Luiss University, Rome, Italy

ARTICLE INFO

Keywords:
Open innovation
Practices
Topic modelling
Natural language processing
Financial performance

ABSTRACT

Despite the popularity of open innovation in recent years, studies examining the impact of open innovation upon firm performance have shown mixed results. Previous empirical work on this topic is often based on surveys or archival sources, usually done either in isolation or in aggregate through employing proxy measures. In contrast, we employ an unsupervised learning technique (i.e., topic modelling) utilizing natural language processing to extract information on companies' open innovation practices, creating an initial keyword basket for future development. We then revisit the relationship between open innovation practices and financial performance of firms. The results show that a firm's overall openness level is associated with improved financial performance. More granular practices developed from our approach, however, show variations. The inverted U-shaped relationships are observed in specific open innovation practices but not in all, partly supporting the existence of the openness paradox from prior literature. The complementarity between internal R&D and individual open innovation practices also varies by practice. Further, the influence of these open innovation practices also varies by sector. Our findings prompt us to conclude that open innovation's impact on financial performance is nuanced, and that there is no uniform set of best practices to practice open innovation effectively.

1. Introduction

With firm boundaries becoming more permeable, companies are increasingly embracing open innovation as part of their innovation strategies (Chesbrough, 2019; Shaikh and Levina, 2019). Defined as "a distributed innovation process based on purposively managed knowledge flows across organizational boundaries, using pecuniary and non-pecuniary mechanisms in line with the organization's business model", open innovation has attracted considerable academic attention in recent years (Bogers et al., 2017; Chesbrough and Bogers, 2014). Its impact has spread well beyond academic studies of industrial innovation processes. For instance, LinkedIn now lists more than 500,000 people with open innovation as part of their job, and more than 4600 job postings that involve open innovation.[1] Surveys of large firms practicing open innovation also showed that, while firms were still early in their understanding of how best to practice open innovation, their adoption of many of its practices was on the increase (Brunswicker and Chesbrough, 2015, 2018).

Many studies of open innovation look at broad dimensions of open innovation—i.e., outside-in open innovation, inside-out open innovation and coupled open innovation (e.g., Bianchi et al., 2016; Cassiman and Valentini, 2016; Greco et al., 2016; Lichtenthaler, 2009). Other empirical studies mainly focus on one or two open innovation practices (OIPs), like crowdsourcing (Liu et al., 2020), search breadth and search depth (Laursen and Salter, 2014), or external technology acquisition and exploitation (Hung and Chou, 2013). We define an OIP in this paper as the actual application or use of the ideas or methods based upon open innovation that companies implement for their innovation-related processes (Ebersberger et al., 2012; Spithoven, 2013; Spithoven et al., 2013), a definition we operationalize below in our methods section. There are only a limited number of studies that compare fine-grained OIPs (see examples from Ahn et al., 2015; Cheng and Huizingh, 2014; Mazzola et al., 2012). Moreover, these assessments of different OIPs are largely based on survey methods (e.g., Brunswicker and Chesbrough, 2018; Chen et al., 2016), and the relevant constructs derived from surveys are more about measuring innovation performance than firm

* Corresponding author.
E-mail addresses: qinli.lu@unisg.ch (Q. Lu), chesbrou@berkeley.edu (H. Chesbrough).

[1] https://www.linkedin.com/jobs/search/?keywords=Open%20Innovation (last accessed August 11, 2020).

https://doi.org/10.1016/j.technovation.2021.102434
Received 1 February 2021; Received in revised form 9 November 2021; Accepted 20 November 2021
Available online 3 December 2021
0166-4972/

financial performance. Survey methods also involve some limitations, including self-reported bias (Zobel et al., 2016), quality issues (Köhler et al., 2012),and the scale format and anchors used in survey research can systematically affect results (Podsakoff et al., 2003).

Recently, developments in natural language processing (NLP), and in particular unsupervised topic modelling techniques, have the potential to address some of these limitations. They have been recognized in other areas of social science as powerful tools to analyze large unclassified textual data sources, which can create possible measures for our research interests (Alghamdi and Alfalqi, 2015; Teodoridis et al., 2020). Moreover, growing quantities of unstructured digitized text are available for research with the support of text mining approaches such as topic modelling techniques (Antons et al., 2020). In this paper, we use them to discover fine-grained OIPs, providing insights into the differences among various OIPs and their relationship with firm financial performance.

Our results show that NLP can capture nuanced variation in OIPs between firms. Our primary finding shows that a firm's overall openness is significantly and positively associated with improved firm financial performance. With our more granular treatment of individual OIPs, though, we find that this association varies significantly across practices and across sectors. The overall openness level does not show a curvilinear correlation with financial performance, yet certain OIPs do follow an inverted U-shape pattern with firm financial performance. The complementarity between internal R&D and individual OIPs is sometimes positive, but other times negative. Moreover, practices that are positively and significantly associated with improved financial performance in one sector might show no significant effect in a different sector. And the salience of OIPs themselves varies largely by sector as well. These findings reveal a subtle, nuanced relationship between open innovation, individual practices, and firm performance. They suggest, among other things, that there are likely to be no single set of best practices in open innovation that would work in all sectors of the economy.

2. Theoretical background and literature review

2.1. Heterogeneous perspectives on open innovation

With increasing technological and environmental uncertainty and complexity, firms are relying on open innovation to ensure long-term competitiveness (Chesbrough, 2003). Firms, therefore, are evolving in their practices of open innovation, engaging in a variety of activities, from bilateral transactions to collaborations that involve multiple parties in an interactive relationship (Brunswicker and Chesbrough, 2018). Yet, the adoption of open innovation is not only a source of opportunities for companies as it can also present risks (Marullo et al., 2020).

On the one hand, by enabling knowledge flows across organizational boundaries, one of the main characteristics of open innovation is the involvement of external parties in the innovation process (Laursen and Salter, 2006; Chesbrough and Bogers, 2014). Companies can acquire needed knowledge from third parties (e.g., customers, competitors, collaborators), and use this augmented knowledge to improve innovative capabilities (Huizingh, 2011), to strengthen competitive advantage (Foss et al., 2010) , and to generate greater business performance (Oltra et al., 2018). On the other hand, however, a potential dark side of adopting open innovation exists as well (Frishammar et al., 2015). For instance, cognitive differences with external partners may occur throughout the process of implementing open innovation projects (Du et al., 2014); Organizational and cultural issues may emerge as a consequence of interacting with external partners (Marullo et al., 2020; Vahter et al., 2014); A focal firm may lose important knowledge to other actors in inter-organizational collaborations, which can lead to knowledge leakage and in turn lower firm performance (Easterby-Smith et al., 2008; Frishammar et al., 2015).

These heterogenous views inevitably spur on a growing theme 'paradox of openness' in the open innovation literature (Arora et al., 2016; Bogers, 2011; Foege et al., 2019; Laursen and Salter, 2014; Wang et al., 2017). And correspondingly, extant empirical studies yield mixed results regarding the use of open innovation. For instance, Laursen and Salter (2006) reported both positive effects – but also curvilinear effects – from the searching modes of open innovation, meaning that after a certain point, the costs of additional openness were greater than its benefits. Lokshin et al. (2008) found evidence of complementarity between internal and external R&D activities. Huizingh (2011) reviewed a variety of evidence on open innovation effectiveness, and reports mixed results. Laursen and Salter (2014) found a "paradox of openness", where the benefits of opening up must be balanced against the risks of unwanted expropriation. West and Kuk (2016) proposed an inherent complementarity of selective openness strategies between open and proprietary components. Cassiman and Valentini (2016), by contrast, found evidence that open innovation's inflows and outflows of knowledge were not complementary to one another, and that R&D costs to generate knowledge outflows could rise faster than the benefits of those outflows. Grimpe and Sofka (2016), who contrasted collaborative vs. transactional open innovation activities regarding knowledge search, found that each had a stronger effect on performance in the presence of the other, making them complementary. Brunswicker and Vanhaverbeke (2015) argued that not all OIPs are beneficial in enhancing firm innovation performance.

One critical reason for these mixed perceptions is that prior studies captured different facets of open innovation, while the phenomenon has been suggested to be broader and more comprehensive in scope (Foege et al., 2019). The aim of leveraging open innovation should include market-based, science-based and transaction-based relationships (Marullo et al., 2020; Vanhaverbeke et al., 2018). Moreover, to facilitate a closer connection between open innovation and firm-specific practices, it has been argued that the general advice implied in the open innovation literature lacks specificity, and more nuanced research on firm-level strategy for open innovation is needed to better understand the specific open innovation strategies (Felin and Zenger, 2020).

2.2. Different measures of open innovation practices

In general, innovation itself can be difficult to measure. Previous empirical studies that seek to demonstrate open innovation's impact on firm performance at an aggregate level suffer from some methodological limitations. It has been noted that the most common measures of OIPs are based on surveys (Asakawa et al., 2010; Lichtenthaler, 2009; Podmetina et al., 2014). Some of the best empirical work at the national level, for instance, derives from the Community Innovation Surveys that resulted from the Oslo Innovation Manual (OECD, 1997). This Manual was constructed in 1997, well before open innovation was introduced as a concept. So studies utilized variables contained in the Community Innovation Surveys (e.g., Greco et al., 2016; Grimpe and Sofka, 2016; Laursen and Salter, 2006) to construct proxy measures for openness, such as counts of the number of external sources of knowledge accessed during the innovation process. These variables were originally collected for other purposes, and some OIPs, such as crowdsourcing, were not recorded in specific survey instruments.

There are other limitations to survey data. For instance, it is quite common to see that the scale format and anchors will systematically affect responses in survey research, and respondent characteristics and expectations are also well recognized as potential sources of method biases (Podsakoff et al., 2003). Moreover, the measures derived from surveys are based upon self-reported and largely qualitative data, which may raise quality issues regarding administration, non-response, and response accuracy (Criscuolo et al., 2010; Grimpe and Kaiser, 2010; Köhler et al., 2012). If a single respondent is both the source of the dependent variable and some of the independent variables, this can introduce personal beliefs into the empirical relationships being studied.

Of course there are benefits to conducting surveys in innovation research, as surveys can provide direct and importance-weighted measures (Grimpe and Kaiser, 2010; Podsakoff et al., 2003). More importantly, they can collect information which cannot be accessed through public datasets, such as the turnover of new products to the firm (Laursen and Salter, 2006), or qualitative judgement data like innovation capabilities (Foss et al., 2011) or strategic capabilities (Ovuakporie et al., 2021).

Another common approach to measure open innovation quantitatively is conducting supplementary quantitative interviews together with surveys in specific industries (e.g., Marullo et al., 2020; Pullen et al., 2012). Some empirical research has validated the contribution of open innovation to firm performance at the project level (Du et al., 2014), with observed differences between more scientifically-based collaborations and more market-oriented collaborations. There is a considerable body of archival research that measured specific activities related to open innovation by utilizing secondary datasets, mainly about alliance relationships (Schilling, 2009), such as investigating strategic technology agreements based on the MERIT-CATI dataset (Duysters and Hagedoorn, 2000). However, these archival datasets typically focus on one specific OIP, rather than covering all types of OIPs. Meanwhile, another research stream took patent data as a proxy to investigate specific technological knowledge flows within an open innovation context (e.g., Li and Tang, 2010; Wadhwa and Kotha, 2006). Such datasets are focused on specific industries or sectors, or specific practices, making it hard to generalize their findings to other industries or sectors.[i]

In a broader scope, a common measure of innovation activity has been the issuance of patents, and citations to those patents (Nagaoka et al., 2010). As with our paper, new research on using patents to measure innovation extracts more insights from unclassified texts, using NLP techniques to construct measures of innovation (e.g., Arts et al., 2020; Vakili and Kaplan, 2020). However, the activities related to open innovation have not been developed to this point in time. Considering the desirability of more nuanced research on firm-level strategy for better open innovation activities (Felin and Zenger, 2020), we attempt to introduce this new methodology to try to improve our ability to identify and measure the interactions between OIPs.[ii]

Following the use of topic modelling techniques in other parts of the social science literature (Antons et al., 2020; Corritore et al., 2020; Kang et al., 2020), we seek to utilize business communications that were constructed to report the financial results of that business, and then employ novel NLP methods to uncover OIPs used by those businesses. These communications can provide another lens to observe an organization's innovation practices. Like patents, these communications also are publicly observable to all, which greatly assists the ability of other scholars to replicate and extend the analyses we report below. And we have posted our keyword basket and source code in a publicly accessible GitHub repository, making it readily observable as well. These virtues, however, do come at a cost. We will discuss the limitations of our approach in the following Methods and Discussion sections of our paper.

2.3. Open innovation practices and firm financial performance

The prior literature has conducted numerous studies to investigate the relationship between open innovation and firm performance (i.e., mainly about innovation performance and financial performance). We reviewed relevant empirical analyses and summarized the main datasets they relied on, the scopes of open innovation, the constructs they built to measure firm performance, as well as the main findings in appendix Table A1 [iii]. A large portion of prior research featured a variety of constructs to measure innovation performance, most of which were derived from surveys (e.g., Grimpe and Sofka, 2016; Laursen and Salter, 2006). By contrast, we focus on investigating the impact of open innovation on firm financial performance (see Section 3.2.4 for more details), an approach that views innovation as a means to an end, in this case financial performance (Chesbrough, 2019).

Studying the relationship between open innovation and firm financial performance is not straightforward because open innovation is a relatively broad concept that comes in various different forms (Chesbrough, 2003; Dahlander and Gann, 2010; Huizingh, 2011). Scholars have studied practices as diverse as crowdsourcing, university partnerships, prize competitions, startup collaborations, corporate venture capital, co-creation with customers and/or with suppliers, intermediaries, and user innovation, all under the rubric of open innovation. As we discussed in Section 2.2, the majority of empirical studies were based on survey data to construct variables for different OIPs (e.g., Greco et al., 2016; Köhler et al., 2012; Sisodiya et al., 2013).

Different effects (positive, negative, curvilinear and not significant) of open innovation activities on firm financial performance have been reported in various contexts (see Table A1). Specifically, while individual practices inspired by open innovation have shown positive results in numbers of studies, more negative results were reported when a more nuanced classification of OIPs was conducted (Ahn et al., 2015; Lichtenthaler, 2015; Mazzola et al., 2012). These disparate findings may be caused by a lack of more nuanced instruments to measure these effects in prior literature, apart from surveys (e.g., Chen et al., 2016; Greco et al., 2016). Owing to the complexity and heterogeneity of the concept of open innovation, which encompasses a variety of innovation activities, there have been substantial challenges to the measurement of open innovation (Ahn et al., 2015; Podmetina et al., 2014; Schroll and Mild, 2012). Thus, one fundamental research direction is to explore new methods to identify individual OIPs in a more fine-grained way, to see the interactions between these practices and examine how these practices influence firm financial performance.

2.4. Internal R&D and open innovation practices

Internal R&D has long been treated as the most prominent resource to firms' competitive advantage (Chen et al., 2016). It not only helps firms build their knowledge base (Veugelers and Cassiman, 1999), contributing to the development of firms' internal knowledge, but also strengthens their absorptive capacity and the resulting ability to use external knowledge (Cohen and Levinthal, 1990; Hung and Chou, 2013; Laursen and Salter, 2006), which is essential for open innovation effectiveness.

In the open innovation process, utilizing internal R&D enables an organization to develop its technological knowledge. Thus, a firm with high R&D capacity has a sufficient degree of relevant knowledge to recognize the value of new ideas and external knowledge flows, and vice-versa (Cohen and Levinthal, 1990; Chesbrough, 2003; Spithoven et al., 2010). By effectively identifying relevant external technological and market-related knowledge, a firm can choose appropriate practices (i.e. license out, co-develop, or cross-license) to advance its innovation process or to gain more profits (Brunswicker and Chesbrough, 2018; Hung and Chou, 2013; Spithoven et al., 2010). Therefore, when a firm invests at a high level of internal R&D activities, it acquires sufficient internal knowledge to appropriate benefits for better performance. These arguments strongly suggest that open innovation would be highly complementary to internal R&D, and associated with better firm financial performance.

These findings, however, become less consistent, once researchers examine more specific OIPs. Prior research reports diverse findings on the relationship between internal R&D and different types of OIPs (Chen et al., 2016). For instance, some studies revealed a complementary relationship between internal research and leveraging external knowledge (e.g., Cassiman and Veugelers, 2006; Rothaermel and Hess, 2007; Schmiedeberg, 2008). Yet some empirical research showed a substitutability relationship between them (e.g., Hess and Rothaermel, 2011; Laursen and Salter, 2006). And some report no significant results about their interactions (Lokshin et al., 2008). One possible reason for the divergent observations may be attributed to the diversity of different OIPs. Different OIPs or knowledge sources imply distinct characteristics

(Chen et al., 2016), wherein the heterogeneity may cause disparate results (Hagedoorn and Wang, 2012). Given the importance of R&D on the adoption of open innovation, as well as its role on firm performance, we also re-evaluate its moderating effect on the relationship between open innovation and firm financial performance, using more fine-grained measures about OIPs in this paper.

3. Data and method

3.1. Data sources

Our empirical analysis is based on unstructured data of how companies describe their business operations. The source of our corpus is the annual reports of U.S. publicly traded companies. We examine historical year-end 10-K filings of 2017, 2018, and 2019 (for the years 2016, 2017, and 2018 fiscal years' business operations, respectively), which are pooled for cross-sectional analysis. We obtained these data from the Security and Exchange Commission (SEC) EDGAR database for Russell 3000 stocks. The Russell 3000 measures the performance of the 3000 largest publicly held companies incorporated in America by considering total market capitalization, and represents approximately 98% of the American public equity market by market value. Since the Sarbanes-Oxley legislation of 2009, CEOs and CFOs must personally sign these statements. In theory, they can be held personally liable for material errors and omissions in these reports. Therefore, there is a strong motivation for accuracy in these reports, and the top management of each company pays close attention to the content in these communications. They also represent one of the key sources of financial information for outside investors, and the SEC pays close attention to guard against misrepresentation or insider disclosure of material non-public information. These communications, thus, are carefully constructed, closely reviewed and broadly disseminated.

In the 10-K fillings, we chose the Business section for text extraction. The Business Section describes a company's business, including its main products and services, and what markets it operates in. It also includes information about recent events, competition the company faces, and its strategy for competing in its environment, which is a good place to understand how the company operates its business (Li et al., 2013). More studies are starting to develop new measures to evaluate firms' strategy and performance based on 10-K filings (e.g., Li et al., 2013; Qiu and Wang, 2018). The code of Global Industry Classification Standard (GICS), an industry taxonomy developed in 1999 by the global financial community, is employed for our sectoral analysis in this paper. The data extraction is based on the 'edgar' package, which helps in bulk data gathering and textual analysis of EDGAR filings in the R language environment.[2]

We extracted the financial data on all of the Russell 3000 firms from the WRDS CRSP Database. And we further restricted the firm sample to: 1) publicly-traded companies for which we have access to financial performance data from WRDS CRSP Database; and 2) firms with at least five extracted sentences based on our key-word basket to ensure that there is sufficient content to explore the distribution of OIPs. Because of this restriction, companies which are not active in adopting OIPs were not included in our dataset. These resulted in 6624 observations pooled over a three-year period. We also included measures of internal R&D effort, i.e., R&D intensity (RDI, dividing company R&D expenditures by its sales) among these companies when those data were available.

3.2. Measuring open innovation practices

We develop language-based measures of open innovation to capture variations among different types of practices. The measurement corpus is based on the business section in 10-K filings of the US. publicly traded companies, which discusses the business results obtained by those companies in that year. We employ topic modelling to discern the underlying open innovation activities involved in the conduct of the firm's business. Fig. 1 provides the main steps we take to construct our approach, which we will develop in the following sub-sections.

3.2.1. Key-word basket building for open innovation practices

Although the premise of NLP is to develop automated approaches to process free-text data, building those approaches requires a substantial amount of manual analysis and annotation of data (Xia and Yetisgen-Yildiz, 2012). Annotation is an important prerequisite for NLP, as annotated data serves as a resource for creating statistical models by the application of machine learning approaches (Roberts et al., 2009). Based on human annotations of natural language data, machine learning algorithms can replicate directly a human classification task, thereby reducing noise in labels and features for a real-world dataset (Passonneau et al., 2009). Therefore, to improve the quality and accuracy of algorithms, producing high-quality annotations is essential to measure OIPs in our study.

Since there has not been an annotated dataset available regarding OIPs, we created a primary keyword basket (shown as the first step in Fig. 1, and the detailed steps we employed are shown in Fig. 2). This both informed our own analysis, and also can lay a foundation for future studies by other academics to build upon our keyword basket. To construct this basket, we started with the US publicly traded companies that were selected as the world's most innovative companies from Forbes (2018) [3] as the samples to primarily test different OIPs that are implemented in practice. 52 companies were selected in this step. Then, two scholars with two research assistants annotated the sentences and the keywords in the sentences which were judged to represent OIPs. We obtained 119 phrases that we used to represent OIPs.

To finally check whether the keywords we selected can in fact represent OIPs, we stemmed the keywords and used specific algorithms to extract all sentences in the business section of our 52 sample companies that include these stemmed keywords. A total of 1764 sentences were obtained. We randomly selected 500 sentences and divided them into 10 groups. Following common NLP research practice from other fields, we recruited a panel of 20 open innovation experts to help validate the labeling for 50 sentences each (Ghanem et al., 2019; Khodak et al., 2017). These experts were selected based on following criteria: 1) their core research has been related to open innovation, 2) their papers have been cited by others in the development of the field. As much as possible we tried to balance the seniority and gender of the researchers. The experts were invited to annotate the words (0, not related to OIPs; 1, related to OIPs) by relying on the context of the sentences that we extracted from the business section of annual reports. Table 1 exhibits some examples how invited experts helped validate the phrases which we put into the word basket. For instance, one sentence extracted from the business section of a random firm *(#1* in Table 1), "to date, these alliances have taken several forms, including cooperation in the areas of product development, training, procedure development, and marketing activities", which includes a stemmed keyword (allianc) that we put in the keyword basket. Two experts and the authors of this paper then labeled this sentence at the same time. Only when the 2 experts both chose the 1 option (related to open innovation), did we consider the keywords contained in the sentence to represent OIPs. Similar annotation approaches were conduct by other studies to form labeled datasets in other research fields (Kang et al., 2020; Karoui et al., 2017; Skalicky and Crossley, 2018).

With the help of these expert annotations, we verified 115 keywords in our key-word basket regarding OIPs. We removed 4 keywords, because the agreement rates among the experts (the inter-rater

[2] More information and the codes about edgar package can be referred to https://cran.r-project.org/web/packages/edgar/index.html.

[3] https://www.forbes.com/innovative-companies/list/.

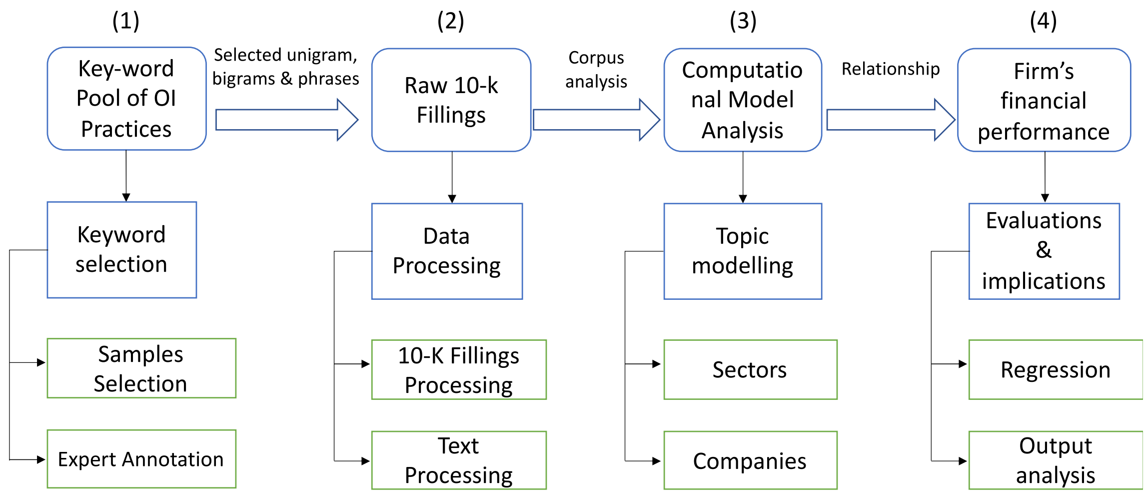

Fig. 1. Overview of measuring OIPs based on topic modelling.

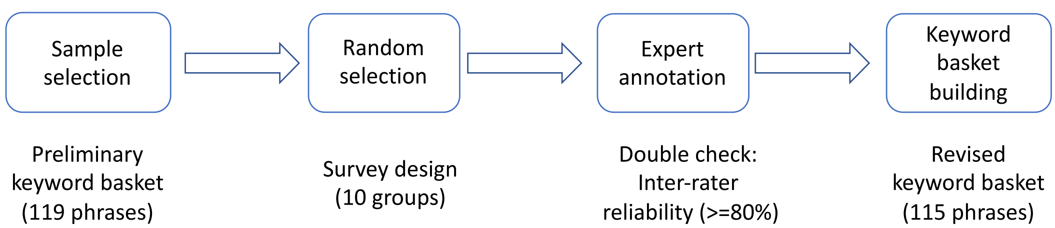

Fig. 2. Steps of keyword basket building.

Table 1
Example exhibition of expert annotation.

#	Sentences	Key Words (stemmed)	Expert 1	Expert 2	Expert 0 Authors
1	To date, these alliances have taken several forms, including cooperation in the areas of product development, training, procedure development, and marketing activities	**allianc**	1	1	1
2	The company also licenses certain of its brands to third parties	**license, third-parti**	1	1	1
3	We provide customer service through a mix of in-house call centers and outsourced third-party services.	**outsourc, third-parti**	0	1	0
4	We help our customers create and deliver the most compelling experiences in a streamlined workflow and optimize those experiences for greater return on investment.	**custom creat**	1	1	1

0: Not related to OIPs; 1, Related to OIPs.

reliability) were lower than 80% for them (Ghanem et al., 2019; Landis and Koch, 1977). We then used this revised keyword basket to extract information related to OIPs in the business section of annual reports of the Russell 3000 companies by providing selected unigram, bigrams & phrases. We share the keyword basket together with our analysis codes and our LDA results on a publicly accessible GitHub platform (Measuring_OIPs), so that other scholars can utilize and further develop the keyword basket for future research.

3.2.2. Topic modelling technique –LDA

Topic modelling is an unsupervised machine learning technique for text-mining unstructured data, which can discover latent topics in documents (Blei, 2012). It is capable of scanning a set of documents, detecting word and phrase patterns within them, and automatically clustering word groups and similar expressions that best characterize a set of documents. This method assumes that a document is composed of words, and the topics covered in more than one document can be expressed by a combination of closely related words. Any given document can be associated with more than one topic. Therefore, topic modelling is a technique that can be used to infer latent topics in a collection of text documents (Westerlund et al., 2018).

There are multiple techniques and algorithms that can be used to data-mine text documents. Among them, the Latent Dirichlet Allocation (LDA) method has gained popularity, due to its great advantage of dealing with large-scale documents and interpreting identified latent topics (Blei, 2012; Blei et al., 2003). It allows sets of observations to be explained by unobserved word groups that explain why some parts of the data are similar, and it does this without imposing any prior structure on the data. For example, if observations are words collected into documents, it posits that each document is a mixture of a small number of topics and that each word's presence is attributable to one or more of the document's topics. A handful of studies have started to take advantage of such techniques in innovation studies (e.g., Choudhury et al., 2019; Kaplan and Vakili, 2015; Westerlund et al., 2018).

To put it more technically, the LDA approach inputs a document-term matrix, for which the rows are unigram (single keywords) or bigram (two consecutive words) counts and the columns are annual report documents. LDA identifies distinct topics across the corpus by observing words that tend to co-occur frequently within each text. LDA then outputs a document-topic matrix, for which each document is assigned to a probabilistic mixture of topics, or a probability distribution giving the percentages across all topics. The number of topics to be used in the analysis is itself a choice that researchers must specify. Selecting too narrow a number of topics can lead to a lack of coverage of all the relevant information in the corpus, while choosing too large a number of topics can increase redundancy and complicate interpretation of the results. Because this is a subjective choice, we used a range of topics in our analysis. More details and original algorithms can be found in Blei et al. (2003) and Blei (2012). Fig. 3 illustrates Dirichlet-distributed topic-word distributions using matrix notation.

3.2.3. Measures for open innovation practices – research procedure

In our study, we first extracted the annual reports (2016, 2017 and 2018 fiscal years) of publicly traded companies included in the Russell 3000 Index of publicly traded stocks in the United States, and then extracted their business sections. Then we removed common stop words and punctuations, discarded word order, and stemmed words by using the Porter stemming algorithm (Cao et al., 2009). After these steps, we constructed a document-term matrix for which the rows represent distinct sentences observed across business sections that contain the word from keyword basket we created. To ensure each document contains enough text for LDA analysis, we removed the samples (companies' annual reports) that do not include more than five sentences (5% of total documents). We ended up with 7925 documents, with 381,261 sentences in total.

Our model training approach requires a key assumption: when companies describe their open innovation activities in their reports, they sometimes explicitly use some of the words in our keyword basket or a synonym and sometimes they do not. Since our research field lacks a mature labeled dataset about open innovation activities that could be used to analyze a corpus of data, we instead use the presence of our keyword basket as the labels that indicate a given phrase contains content relevant to open innovation [iv]. Training the LDA model with the extracted texts allows us to identify a set of topics related to OIPs.

We need to select the parameter for the number of topics for LDA analysis. As we aim at investigating the distribution of different OIPs, rather than maximizing the coherence or distinctiveness of the topics, we do not expect the topic numbers to influence significantly our results. We output the topics with different numbers (i.e., 25, 50, and 100 topics), and the results are highly consistent with those we report below. By inspection, 100 topics appeared to include a number of redundant topics, while 25 topics did not appear to cover all of the information in the corpus of data. Thus, we took the number of 50 topics each year as the main output for analysis we report.

Our main point is to derive different OIPs regarding the coordination and knowledge sharing in inter-organizational relationships (Brunswicker and Chesbrough, 2018), rather than investigating the types of partners with which firms conduct open innovation activities together (e.g., Du et al., 2014). Thus, to keep consistent with prior literature, we mainly refer to the open innovation classifications in limited existing studies to define clustered topics from LDA analysis. Mazzola et al. (2012), for instance, summarized twelve common OIPs that firm usually adopted, covering the activities from collaboration, licensing, to knowledge commercialization. Ahn et al. (2015) proposed an open innovation taxonomy with seven types of open innovation activities and three open innovation modes based on the dominant changes involved, from user involvement to spin-off. Brunswicker and Chesbrough (2018) distinguished OIPs into four major modes using two dimensions for classification, bilateral versus multi-actor and transactional versus collaborative, with seven individual OIPs.

Based on the defined practices from these studies, we primarily propose eight individual OIPs by taking verified open innovation keyword baskets into account as well: network & community; customer engagement; crowdsourcing; open innovation intermediaries; partnership & joint venture activities; industry-academia collaboration; contract & IP licensing; bilateral transactional activities. The final OIP classification in this study will be defined by the topics generated from LDA analysis.

3.2.4. Measuring firms' financial performance

We chose financial performance as our dependent variable rather than innovation performance with following reasons. First of all, while prior studies featured a variety of constructs used to measure innovation performance, (e.g., Brunswicker and Vanhaverbeke (2015); Chen et al. (2016); Grimpe and Sofka (2016); see more studies in Table A1), they remain relatively silent on how individual OIPs will influence the financial performance of the firm. We regard innovation performance as an intermediate measure in the pursuit of overall improved firm performance (i.e., not "innovation for its own sake") (Chesbrough, 2019)., Innovative activities in firms usually are intended to stimulate growth (Artz et al., 2010), and investors tend to reward companies that demonstrate creative business strategies and growth with higher valuations (Kogan et al., 2017). Firms can achieve a more reasonable and efficient allocation of resources by adopting different types of OIPs, which is beneficial for long-term financial performance. Recent publications have argued that open innovation needs to be managed to achieve positive business results (Chesbrough, 2019).

We measure firm financial performance by using Tobin's Q– i.e., the ratio of the market value of the firm to the book value of its assets as our dependent variable. Tobin's Q has been used to explain a wide variety of phenomena since it was first introduced, especially as a useful measure of firm financial performance (Bharadwaj et al., 1999; Lang and Stulz, 1994). Applying it to performance has been proved robust in various settings including diversification strategy and innovation (Hung and Chou, 2013). Regarded as a forward-looking measure, Tobin's Q is a good proxy for a firm's competitive advantage and its potential long-term profitability (Chung and Pruitt, 1994), which has been suggested as an appropriate indicator for open innovation related research, as the gains of open innovation will unfold over time in ways that may not always be directly observable (Hung and Chou, 2013; Lin et al., 2006; Sisodiya et al., 2013). Therefore, compared to other commonly used financial performance indicators, such as return on assets (ROA), or return on investment (ROI), Tobin's Q has a smaller average error and higher average correlations (Anderson et al., 2004), and its characteristics can represent open innovation better. When being used as a dependent variable, the potential measurement error of Tobin's Q is less of a concern (Lu and Beamish, 2004). Furthermore, other recent research has linked Tobin's Q to a firm's creativity and innovation, as judged by the market (Corritore et al., 2020). Following other studies to

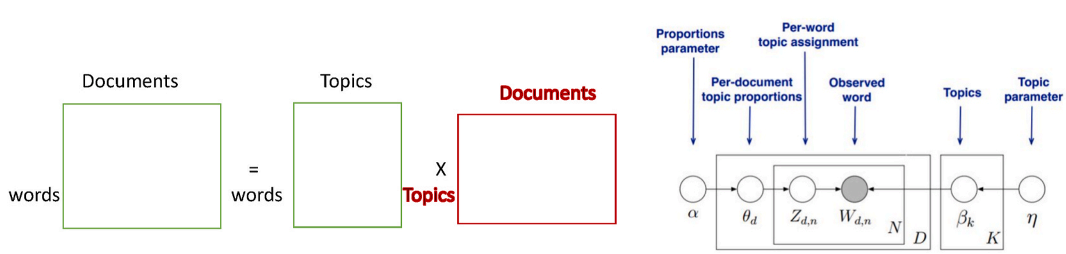

Fig. 3. Matrix notation for LDA with Dirichlet-distributed topic-word distributions.

reduce the impact of outliers on the results (Alti, 2006; Deb et al., 2017), we drop observations where the value of Tobin's Q exceeds 10. We also put the analysis of using other financial indicators (i.e., profitability, revenue growth and ROA) which are commonly used in prior studies as dependent variables in the supplementary material on GitHub.

3.2.5. Control variables

There are many factors that influence a firm's financial performance, so it is critical to identify other control variables that influence this relationship, in order to reduce the risk of omitted variable bias in estimating this relationship. Several variables have been identified in the empirical literature as influencing a firm's financial performance as control variables. First, we control for firm size, measured as the number of employees (expressed in log form), as firm sizes are commonly associated with growth prospects (Josefy et al., 2015). Further, we control firm prior performance for firm long-term financial performance, measured by the firm's return on assets (ROA) with a one-year lagged variable (Li and Tang, 2010). Another measure of a firm's investment intangible assets is capital intensity, measured by a firm's capital expenditures (net plant, property and equipment) divided by total assets, as higher expenditures are a natural trigger of firm performance (Li and Tang, 2010). We also control the year fixed effects, as well as including a set of 11 dummy variables for sectors (Chen et al., 2016).

To measure the importance of internal R&D for OIP adoption, we use firm-level R&D intensity (RDI, R&D expenditure divided by sales) (Cohen and Levinthal, 1990). We replace missing values of R&D expenditure with 0 and cap the upper limit at 1, and then include a dummy variable for missing R&D in the analysis based on prior studies (Blagoeva et al., 2020; Deb et al., 2017). All the financial data is taken from the WRDS COMPUSTAT Database.

3.2.6. Potential method limitations

A key common issue with topic models is the reliability of topics, as it is based on unsupervised learning approaches to cluster topics. So we employed some pretreatment steps (following earlier LDA analyses in other social science fields) to improve the reliability of our topics. First, we manually read the annual reports to select relevant keywords by choosing the most innovative companies, which can help us observe more OIPs in real practices. Second, we relied on the language that companies are using to describe their real practices, rather than academic papers, to construct the keyword basket. Furthermore, we adopted an expert annotation approach to double check the initial keyword basket, removing the words whose inter-rater reliability is lower than 80%. By setting up these pretreatment steps before LDA analysis, we can reduce the noise from the original corpus and improve the reliability of clustered topics related to OIPs. We have to acknowledge, however, that our resulting keyword basket is likely to be incomplete and may not be comprehensive in its identification of OIPs. We do see it as an important initial step to build up a foundation for future research, which might enhance and extend the keyword basket we report here.

4. Results

4.1. Selected topics for open innovation practices

Even though we proposed eight individual OIPs based on the defined practices from prior studies (see 3.2.3 section), the final OIPs were derived from our LDA results. Although the presence of keywords represents the probability of OIP adoption by firms in our setting, we still followed the overall meaning of each topic closely. Therefore, only the topics including not less than two key-words were treated as the topics related to open innovation (each topic contains 10 words in our model setting), and other topics are regarded as unrelated topics to open innovation (i.e., the probability of adopting OIP = 0). We sum the probability of topics if they are relevant to any type of these OIPs. All topics generated from LDA analysis are shown on our keyword repository on GitHub. From the analysis, we characterized six main OIPs based on the word distributions among topics. These OIPs are: 1) Network & community; 2) Customer engagement; 3) Partnership & joint venture activities; 4) Industry-academia collaboration; 5) Contracts & IP licensing; and 6) Bilateral transactional activities. Two OIPs were not found from our LDA analysis. These were the practices of crowdsourcing and of open innovation intermediaries. While managers did respond to survey prompts for these two practices in Brunswicker and Chesbrough (2015), our LDA topic modelling did not discern evidence for either practice from our corpus of Russell 3000 companies. We return to this finding in the Discussion section below.

Table 2 shows the selected topics and their associated words, where bold words are from the open innovation keyword baskets. We find that certain keywords appear together under several topics, and the simple labels we chose generally capture the underlying meanings of the topics regarding different OIPs. For instance, the words in Topic 5 are related to patent licensing, which we manually label this topic as contracts & IP licensing activities. We visually represent an example in Fig. 4, to show how words are associated with different topics, represented by different shades of colors, and also show their topic probability distribution (with 4 topics in our chosen example). Specially for the OIP of bilateral transactional activities, even though some clustered topics are related to it, this type of OIP lacks a sufficient number of keywords to support its distribution among companies based on the keyword basket (see the limitations on Section 5.3). We remove it in the following analysis.

4.2. The distribution of open innovation practices across industries

Table 3 presents the descriptive statistics and simple correlations among our variables. We took the probability of all OIP adoptions (the sum of the probabilities of OIPs, OIP_sum) as the proxy measure for each firm's openness level. The correlation table indicates there are no strong correlations between the independent variables, and we also examined variance inflation factor (VIF) values for the variables, and all the VIF values are below the thresholds suggested by references (Hair et al., 2011), both of which indicate that multi-collinearity is not a problem for this study.

Consistent with the findings proposed by prior studies, the openness level (OIP_sum) has a significant positive association with Tobin's Q, but not all individual OIPs are significantly positively associated with Tobin's Q. And the overall openness level is significantly positively associated with internal R&D intensity. Meanwhile, it is interesting to note even in the descriptive statistics we find a result that we did not expect: the five individual OIPs do not all correlate positively to one another, but on the contrary, some of them are negatively correlated to each other. The correlation table shows a relatively strong positive, statistically significant correlation between OIP1 (Network & communities) and OIP2 (Customer engagement), and also between OIP4 (Industry-academia collaboration) and OIP5 (Contracts & IP licensing). Other OIPs overall show negative correlations to each other, especially OIP3 (Partnership & joint venture activities) to the other OIPs.

Table 2
Selected topics and their associated keywords from LDA analysis.

Topics	Key words
1. Network & community	Data, advertis, **campaign**, measur, **platform**, buyer, collect, technolog, marketplace, **third-parti**
2. Customer engagement	**Custom**, commun, **engag**, data, provid, market, **platform**, collect, busi, inform
3. Partnership & joint venture activities	**Properti**, **partnership**, oper, interest, real, **joint_ventur**, estat, **partner**, manag, million
4. Industry-academia collaboration	Program, **institute**, **educ**, student, **author**, school, univers, titl, require, **educ_program**
5. Contracts & IP licensing	**Licens**, **agreement**, **patent**, product, develop, commerci, certain, **collabor**, **grant**, exclus
6. Bilateral transactional activities	**Franchise**, restaur, oper, develop, **agreement**, **franchis**, sale, market, local, **licens**

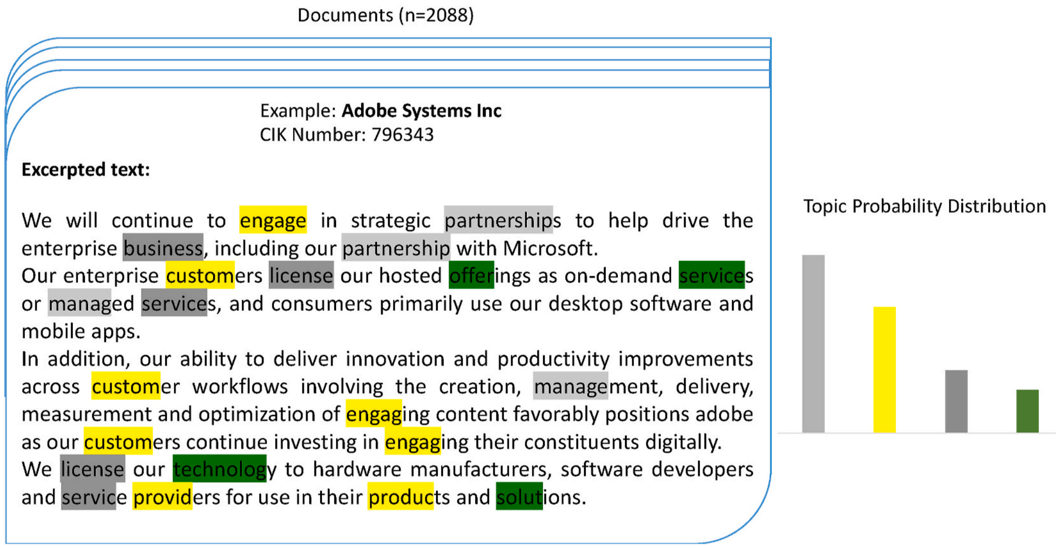

Fig. 4. Example of LDA's topic assignment
(Note: Excerpted text can be viewed as a mixture of various topics, where their probabilities are assigned via LDA. Each topic contains different words, and the words highlighted with different colors belong to different topics.).. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

We can also observe that the adoption of OIPs varies among firms from the descriptive statistics in Table 3. Some OIPs are adopted by specific firms while they are not adopted by other firms (e.g., the max values of all OIPs are all larger than 0.9 and the min values are 0). Therefore, we further display the average probabilities for different OIPs by industry sector in Table 4. It is clear even from these descriptive statistics that OIPs are not evenly distributed across different sectors of the economy. Overall, OIP5 (Contract & IP licensing) and OIP3 (Partnership & joint venture activities) are adopted by firms for their business operations more frequently compared to other OIPs. Different sectors tend to adopt different OIPs. The sectors of Information Technology and Communication Services adopted more practices related to Network & community (OIP1) and Customer engagement (OIP2), while Real Estate and Energy sectors took more partnership & joint venture activities (OIP3) compared to others. The sectors of Health Care, Consumer Discretionary and Consumer Staples adopted more Contracts & IP licensing practices (OIP5). Overall, the Utilities and Financials sectors launched fewer open innovation activities than other sectors by considering the total sum of probabilities of OIPs (OIP_sum). This pattern of differences across sectors provides new insights in the open innovation performance literature.

OIP1: Network & communities; OIP2: Customer engagement; OIP3: Partnership & joint venture activities; OIP4: Industry-academia collaboration; OIP5: Contracts & IP licensing. OIP_sum: the sum of the probabilities of OIPs, here we take it as a measure of the openness level of a firm.

4.3. The relationship between open innovation and financial performance

Table 5 shows OLS linear regression results for models of firm financial performance (with a one-year lag) on the adoption of OIPs as well as the effects of R&D efforts on their relationship. Whereas Model 1 shows that the overall openness level (the sum of probabilities of all OIPs, named as OIP_sum in the table) has a significant positive association with Tobin's Q, Model 2 reports that not all OIPs exhibit the same positive relationship. The practice regarding OIP2 (Customer engagement) and OIP4 (Industry-academia collaboration) exhibit a particularly significant and positive association with Tobin's Q, compared to other types of open innovation practices. Item OIP5 (Contracts & IP licensing) and OIP1 (Network & communities) also show a positive and weak significant association with Tobin's Q, while OIP3 (Partnership & joint venture activities) is significantly negatively associated with financial performance.

In these models, we can also see the contribution of internal R&D effort to Tobin's Q, along with individual OIPs and aggregate openness measure. Internal R&D intensity (shown as RDI in the table) remains significantly and positively associated with Tobin's Q. To further investigate the role of in-house R&D capabilities in amplifying the effect of OIPs on firms' financial performances, we use a set of interaction terms between internal R&D intensity and the overall openness level as well as different OIPs in Model 3 & Model 4. The results about the relationships between open innovation with both aggregate and nuanced levels and financial performance remain consistent with Model 1 and Model 2. In Model 3, the overall openness level is still positively and significantly associated with Tobin's Q, while the interaction term (RDI*OIP_sum) has a negative and significant effect. Model 4 shows that the sign of the effect varies across the OIPs. R&D intensity positively moderates the relationship between OIP2 (Customer engagement) as well as OIP3 (Partnership & joint-venture activities) and firm financial performance, whereas R&D intensity negatively moderates the relationship between OIP1 (Network & community) as well as OIP5 (Contract & IP licensing) and financial performance. The interaction of OIP4 (Industry-academia collaboration) with R&D intensity is not significantly related.

Inspired by prior studies which conducted empirical investigations on the curvilinear relationship between open innovation and firm financial and innovation performance (e.g., Belderbos et al., 2010; Laursen and Salter, 2006, 2014), we further test curvilinear effects of the overall openness level and individual OIPs on firm financial performance in Table 6 [v]. Among these models, the coefficient of OIP adoption sum (OIP_sum) is significant and positive to financial performance (Tobin' Q), and the coefficients of most of the individual OIPs are significant and positive to financial performance except OIP3, all of which keep consistent with Table 5's observation.

The parameters for the overall openness level (OIP_sum) and almost all individual OIPs except OIP3 are significant and positive, whereas the parameters for their squares are significant and negative. To investigate a curvilinear relationship (an inverted U-shaped relationship for the paradox of openness in our study), a significant and positive coefficient for the independent variable and a significant and negative coefficient

Table 3
Descriptive statistics and bivariate correlations.

Variables	Mean	S.D.	Min	Max	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) **Tobin's Q**	1.320	1.506	0	10.00	1.000										
(2) **Firm size**	0.213	1.076	−3	3.362	−0.011	1.000									
(3) **R&D intensity**	0.093	0.238	−0.013	1	0.366*	−0.327*	1.000								
(4) **ROA**	-.0031	.0389	−13.17	4.023	-0.066*	0.259*	−0.487*	1.000							
(5) **Cap intensity**	0.608	2.371	0	78.225	0.010	−0.114*	0.068*	−0.073*	1.000						
(6) **OIP1**	0.062	0.133	0	0.989	0.053*	0.155*	−0.077*	0.045*	−0.065*	1.000					
(7) **OIP2**	0.088	0.159	0	0.997	0.149*	0.150*	−0.046*	0.049*	−0.085*	0.089*	1.000				
(8) **OIP3**	0.098	0.176	0	0.999	−0.075*	−0.180*	−0.038*	0.005	0.050*	−0.094*	−0.110*	1.000			
(9) **OIP4**	0.015	0.073	0	0.966	0.111*	−0.060*	0.205*	−0.106*	0.023	−0.026*	−0.042*	−0.070*	1.000		
(10) **OIP5**	0.149	0.245	0	1	0.251*	−0.209*	0.588*	−0.282*	0.057*	−0.152*	−0.167*	−0.147*	0.096*	1.000	
(11) **OIP_sum**	0.411	0.311	0	1	0.279*	−0.138*	0.432*	−0.200*	0.008	0.292*	0.344*	0.338*	0.239*	0.575*	1.000

Note: Correlations equal to or above |0.04| are significant at $p < 0.05$. Two-tailed tests.

for its square are necessary. Based on the three-step procedure proposed by Lind and Mehlum (2010) to establish a quadratic relationship, other two conditions should be met as well: the slope must be sufficiently steep at both ends of the data range and the turning points should be located well within the data range (Haans et al., 2015). We follow the above three steps to formally test the quadratic relationships between open innovation and financial performance (See Table A2 in the appendix).

We find that the firm's overall openness level is positively associated with firm financial performance (Model 5), rather than showing an inverted U-shape, given its turning point out of its data range. The same situation exists in OIP2 (Customer engagement), which means more customer engagement for open innovation activities is positively associated with improved financial performance (Model 7). For OIP1 (Network & communities), OIP4 (Industry-academia collaboration), and OIP5(Contract & IP licensing), we find strong support for asserting that these individual OIPs are curvilinear – showing an inverted U-shape – associated with firm financial performance (Model 6, 9 and 10). These models partially support the openness paradox, –when firms adopt these specific OIPs too extensively in their business operations, they will decrease financial gains.

4.4. Robustness checks

We conducted several additional analyses to assess the robustness of our findings. First, we varied the number of topics (i.e., 25, 50, 100) to ensure using companies' annual report to extract information can be a good proxy for measuring OIP adoption. And we also pooled three years of data together into one corpus to enable the LDA algorithm to expand its analysis to 150 topics. The results remain robust and consistent with our findings. Second, we have done further regression analysis for firms with the probability of OIP distribution = 0 and the probability OIP >0 also for robustness checks. Our results remain quite consistent, implying that this is not biasing the estimates we report in our paper. Third, we undertook a manual coding approach to identify the OIPs from our samples (10 random cases of 10 each), and then compared them to the OIPs from LDA results. For instance, we manually read the business section of Adobe Systems Inc and extracted the OIPs that Adobe adopted, comparing these OIPs we identified manually to LDA analysis. Our comparison shows that algorithmically-derived measures and outcomes were nearly identical to the results based on our manual analysis.[4] Fourth, we include several other control variables in our analysis, firm age, firm growth (measured as the natural logarithm of sales in *t* year divided by sales in *t-1* year). Including these control variables restricts the size of our samples, but the results obtained were consistent with the results reported here. Finally, we use Tobit models, which are commonly used in the open innovation research stream (e.g., Laursen and Salter, 2006; Du et al., 2014; Greco et al., 2016) for single or double censored continuous dependent variables (i.e., Tobin's Q as a financial performance indicator in our study), and we found the results are robust and consistent.

5. Discussion

Our methods allow us to build upon the prior literature showing mixed results regarding the performance impact of open innovation (Cheng and Huizingh, 2014). Our aggregate measure of openness, built from the sum of OIPs, shows that firms benefit from the overall openness level, which is consistent with prior research (e.g., Du et al., 2014; Kafouros and Forsans, 2012; Mazzola et al., 2012). Due to our large corpus and our construction of a keyword basket, we can make more fine-grained measures of individual OIPs, without incurring the limitations noted above in self-reported survey data or archival data. This

[4] We thank an anonymous reviewer for the suggestion to compare a manual coding of the keywords to the topic modelling result for selected companies.

Table 4
Probability distributions of different OIPs and industry average.

GIC sector	Obs	OIP1	OIP2	OIP3	OIP4	OIP5	OIP
10 Energy	276	0.033	0.034	0.214	0.006	0.087	0.374
15 Materials	280	0.049	0.029	0.09	0.006	0.102	0.276
20 Industrials	938	0.109	0.11	0.075	0.006	0.076	0.376
25 Consumer Discretionary	833	0.061	0.108	0.074	0.038	0.16	0.441
30 Consumer Staples	240	0.073	0.054	0.063	0.005	0.198	0.393
35 Health Care	987	0.028	0.038	0.058	0.046	0.492	0.662
40 Financials	1276	0.024	0.042	0.036	0.004	0.083	0.189
45 Information Technology	901	0.127	0.228	0.074	0.003	0.062	0.494
50 Communication services	259	0.104	0.173	0.046	0.003	0.045	0.371
55 Utilities	180	0.014	0.016	0.052	0.004	0.041	0.127
60 Real Estate	415	0.029	0.029	0.533	0.004	0.025	0.62
All Sectors	6585	0.062	0.088	0.098	0.015	0.149	0.412

Table 5
OLS regression: the relationship between firm financial performance and OIP adoptions and the effects of R&D efforts.

	Model 1	Model 2	Model 3	Model 4
Dependent variable	Tobin's Q	Tobin's Q	Tobin's Q	Tobin's Q
OIP1		0.156*		0.287**
		(0.133)		(0.144)
OIP2		1.146***		0.642***
		(0.115)		(0.135)
OIP3		−0.258**		−0.276**
		(0.124)		(0.133)
OIP4		0.712***		0.496*
		(0.231)		(0.263)
OIP5		0.141*		0.332***
		(0.091)		(0.097)
OIP_sum	0.340***		0.424***	
	(0.063)		(0.066)	
RDI*OIP_sum			−1.553***	
			(0.361)	
RDI*OIP1				−2.099**
				(1.052)
RDI*OIP2				5.533***
				(0.927)
RDI*OIP3				−0.867*
				(0.504)
RDI*OIP4				0.320
				(0.657)
RDI*OIP5				−1.451***
				(0.370)
Firm size (log)	−0.031*	−0.043**	−0.029	−0.036**
	(0.018)	(0.018)	(0.018)	(0.018)
ROA	0.681***	0.687***	0.685***	0.679***
	(0.059)	(0.058)	(0.059)	(0.058)
Capital intensity	0.017**	0.019***	0.018**	0.019***
	(0.007)	(0.007)	(0.007)	(0.007)
RDI	1.378***	1.472***	2.679***	2.316***
	(0.109)	(0.112)	(0.322)	(0.323)
Missing R&D value	0.243***	0.243***	0.223***	0.174***
	(0.041)	(0.041)	(0.042)	(0.042)
Constant	0.571***	0.699***	0.539***	0.685***
	(0.088)	(0.089)	(0.088)	(0.089)
Sector-fixed effects	Yes	Yes	Yes	Yes
Year-fixed effect	Yes	Yes	Yes	Yes
No. of Obs.	6541	6541	6541	6541
R-squared	0.275	0.285	0.277	0.295

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

allows us to go further into the paradox of openness, by unpacking the role of specific OIPs, and whether and how those contribute to firm financial performance. Firms can benefit from collaborations with several types of external knowledge sources (Chen et al., 2016), but not all OIPs are equally important to improve firm financial performance. For instance, we find the practice related to customer engagement has a stronger positive association with improved financial performance, while the practice related to market-based partnership & joint venture shows a negative effect (Du et al., 2014). Indeed, some practices are even negatively associated with other practices. So this provides one explanation for earlier mixed findings: different practices have different effects on financial performance.

At the same time, we also find evidence for the curvilinearity of the impact of open innovation – the inverted U-shape effect not only exists in the relationship between open innovation and firm innovation performance (e.g, Laursen and Salter, 2006, 2014), but also exists in the relationship between some (but not all) specific OIPs and firm financial performance. Different OIPs show various patterns, with some being associated with improved financial performance while others are not. The findings partly support the existence of paradox of openness for some OIPs, as some innovation activities usually require openness, but gaining business benefits require protection (Laursen and Salter, 2014). Similar to many relationships in strategic management – too much can be as bad as too little (Haans et al., 2015), the adoption of some specific OIPs also follows this pattern, such as building network and communities, or conducting IP licensing contracts (Laursen and Salter, 2014).

Our results also suggest that a richer, more granular set of measures for how open innovation is practiced within firms yields further important findings. While most OIPs exhibit a positive association with firm financial performance, some OIPs display a positive association only in specific sectors. The overall openness level and all individual OIPs can only play a positive role in the growth of financial performance to a certain extent (Laursen and Salter, 2014). And the relationship between individual practices is nuanced, with some of them exhibiting a complementary relationship to one another, and others showing a substitutive relationship instead. Still other OIPs involving crowdsourcing, intermediaries and bilateral transactional activities were not discernible from our examination of the corpus of Russell 3000 firms.

We similarly find nuanced results for the role of internal R&D effort on firm performance: a positive moderation effect of internal R&D on the relationship between open innovation adoption and firm financial performance suggested by prior research overall is not supported. By contrast, we observe a negative moderation effect. The usual story of in-house R&D capability supporting external research (e.g., Chen et al., 2016) appears to be more complex than a simple positive relationship between internal R&D and external innovation. Individual OIPs are moderated by R&D intensity, but the effect is sometimes positive and other times negative, and also varies by the sector of the economy, making an aggregate measure indeterminate. Thus, there is not a clear-cut answer to the question of whether internal R&D and external searching are complementary or substitutive innovation activities (Hagedoorn and Wang, 2012). This suggests that the complementarity between internal R&D effort and external innovation activities (Cassiman and Veugelers, 2006; Hung and Chou, 2013), may vary with specific OIPs and particular economic sectors. Our results indicate that, if we examine open innovation's impact on firm performance at too aggregate a level, we inadvertently might conflate practices that do influence performance with practices that do not. The interactions between individual OIPs and R&D intensity help to show how nuanced the effect of R&D intensity is with open innovation, and that the effect varies

Table 6
OLS regression: exploring possible curvilinear effects of open innovation on financial performance.

	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10
Dependent variable	Tobin's Q	Tobin's Q	Tobin's Q	Tobin's Q	Tobin's Q	Tobin's Q
OIP1		0.995***				
		(0.332)				
OIP1ˆ2		−1.592***				
		(0.568)				
OIP2			1.471***			
			(0.282)			
OIP2ˆ2			−0.546			
			(0.437)			
OIP3				−0.223		
				(0.264)		
OIP3ˆ2				−0.324		
				(0.370)		
OIP4					2.489***	
					(0.613)	
OIP4ˆ2					−2.576***	
					(0.816)	
OIP5						1.065***
						(0.236)
OIP5ˆ2						−1.391***
						(0.294)
OIP_sum	0.608***					
	(0.212)					
OIP_sumˆ2	−0.290*					
	(0.220)					
Firm size (log)	−0.031*	−0.038**	−0.043**	−0.038**	−0.036**	−0.036*
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
ROA	0.682***	0.689***	0.684***	0.690***	0.690***	0.695***
	(0.059)	(0.059)	(0.058)	(0.059)	(0.059)	(0.059)
Capital intensity	0.017**	0.017**	0.019***	0.017**	0.016**	0.017**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
RDI	1.409***	1.552***	1.557***	1.550***	1.443***	1.601***
	(0.112)	(0.106)	(0.105)	(0.106)	(0.108)	(0.113)
Missing R&D value	0.241***	0.227***	0.227***	0.224***	0.236***	0.217***
	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
Constant	0.529***	0.667***	0.679***	0.777***	0.684***	0.660***
	(0.093)	(0.086)	(0.084)	(0.089)	(0.084)	(0.085)
Sector-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year-fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
No. of Obs.	6541	6541	6541	6541	6541	6541
R-squared	0.275	0.273	0.283	0.273	0.274	0.274

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

with the OIPs being studied, and the sector in which it is studied.

Our research design does not allow us to demonstrate causality in our findings as topic models are unsupervised learning approaches, but our findings are quite consistent with other evidence reported at the project level on open innovation and firm performance where causality is more able to be established. For instance, Du et al. (2014) examined 489 projects of a European manufacturer, and reported that scientifically-based collaborations with universities resulted in different performance outcomes when compared to market-based collaborations. They reported that different kinds of collaborations used different practices, leading to different results. The "open innovation paradox" (Laursen and Salter, 2014) also likely works differently in different contexts. A scientifically-based collaboration with a university, for example, might involve a paradox of whether and when to disclose important research outcomes from technical experiments. Here, IP protection may be of critical importance. A market-based collaboration (Oltra et al., 2018), by contrast, likely focuses on whether and when to disclose information on pricing, availability and servicing, instead of technical experimental data. Therefore, IP may be far less relevant in that context. Our results also showed that this set of practices (Contract & IP licensing OIP) has a smaller association with better financial performance, compared to our other observed OIPs. Different OIPs would therefore yield different results. Moreover, these results might vary by the sector of the economy.

5.1. The salience of open innovation practices varies by sector

Another important finding from our empirical analysis is that the specific OIPs vary in their salience by sector. Contracts and IP licensing (OIP5), for instance, is commonly used by most of the sectors, whereas Industry-academia collaboration (OIP4) is found extensively in some sectors (e.g., Heath Care sector), and not at all in other sectors. Technologically intensive sectors like Health Care, for example, tend to conduct research collaborations with universities (OIP4). Sectors that place more emphasis on market knowledge flows than technological knowledge flows, such as the Real Estate sector, on the other hand, seem rather to adopt more partnership & joint venture activities (OIP3). Prior research also showed that different external knowledge search patterns exist in different industries and sectors (e.g., Grimpe and Sofka, 2009).

One explanation for this variance is that different sectors might be in different phases of their industry life cycle (Miles et al., 1993). Sectors in the growth stage of their life cycle (e.g., the Information Technology sector), might experience a high rate of new firm entry, creating additional between-firm heterogeneity (Madsen and Walker, 2002). These differences ultimately lead to substantial variance in profitability (Knott, 2003) across competitors, and thus it is not surprising to see the sectors in the growth stage use more multi-collaborative OIPs (i.e. network & community; customer engagement) with more interactive and dynamic characteristics than other sectors in later stages of the life cycle (Chesbrough, 2011; Radziwon and Bogers, 2019). Sectors in the maturity and decline stages of the life cycle (e.g., Real Estate sector), might experience changes that are less radical and more incremental with more standardized established industry norms (Karniouchina et al., 2013). Partnerships and joint venture activities are commonly adopted to further consolidate their positions as growth diminishes and the industry matures. Therefore, we argue that industry life cycle conditions may be important reasons for these sectoral differences.

Given that the definition of open innovation involves knowledge flows across organizational boundaries (Chesbrough and Bogers, 2014), it is likely that the speed of those knowledge flows and the rate of knowledge iteration (Eppinger, 2001; Escribano et al., 2009; Schulz, 2001) also influence the salience of individual OIPs. For instance, building strong network & communities (OIP1) and putting customers directly into the innovation process (OIP2) requires more complex and dynamic knowledge flows than other OIPs, because they require close and continuing collaborations (Chesbrough, 2011; Maslowska et al., 2016; Sawhney et al., 2005). In our results, we observe that these two types of OIPs are adopted more by specific sectors such as Information Technology and Communication Service, where the speed of knowledge flows across firm boundaries occurs rapidly in these sectors. There is an important implication from these variations across economic sectors: different OIPs work differently in different parts of the economy. Firms have to make the most appropriate choices from among various options to suit their particular situation. This means that there is likely no single set of Best Practices for open innovation that will be effective across all of the economy.

Also of interest is what we did not find in our analysis. As noted above, two OIPs, crowdsourcing and use of intermediaries, did not emerge from the NLP analysis of the corpus of annual reports (10-Ks) from the Russell 3000 companies. These non-findings are worth discussing, because both practices are actively studied in the academic literature on open innovation (e.g., Katzy et al., 2013; Lee et al., 2010; Liu et al., 2020; Terwiesch and Xu, 2008). A recent search on Google Scholar, for example, returned more than 53,000 citations for "open innovation crowdsourcing". Another search from the same source for "open innovation intermediaries" returned even more citations, 270, 000. This lack of salience in the corpus of the Russell 3000 companies' annual reports (10-Ks) suggests that academics might reflect on where they are allocating their research attention. Our non-finding is consistent with other survey results of the adoption of open innovation in large companies (Brunswicker and Chesbrough, 2015, 2018). Those surveys reported that the practices of crowdsourcing and intermediaries were the least important, least used practices of large firms that had adopted open innovation. We have to note that one potential reason is that many firms do not necessarily consider crowdsourcing as open innovation but rather a means for marketing, or perhaps they describe the practices related to crowdsourcing and intermediaries in different ways when reporting financial results. [vi]

Because these NLP methods are new to the study of open innovation, we must be cautious in the interpretation of our non-findings. With a more mature keyword basket that has been tested and validated by other scholars, for example, we might see evidence emerge for these practices. Similarly, if scholars apply our methods to other corpora of text, evidence again might also emerge. We would observe, though, that the extensive use of crowdsourcing did not save Quirky from going bankrupt in 2015. And the large Solver community built by the intermediary InnoCentive nonetheless resulted in the acquisition of InnoCentive at less than 5% of its invested capital in 2019 (Chesbrough, 2019). Such negative outcomes have not been much discussed in the academic literature, yet are consistent with both the survey results of open innovation adoption in large firms (Brunswicker and Chesbrough, 2015, 2018), and our own NLP analysis reported here. Therefore, more research is clearly needed to resolve the contrast between the robust academic study of these practices, and the lack of empirical support to date for their effects on financial performance.

5.2. How can NLP techniques advance the study of innovation and strategy?

In this paper, we employ NLP and machine learning techniques, which are now actively deployed in other social sciences, and recently have arrived in the field of innovation studies (see Arts et al. (2020) and He et al. (2020) for some of the first analyses as well as Antons et al. (2020) for an overview). We believe that our field should welcome these new methods. These approaches do not require a priori specification of particular relationships, but instead can detect complex patterns in unstructured data. As we have already noted, these methods also avoid some of the biases inherent in collecting innovation data with survey methods, and can extend to more practices than most archival sources allow. While there is a valid concern about the ability to discern causal relationships from these methods, new research is developing that employs NLP methods to support traditional hypothesis testing based upon inspection of data patterns that emerge from an initial analysis (Shrestha et al., 2020; Tidhar and Eisenhardt, 2020). These new methods arrive at a time when our prior research on open innovation and performance have employed measures that have many deficiencies. Not least among these deficiencies are self-reporting bias, common method bias, and (depending on the method) an inability for other scholars to observe and replicate the findings. The NLP methods used in our study yield that results are consistent with earlier research on open innovation, making our findings more plausible. To be sure, NLP methods have their own limitations, as we will discuss below.

It is also worth remembering the source of the textual data that forms the corpus for our analysis. Annual reports (10-Ks) or quarterly reports (10-Qs) are public documents, and top managers bear personal responsibility for their accuracy and completeness. The SEC regulates the disclosure of public financial information, and there is a deep infrastructure of accounting and auditing processes that enforce a certain amount of consistency in reporting. External stock market analysts further enforce certain reporting norms and expectations. These documents are also painstakingly constructed every quarter and every year. They form the base of all communications between publicly held companies and their investors. These reports are critical documents for companies to detail their financial performance to external investors.

Our analysis reveals that a basket of these words contains many keywords and sentences that allow researchers to compile specific OIPs from the words in these documents. These new methods allow us to construct more fine-grained measures of OIPs that can be observed and replicated by others. We offer our keyword basket for inspection to other scholars, with the expectation that these keywords will be refined and improved with additional research efforts. Hopefully, those improvements also will be made publicly available to scholars. In so doing, we may advance our understanding of how open innovation is practiced, and which practices are associated with better business results. NLP techniques can also be applied to other secondary datasets which have been used to extract specific open innovation activities, such as strategic technology agreements (MERIT-CATI dataset), SDC and industrial partnerships (Schilling, 2009).

We view our results as promising, initial findings. Much more can be done with these methods. In classification tasks, labeled data is a prerequisite for analysis. However, there exist very few mature labeled datasets in management (Kang et al., 2020) and none to our knowledge in open innovation, that would allow the use of labeled data. To improve the quality and accuracy of algorithms, manually coding data is necessary as the first step, which is time-consuming and subject to coding errors. Fortunately, we can combine manual coding and training algorithms together to ensure quality and efficiency (Song et al., 2019). Our study not only introduces these methods to the open innovation literature, it also offers an initial key-word basket for open innovation practices by using companies' language in their financial reporting. Creating, sharing, and augmenting the keyword basket, and deploying these baskets across other business documents, will doubtless provide additional findings about open innovation and financial performance. Once we have robust classification data, we can begin to employ supervised learning methods, in addition to the topic modelling methods shown here.

5.3. Limitations and future directions

Our study and these new methods have some important limitations that must be kept in mind. While we are studying the relationship between open innovation and financial performance, we mainly focus on the initial steps of carefully introducing the novel text-based approach as well as creating a key-word basket foundation for future research. Our analysis lacks the identification needed to demonstrate causality for our claims. Another limitation is that many companies do not provide detailed discussion of their innovation practices and performance in their annual reports (10-Ks). Indeed, many do not even report R&D expenditures. Other corpus of text might provide greater evidence of the value of crowdsourcing, open innovation intermediaries and bilateral transactional activities, for example. Our corpus has its own limitation, as firms will mainly talk about the important practices in their business sections, and they may remove other activities such as crowdsourcing in their reports. Other textual sources of data would be helpful for scholars to further investigate companies' innovation practices.

The modelling strategy we used with NLP methods and topic modelling have the usual risk of "garbage in, garbage out". This means that scholars employing these methods must pay close attention to their inputs, lest they generate more garbage than they intend. In our study, we set up a primary annotated word basket which can advance the research of measuring OIPs. We will share our keyword basket with all interested researchers, and hope to construct an open-source resource (see the repository on GitHub: Measuring_OIPs) that would allow others to share and test their own keyword baskets with academic researchers. Due to the laborious process of labeling work, our initial keyword basket is likely to be incomplete and cannot cover all OIP-related phrases comprehensively. Better, more widely shared keyword baskets will reduce the amount of "garbage in", and therefore reduce the amount of "garbage out" in future analyses.

A third limitation is that this analysis is essentially a cross-sectional analysis, limited to three years of pooled observational data (2016–2019) with fixed year effects. Our approach can identify the incidence and salience of open innovation practices, but cannot demonstrate the causal effects of these practices. A more longitudinal analysis would allow researchers to track the patterns of adoption of OIPs over time by firm, and better identify the relationship of those practices to improved firm performance.

A final limitation is to note that NLP models are rapidly advancing. It will be important to ensure that our studies are robust to alternative specifications of the keyword basket for OIPs and alternative methodologies to train algorithms to analyze those practices across other datasets. To the best of our knowledge, the prior literature in open innovation literature has not yet built up a research foundation by using NLP algorithms to develop new ways to measure open innovation. For example, no annotated text-based dataset exists for NLP analysis, where labeled data plays a central role in supervised learning. Therefore, we consider our work as an initial step, which, while it can be improved further, is a valuable starting point. We encourage scholars to further develop the labeled data (i.e., labelling sentences rather than just words or phrases) based on our work, which would enable the use of more advanced NLP models (i.e., supervising learning algorithms) to do further analyses. It will be important to experiment with these other techniques to identify those that generate the best signal-to-noise ratio in analyzing textual data to extract information on innovation practices.

6. Conclusion

Despite the popularity of open innovation in recent years, performance results from the practice of open innovation in the academic literature have been mixed. Measuring and observing open innovation is usually done either in isolation, such as studying a single practice, or in the aggregate, such as employing proxy measures, whereby most of them are based on a survey approach. Based on these proxies, there is not sufficient research investigating the relationship between nuanced open innovation practices and firm financial performance. There are differing claims in the literature on the performance of open innovation, and the evidence used to support these claims often suffers from a number of biases or other limitations.

We employ a new technique that utilizes natural language processing operating on unstructured data to extract information on companies' open innovation practices, to provide a complementary data source and a new method to revisit the relationship between open innovation and firm financial performance. These methods can avoid some of the limitations of earlier methods, utilizing textual data from their annual financial reports. These reporting documents are taken quite seriously both inside the firm and outside the firm. By using these reports as the corpus, the findings overall are quite consistent with prior empirical research on open innovation – open innovation practices in the aggregate level are indeed associated with improved firm performance. We also find evidence for the curvilinearity of the impact of some of these practices (but not all). Our use of NLP allows us to go further, and shed additional light on what may be driving the disparate results in the prior literature, and the open innovation paradox. The plausibility of our findings should provide some comfort to skeptical readers, who might question the validity of NLP methods in understanding open innovation, while also providing more deeper understanding about open innovation adoption. Our results also confirm the complementarity between internal R&D and some specific OIPs but not others, which suggests there is not a clear-cut answer to the relationship between internal R&D effort and open innovation activities. Moreover, the effect of these practices upon firm financial performance varies with the industrial sector in which they are observed. Thus, it is unlikely that there will be a single set of Best Practices for open innovation that would be equally effective across all sectors in the economy. We invite other scholars to build on these methods, to advance our understanding of whether, how and where open innovation influences firm performance, across different practices and across different parts of the economy.

i More information about different measures in prior empirical studies can be found in the Appendix Table A1. We received a number of queries from anonymous reviewers to substantiate certain claims in our paper, and have developed the Appendix Table to provide this information.

ii We were also inspired by recent research in the analysis of ESG influences upon firm performance. While early studies utilized third-party ratings of ESG behavior to evaluate a firm's governance, more recent work has deployed NLP techniques to extract this information from firms' communications directly.

iii We are indebted to an anonymous reviewer for many of the studies shown in this table, which expanded an already large number of studies that were included in an earlier draft of the manuscript.

iv One of our intentions for this paper is to start the process of building a more mature labeled dataset of open innovation practices. Such a dataset will allow other researchers to examine other corpuses of data with a commonly shared set of labels for open innovation practices, as well as replicating the analyses we report here

v We wish to express our thanks to an anonymous reviewer for the suggestion to test curvilinear effects of open innovation on firm financial performance, so that we can connect more directly with prior literature about the mixed results on financial performance.

vi Thanks to an anonymous reviewer for providing a possible reason of why companies don't report crowdsourcing activities in their annual report, or why we cannot capture the crowdsourcing practices by using topic models.

Acknowledgements

We would like to thank Ann-Kristin Zobel, Agnieszka Radziwon, Marcel Bogers, and Christopher Tucci, as well as the community of the Open Innovation Seminar at UC Berkeley for their helpful comments. We would also like to thank the 20 open innovation experts who helped us validate our keyword basket. Special thanks to an NLP expert Hua Yan, for his great support and feedback during the model building process. We are especially grateful to Wim Vanhaverbeke, the editor of Technovation who handled this paper, and four anonymous reviewers for their guidance and feedback.

Appendix

Table A1
Prior empirical studies on open innovation -firm performance relationship

OIP type	Studies	Data source		Performance	Relationship
Technology, market, organizational-oriented OIPs (seven types)	(Ahn et al., 2015)	Survey	Survey	innovation performance & financial performance	+ & n.s
Outside-in OI	(Bagherzadeh et al., 2019)		Survey (European Network of Excellence on Open and Collaborative Innovation)	Innovation performance	+
Outside-in OI	(Bianchi et al., 2016)		Spanish Business Strategy Survey	Innovation performance	Inverted-U
Different types of partners	(Chen et al., 2016)		Surveys in China	Innovation performance (new product sales)	+ & n.s
Outside-in (5) & inside-out OI (4); coupled (3)	(Cheng and Huizingh, 2014)		Asian service firm survey	Innovation performance	+
Outside-in OI (search depth and search breadth)	(Chiang and Hung, 2010)		Survey (Taiwanese electronic manufacturing)	Innovation performance	+
OIPs (search, external sourcing, commercilization, and collaboration)	(Ebersberger et al., 2012)		Survey (CIS)	Innovation performance	+
Technology alliance vs. internal innovation effort	(Faems et al., 2010)		Survey (CIS)	Financial performance (personnel costs in value added and profit margin)	-
Partnering and contracting; openness to knowledge (outside-in)	(Fey and Birkinshaw, 2005)		Interview and questionnaire	R&D performance	+
Customer interaction	(Foss et al., 2011)		Questionnaire	Innovation performance	+
External search breadth and depth (outside-in OI); coupled OI	(Greco et al., 2016)		Survey (CIS)	Industrial innovation performance; economic-financial innovation performance	Curvilinear; some OIPs are more effective than others
Internal R&D and external R&D	(Grimpe and Kaiser, 2010)		Survey (CIS)	Innovation performance	inverse U-shaped
Collaborative vs. transactional open innovation activities	(Grimpe and Sofka, 2016)		Survey (CIS)	Innovation performance	complementarity
Networking, technology buy-in	(Huang and Rice, 2018)		Survey	Innovation performance (RDI)	+ & -
External technology acquisition and exploitation	(Hung and Chou, 2013)		Survey	Financial performance (Tobin's Q)	+
External search breadth and depth	(Hwang and Lee, 2010)		Survey (Korean Innovation Survey)	Innovation performance	+
Outsid-in process in R&D management	(Inauen and Schenker-Wicki, 2011)		Survey	Innovation performance (product, process, sales share)	+
External knowledge sourcing (information transfer from informal network, technology acquisition, R&D collaboration)	(Kang and Kang, 2009)		Survey	Innovation performance	+ & inverted-U-shape
Knowledge sources from different partners	(Köhler et al., 2012)		Survey (CIS)	Innovation efforts (new-to-market innovations and imitations); innovation success	+
External relationship	(Lasagni, 2012)		Survey	Innovation performance	+
External search breath and depth	(Laursen and Salter, 2006)		Survey (CIS)	Innovation performance	+
Outside-in OI	(Lichtenthaler, 2009)		Survey	Financial performance (ROS)	+
External linkages (customers, suupliers, strategic alliances etc)	(Love and Mansury, 2007)		Questionnaire	Innovation activities	+
Outside-in (breadth & depth) and coupled OIPs; operational reconfiguration capabilities	(Ovuakporie et al., 2021)		Survey (CIS)	Innovation performance	+
Technology scouting & sourcing, vertical and horizontal technology collaboration	(Parida et al., 2012)		Survey	Innovation performance	+
External knowledge sharing, accidental & intentional knowledge leakage	(Ritala et al., 2015)		Survey	Innovation performance	+ & dilemma
Collaborations with different technological partners	(Santamaria and Surroca, 2011)		Survey	Innovation performance	+

(*continued on next page*)

Table A1 (*continued*)

OIP type	Studies	Data source		Performance	Relationship
OI; relational capability; network spillover	(Sisodiya et al., 2013)		Mixed method (qualitative &quantitative)	Financial performance (Tobin's Q)	+ (but not straightforward)
OIPs (search strategy; external R&D; cooperation; protection)	(Spithoven et al., 2013)		Survey (CIS)	Innovative sales	+
R&D outsourcing activities, and external technology acquisition	(Tsai and Wang, 2009)		Survey	Technological innovation performance	n.s.
R&D colloabrations with uni, suppliers, customers and competitors	(Un et al., 2010)		Survey	Production innovation	+ & - & n.s
external knowledge (buying and cooperating); external sources (industrial agents & scientific agents)	(Vega-Jurado et al., 2009)		Survey	Innovation outcome	+ & - & n.s
Highly and weakly interactive OIPs	(Zacharias et al., 2020)		Survey	Innovation success (technological and market)	+ (depends on the
Science-based and market-based OI partnerships of R&D project	(Du et al., 2014)	Other datasets	cross-sectional dataset on R&D projects (manufacturing company)	Financial performance (the total revenues from the 'transferred' outcomes of an R&D project)	+ & - & n.s
Outside-in OI (external scientific knowledge/technologies)	(Kafouros and Forsans, 2012)		firm-level operating information	Financial performance (profitability)	+
level of vertical integration; alliance activity; level of internationalization	(Li and Tang, 2010)		Compustat database, USPTO, SDC	The quality of innovation output (patent citations)	inverted U-shape
Different types and stages of OI projects	(Marullo et al., 2020)		Exploratory approach (multiple case study)	Value creation	+&-
Outside-in, inside-out, coupled OIPs (fine-grain)	(Mazzola et al., 2012)		10K	Innovation performance & financial performance	+ & - & n.s
Outside-in, inside-out, coupled OIPs (fine-grain)	(Mazzola et al., 2016)		10k	Innovation performance & financial performance	+ & - & n.s
Inward technology licensing	(Tsai and Wang, 2009)		Panel dataset	Firm performance	+
Technological knowledge diversity	(Wadhwa and Kotha, 2006)		USPTO	Successful patent applications	+
Technology licensing	(Wang et al., 2013)		SIPO of China (patent dataset)	Innovation performance	+
Six OIPs	(Xie and Wang, 2020)		QCA	Product innovation	+ & n.s

Table A2
Three-step test for the inverted-U shape between open innovation and financial performance

Three-step test procedure		OIP_sum	OIP1	OIP2	OIP3	OIP4	OIP5
1	The coefficient of independent variable (IV)	0.608	0.995	1.471	−0.223	2.489	1.065
		(p = 0.000)	(p = 0.003)	(p = 0.000)	(p = 0.398)	(p = 0.000)	(p = 0.000)
	The coefficient of IV square	−0.290	−1.592	−0.546	−0.324	−2.576	−1.391
		(p = 0.1)	(p = 0.005)	(p = 0.212)	(p = 0.381)	(p = 0.000)	(p = 0.000)
2	Turning point (TF) (Data range)	1.048	0.313	1.347	−0.34	0.483	0.383
		[0,1]	[0, 0.989]	[0, 0.997]	[0, 0.999]	[0.0.966]	[0, 1]
	Slope of *IV_low*	–	0.995	–	–	2.489	1.065
	Slope of *IV_high*	–	−2.153	–	–	−7.465	−1.717
3	Fieller test (95% confidence interval)	–	[0.226, 0.486]	–	–	[0.392, 0.744]	[0.316, 0.454]
Overall test of presence of an inversed-U shape		–	0.005	–	–	0.008	0.000

References

Ahn, J.M., Minshall, T., Mortara, L., 2015. Open innovation: a new classification and its impact on firm performance in innovative SMEs. J. Innovat. Manag. 3, p33–54.

Alghamdi, R., Alfalqi, K., 2015. A survey of topic modeling in text mining. Int. J. Adv. Comput. Sci. Appl. 6.

Alti, A., 2006. How persistent is the impact of market timing on capital structure? J. Finance 61, 1681–1710.

Anderson, E.W., Fornell, C., Mazvancheryl, S.K., 2004. Customer satisfaction and shareholder value. J. Mark. 68, 172–185.

Antons, D., Grünwald, E., Cichy, P., Salge, T.O., 2020. The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. R D Manag. 50, 329–351.

Arora, A., Athreye, S., Huang, C., 2016. The paradox of openness revisited: collaborative innovation and patenting by UK innovators. Res. Pol. 45, 1352–1361.

Arts, S., Hou, J., Gomez, J.C., 2020. Natural language processing to identify the creation and impact of new technologies in patent text: code, data, and new measures. Res. Pol. 50, 104144.

Artz, K.W., Norman, P.M., Hatfield, D.E., Cardinal, L.B., 2010. A longitudinal study of the impact of R&D, patents, and product innovation on firm performance. J. Prod. Innovat. Manag. 27, 725–740.

Asakawa, K., Nakamura, H., Sawada, N., 2010. Firms' open innovation policies, laboratories' external collaborations, and laboratories' R&D performance. R D Manag. 40, 109–123.

Bagherzadeh, M., Markovic, S., Cheng, J., Vanhaverbeke, W., 2019. How does outside-in open innovation influence innovation performance? Analyzing the mediating roles of knowledge sharing and innovation strategy. IEEE Trans. Eng. Manag. 67, 740–753.

Belderbos, R., Faems, D., Leten, B., Van Looy, B., 2010. Technological activities and their impact on the financial performance of the firm: exploitation and exploration within and between firms. J. Prod. Innovat. Manag. 27, 869–882. https://doi.org/10.1111/j.1540-5885.2010.00757.x.

Bharadwaj, A.S., Bharadwaj, S.G., Konsynski, B.R., 1999. Information technology effects on firm performance as measured by Tobin's q. Manag. Sci. 45, 1008–1024.

Bianchi, M., Croce, A., Dell'Era, C., Di Benedetto, C.A., Frattini, F., 2016. Organizing for inbound open innovation: how external consultants and a dedicated R & D unit influence product innovation performance. J. Prod. Innovat. Manag. 33, 492–510.

Blagoeva, R.R., Mom, T.J.M., Jansen, J.J.P., George, G., 2020. Problem-solving or self-enhancement? A power perspective on how CEOs affect R&D search in the face of inconsistent feedback. Acad. Manag. J. 63, 332–355. https://doi.org/10.5465/amj.2017.0999.

Blei, D.M., 2012. Probabilistic topic models. Commun. ACM 55, 77–84.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022.

Bogers, M., 2011. The open innovation paradox: knowledge sharing and protection in R&D collaborations. Eur. J. Innovat. Manag.

Bogers, M., Zobel, A.-K., Afuah, A., Almirall, E., Brunswicker, S., Dahlander, L., Frederiksen, L., Gawer, A., Gruber, M., Haefliger, S., 2017. The open innovation research landscape: established perspectives and emerging themes across different levels of analysis. Ind. Innovat. 24, 8–40.

Brunswicker, S., Chesbrough, H., 2018. The Adoption of Open Innovation in Large Firms: practices, Measures, and RisksA survey of large firms examines how firms approach open innovation strategically and manage knowledge flows at the project level. Res. Technol. Manag. 61, 35–45. https://doi.org/10.1080/08956308.2018.1399022.

Brunswicker, S., Chesbrough, H., 2015. Open Innovation Executive Survey. Cent. Digit. Open Innov. Purdue Univ.

Brunswicker, S., Vanhaverbeke, W., 2015. Open innovation in small and medium-sized enterprises (SMEs): external knowledge sourcing strategies and internal organizational facilitators. J. Small Bus. Manag. 53, 1241–1263.

Cao, J., Xia, T., Li, J., Zhang, Y., Tang, S., 2009. A density-based method for adaptive LDA model selection. Neurocomputing 72, 1775–1781.

Cassiman, B., Valentini, G., 2016. Open innovation: are inbound and outbound knowledge flows really complementary? Strat. Manag. J. 37, 1034–1046.

Cassiman, B., Veugelers, R., 2006. In search of complementarity in innovation strategy: internal R&D and external knowledge acquisition. Manag. Sci. 52, 68–82.

Chen, Y., Vanhaverbeke, W., Du, J., 2016. The interaction between internal R & D and different types of external knowledge sourcing: an empirical study of C hinese innovative firms. R D Manag. 46, 1006–1023.

Cheng, C.C.J., Huizingh, E.K.R.E., 2014. When is open innovation beneficial? The role of strategic orientation. J. Prod. Innovat. Manag. 31, 1235–1253.

Chesbrough, H., 2019. Open Innovation Results: Going beyond the Hype and Getting Down to Business. Oxford University Press.

Chesbrough, H., 2011. Open Services Innovation: Rethinking Your Business to Grow and Compete in a New Era. John Wiley & Sons.

Chesbrough, H., Bogers, M., 2014. Explicating Open Innovation: Clarifying an Emerging Paradigm for Understanding Innovation. New Front. Open Innov. Oxford Oxford Univ. Press. Forthcom. 3–28.

Chesbrough, H.W., 2003. Open Innovation: the New Imperative for Creating and Profiting from Technology. Harvard Business Press.

Chiang, Y., Hung, K., 2010. Exploring open search strategies and perceived innovation performance from the perspective of inter-organizational knowledge flows. R&d Manag 40, 292–299.

Choudhury, P., Wang, D., Carlson, N.A., Khanna, T., 2019. Machine learning approaches to facial and text analysis: discovering CEO oral communication styles. Strat. Manag. J. 40, 1705–1732.

Chung, K.H., Pruitt, S.W., 1994. A simple approximation of Tobin's q. Financ. Manag. 70–74.

Cohen, W.M., Levinthal, D.A., 1990. Absorptive capacity: a new perspective on learning and innovation. Adm. Sci. Q. 128–152.

Corritore, M., Goldberg, A., Srivastava, S.B., 2020. Duality in diversity: how intrapersonal and interpersonal cultural heterogeneity relate to firm performance. Adm. Sci. Q. 65, 359–394.

Criscuolo, C., Haskel, J.E., Slaughter, M.J., 2010. Global engagement and the innovation activities of firms. Int. J. Ind. Organ. 28, 191–202.

Dahlander, L., Gann, D.M., 2010. How open is innovation? Res. Pol. 39, 699–709.

Deb, P., David, P., O'Brien, J., 2017. When is cash good or bad for firm performance? Strat. Manag. J. 38, 436–454. https://doi.org/10.1002/smj.2486.

Du, J., Leten, B., Vanhaverbeke, W., 2014. Managing open innovation projects with science-based and market-based partners. Res. Pol. 43, 828–840.

Duysters, G., Hagedoorn, J., 2000. Organizational Modes of Strategic Technology Partnering.

Easterby-Smith, M., Lyles, M.A., Tsang, E.W.K., 2008. Inter-organizational knowledge transfer: current themes and future prospects. J. Manag. Stud. 45, 677–690.

Ebersberger, B., Bloch, C., Herstad, S.J., Van De Velde, E.L.S., 2012. Open innovation practices and their effect on innovation performance. Int. J. Innovat. Technol. Manag. 9, 1250040.

Eppinger, S.D., 2001. Innovation at the speed of information. Harv. Bus. Rev. 79, 149–158.

Escribano, A., Fosfuri, A., Tribó, J.A., 2009. Managing external knowledge flows: the moderating role of absorptive capacity. Res. Pol. 38, 96–105.

Faems, D., De Visser, M., Andries, P., Van Looy, B., 2010. Technology alliance portfolios and financial performance: value-enhancing and cost-increasing effects of open innovation. J. Prod. Innov. Manag. 27, 785–796.

Felin, T., Zenger, T.R., 2020. Open innovation: a theory-based view. Strateg. Manag. Rev. 1, 223–232.

Fey, C.F., Birkinshaw, J., 2005. External sources of knowledge, governance mode, and R&D performance. J. Manage. 31, 597–621.

Foege, J.N., Lauritzen, G.D., Tietze, F., Salge, T.O., 2019. Reconceptualizing the paradox of openness: how solvers navigate sharing-protecting tensions in crowdsourcing. Res. Pol. 48, 1323–1339.

Foss, N.J., Husted, K., Michailova, S., 2010. Governing knowledge sharing in organizations: levels of analysis, governance mechanisms, and research directions. J. Manag. Stud. 47, 455–482.

Foss, N.J., Laursen, K., Pedersen, T., 2011. Linking customer interaction and innovation: the mediating role of new organizational practices. Organ. Sci. 22, 980–999.

Frishammar, J., Ericsson, K., Patel, P.C., 2015. The dark side of knowledge transfer: exploring knowledge leakage in joint R&D projects. Technovation 41, 75–88.

Ghanem, B., Karoui, J., Benamara, F., Moriceau, V., Rosso, P., 2019. Idat at fire2019: overview of the track on irony detection in Arabic tweets. In: Proceedings of the 11th Forum for Information Retrieval Evaluation, pp. 10–13.

Greco, M., Grimaldi, M., Cricelli, L., 2016. An analysis of the open innovation effect on firm performance. Eur. Manag. J. 34, 501–516.

Grimpe, C., Kaiser, U., 2010. Balancing internal and external knowledge acquisition: the gains and pains from R&D outsourcing. J. Manag. Stud. 47, 1483–1509.

Grimpe, Christoph, Sofka, Wolfgang, 2009. Search patterns and absorptive capacity: Low- and high-technology sectors in European countries. Research Policy 38, 495–506.

Grimpe, C., Sofka, W., 2016. Complementarities in the search for innovation—managing markets and relationships. Res. Pol. 45, 2036–2053.

Haans, R.F., Pieters, C., He, Z., 2015. Thinking about U: theorizing and testing U- and inverted U-shaped relationships in strategy research. Strat. Manag. J. 37, 1177–1195.

Hagedoorn, J., Wang, N., 2012. Is there complementarity or substitutability between internal and external R&D strategies? Res. Pol. 41, 1072–1083.

Hair, J.F., Ringle, C.M., Sarstedt, M., 2011. PLS-SEM: indeed a silver bullet. J. Mark. theory Pract. 19, 139–152.

He, V.F., Puranam, P., Shrestha, Y.R., von Krogh, G., 2020. Resolving governance disputes in communities: a study of software license decisions. Strat. Manag. J. 41, 1837–1868.

Hess, A.M., Rothaermel, F.T., 2011. When are assets complementary? Star scientists, strategic alliances, and innovation in the pharmaceutical industry. Strat. Manag. J. 32, 895–909.

Huang, F., Rice, J., 2018. The role of absorptive capacity in facilitating "open innovation" outcomes: A study of Australian SMEs in the manufacturing sector. In: Promoting Innovation in New Ventures and Small-and Medium-Sized Enterprises. World Scientific, pp. 477–500.

Huizingh, E.K.R.E., 2011. Open innovation: state of the art and future perspectives. Technovation 31, 2–9.

Hung, K.-P., Chou, C., 2013. The impact of open innovation on firm performance: the moderating effects of internal R&D and environmental turbulence. Technovation 33, 368–380.

Hwang, J., Lee, Y., 2010. External knowledge search, innovative performance and productivity in the Korean ICT sector. Telecomm. Policy 34, 562–571.

Inauen, M., Schenker-Wicki, A., 2011. The impact of outside-in open innovation on innovation performance. Eur. J. Innov. Manag.

Josefy, M., Kuban, S., Ireland, R.D., Hitt, M.A., 2015. All things great and small: organizational size, boundaries of the firm, and a changing environment. Acad. Manag. Ann. 9, 715–802.

Kafouros, M.I., Forsans, N., 2012. The role of open innovation in emerging economies: do companies profit from the scientific knowledge of others? J. World Bus. 47, 362–370.

Kang, Y., Cai, Z., Tan, C.-W., Huang, Q., Liu, H., 2020. Natural language processing (NLP) in management research: a literature review. J. Manag. Anal. 7, 139–172.

Kang, K.H., Kang, J., 2009. How do firms source external knowledge for innovation? Analysing effects of different knowledge sourcing methods. Int. J. Innov. Manag. 13, 1–17.

Kaplan, S., Vakili, K., 2015. The double-edged sword of recombination in breakthrough innovation. Strat. Manag. J. 36, 1435–1457.

Karniouchina, E.V., Carson, S.J., Short, J.C., Ketchen Jr., D.J., 2013. Extending the firm vs. industry debate: does industry life cycle stage matter? Strat. Manag. J. 34, 1010–1018.

Karoui, J., Benamara, F., Moriceau, V., Patti, V., Bosco, C., Aussenac-Gilles, N., 2017. Exploring the Impact of Pragmatic Phenomena on Irony Detection in Tweets: A Multilingual Corpus Study. Association for Computational Linguistics (ACL).

Katzy, B., Turgut, E., Holzmann, T., Sailer, K., 2013. Innovation intermediaries: a process view on open innovation coordination. Technol. Anal. Strat. Manag. 25, 295–309.

Khodak, M., Saunshi, N., Vodrahalli, K., 2017. A Large Self-Annotated Corpus for Sarcasm arXiv Prepr. arXiv1704.05579.

Knott, A.M., 2003. Persistent heterogeneity and sustainable innovation. Strat. Manag. J. 24, 687–705.

Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Q. J. Econ. 132, 665–712.

Köhler, C., Sofka, W., Grimpe, C., 2012. Selective search, sectoral patterns, and the impact on product innovation performance. Res. Pol. 41, 1344–1356.

Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data, pp. 159–174. Biometrics.

Lang, L.H.P., Stulz, R.M., 1994. Tobin's q, corporate diversification, and firm performance. J. Polit. Econ. 102, 1248–1280.

Lasagni, A., 2012. How can external relationships enhance innovation in SMEs? New evidence for Europe. J. small Bus. Manag. 50, 310–339.

Laursen, K., Salter, A., 2006. Open for innovation: the role of openness in explaining innovation performance among U.K. manufacturing firms. Strat. Manag. J. 27, 131–150. https://doi.org/10.1002/smj.507.

Laursen, K., Salter, A.J., 2014. The paradox of openness: appropriability, external search and collaboration. Res. Pol. 43, 867–878.

Lee, S., Park, G., Yoon, B., Park, J., 2010. Open innovation in SMEs-An intermediated network model. Res. Pol. 39, 290–300. https://doi.org/10.1016/j.respol.2009.12.009.

Li, F., Lundholm, R., Minnis, M., 2013. A measure of competition based on 10-K filings. J. Account. Res. 51, 399–436.

Li, H.-L., Tang, M.-J., 2010. Vertical integration and innovative performance: the effects of external knowledge sourcing modes. Technovation 30, 401–410.

Lichtenthaler, U., 2015. A note on outbound open innovation and firm performance. R D Manag. 45, 606–608.
Lichtenthaler, U., 2009. Outbound open innovation and its effect on firm performance: examining environmental influences. R D Manag. 39, 317–330.
Lin, B.-W., Lee, Y., Hung, S.-C., 2006. R&D intensity and commercialization orientation effects on financial performance. J. Bus. Res. 59, 679–685.
Lind, J.T., Mehlum, H., 2010. With or without u? the appropriate test for a U-shaped relationship. Oxf. Bull. Econ. Stat. 72, 109–118. https://doi.org/10.1111/j.1468-0084.2009.00569.x.
Liu, Q., Du, Q., Hong, Y., Fan, W., Wu, S., 2020. User idea implementation in open innovation communities: evidence from a new product development crowdsourcing community. Inf. Syst. J. 30, 899–927.
Lokshin, B., Belderbos, R., Carree, M., 2008. The productivity effects of internal and external R&D: evidence from a dynamic panel data model. Oxf. Bull. Econ. Stat. 70, 399–413.
Love, J.H., Mansury, M.A., 2007. External linkages, R&D and innovation performance in US business services. Ind. Innov. 14, 477–496.
Lu, J.W., Beamish, P.W., 2004. International diversification and firm performance: the S-curve hypothesis. Acad. Manag. J. 47, 598–609.
Madsen, T.L., Walker, G., 2002. The evolution of heterogeneity in performance, 10510. In: Academy of Management Proceedings. Academy of Management Briarcliff Manor, NY. https://doi.org/10.5465/apbpp.2002.7516550. vols. 1–V6.
Marullo, C., Di Minin, A., De Marco, C., Piccaluga, A., 2020. Is open innovation always the best for SMEs? An exploratory analysis at the project level. Creativ. Innovat. Manag. 29, 209–223.
Maslowska, E., Malthouse, E.C., Collinger, T., 2016. The customer engagement ecosystem. J. Market. Manag. 32, 469–501.
Mazzola, E., Bruccoleri, M., Perrone, G., 2012. The effect of inbound, outbound and coupled innovation on performance. Int. J. Innovat. Manag. 16, 1240008.
Mazzola, E., Bruccoleri, M., Perrone, G., 2016. Open innovation and firms performance: state of the art and empirical evidences from the bio-pharmaceutical industry. Int. J. Technol. Manag. 70, 109–134.
Miles, G., Snow, C.C., Sharfman, M.P., 1993. Industry variety and performance. Strat. Manag. J. 14, 163–177.
Nagaoka, S., Motohashi, K., Goto, A., 2010. Patent statistics as an innovation indicator. In: Handbook of the Economics of Innovation. Elsevier, pp. 1083–1127.
OECD, 1997. Oslo Manual: Proposed Guidelines for Collecting and Interpreting Technological Innovation Data. Organization for Economic.
Oltra, M.J., Flor, M.L., Alfaro, J.A., 2018. Open innovation and firm performance: the role of organizational mechanisms. Bus. Process Manag. J.
Ovuakporie, O.D., Pillai, K.G., Wang, C., Wei, Y., 2021. Differential moderating effects of strategic and operational reconfiguration on the relationship between open innovation practices and innovation performance. Res. Pol. 50, 104146.
Parida, V., Westerberg, M., Frishammar, J., 2012. Inbound open innovation activities in high-tech SMEs: the impact on innovation performance. J. small Bus. Manag. 50, 283–309.
Passonneau, R.J., Rudin, C., Radeva, A., Liu, Z.A., 2009. Reducing noise in labels and features for a real world dataset: application of NLP corpus annotation methods. In: International Conference on Intelligent Text Processing and Computational Linguistics. Springer, pp. 86–97.
Podmetina, D., Fiegenbaum, I., Teplov, R., Albats, E., 2014. Towards open innovation measurement system-a literature review. In: ISPIM Conference Proceedings. The International Society for Professional Innovation Management (ISPIM), p. 1.
Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. 88, 879.
Pullen, A.J.J., de Weerd-Nederhof, P.C., Groen, A.J., Fisscher, O.A.M., 2012. Open innovation in practice: goal complementarity and closed NPD networks to explain differences in innovation performance for SMEs in the medical devices sector. J. Prod. Innovat. Manag. 29, 917–934.
Qiu, Y., Wang, T.Y., 2018. Skilled Labor Risk and Compensation Policies. Available SSRN 2935378.
Radziwon, A., Bogers, M., 2019. Open innovation in SMEs: exploring inter-organizational relationships in an ecosystem. Technol. Forecast. Soc. Change 146, 573–587.
Ritala, P., Olander, H., Michailova, S., Husted, K., 2015. Knowledge sharing, knowledge leaking and relative innovation performance: An empirical study. Technovation 35, 22–31.
Roberts, A., Gaizauskas, R., Hepple, M., Demetriou, G., Guo, Y., Roberts, I., Setzer, A., 2009. Building a semantically annotated corpus of clinical texts. J. Biomed. Inf. 42, 950–966.
Rothaermel, F.T., Hess, A.M., 2007. Building dynamic capabilities: innovation driven by individual-, firm-, and network-level effects. Organ. Sci. 18, 898–921.
Santamaria, L., Surroca, J., 2011. Matching the goals and impacts of R&D collaboration. Eur. Manag. Rev. 8, 95–109.
Sawhney, M., Verona, G., Prandelli, E., 2005. Collaborating to create: the Internet as a platform for customer engagement in product innovation. J. Interact. Market. 19, 4–17.
Schilling, M.A., 2009. Understanding the alliance data. Strat. Manag. J. 30, 233–260. https://doi.org/10.2139/ssrn.2533187.
Schmiedeberg, C., 2008. Complementarities of innovation activities: an empirical analysis of the German manufacturing sector. Res. Pol. 37, 1492–1503.
Schroll, A., Mild, A., 2012. A critical review of empirical research on open innovation adoption. J. Betriebswirtschaft 62, 85–118.
Schulz, M., 2001. The uncertain relevance of newness: organizational learning and knowledge flows. Acad. Manag. J. 44, 661–681.
Shaikh, M., Levina, N., 2019. Selecting an open innovation community as an alliance partner: looking for healthy communities and ecosystems. Res. Pol. 48, 103766.
Shrestha, Y.R., He, V.F., Puranam, P., von Krogh, G., 2020. Algorithm supported induction for building theory: how can we use prediction models to theorize? Organ. Sci.
Sisodiya, S.R., Johnson, J.L., Grégoire, Y., 2013. Inbound open innovation for enhanced performance: enablers and opportunities. Ind. Market. Manag. 42, 836–849.
Skalicky, S., Crossley, S., 2018. Linguistic features of sarcasm and metaphor production quality. In: Proceedings of the Workshop on Figurative Language Processing, pp. 7–16.
Song, T., Huang, J., Tan, Y., Yu, Y., 2019. Using user-and marketer-generated content for box office revenue prediction: differences between microblogging and third-party platforms. Inf. Syst. Res. 30, 191–203.
Spithoven, A., 2013. Open innovation practices and innovative performances: an international comparative perspective. Int. J. Technol. Manag. 62, 1–34.
Spithoven, A., Clarysse, B., Knockaert, M., 2010. Building absorptive capacity to organise inbound open innovation in traditional industries. Technovation 30, 130–141.
Spithoven, A., Vanhaverbeke, W., Roijakkers, N., 2013. Open innovation practices in SMEs and large enterprises. Small Bus. Econ. 41, 537–562.
Teodoridis, F., Lu, J., Furman, J.L., 2020. Measuring the Direction of Innovation: Frontier Tools in Unassisted Machine Learning. Available SSRN 3596233.
Terwiesch, C., Xu, Y., 2008. Innovation contests, open innovation, and multiagent problem solving. Manag. Sci. 54, 1529–1543.
Tidhar, R., Eisenhardt, K.M., 2020. Get rich or die tryin finding revenue model fit using machine learning and multiple cases. Strat. Manag. J. 41, 1245–1273. https://doi.org/10.1002/smj.3142.
Tsai, K.-H., Wang, J.-C., 2009. External technology sourcing and innovation performance in LMT sectors: An analysis based on the Taiwanese Technological Innovation Survey. Res. Policy 38, 518–526.
Un, C.A., Cuervo-Cazurra, A., Asakawa, K., 2010. R&D collaborations and product innovation. J. Prod. Innov. Manag. 27, 673–689.
Vahter, P., Love, J.H., Roper, S., 2014. Openness and innovation performance: are small firms different? Ind. Innovat. 21, 553–573.
Vakili, K., Kaplan, S., 2020. Organizing for innovation: a contingency view on innovative team configuration. Strat. Manag. J.
Vanhaverbeke, W., Frattini, F., Roijakkers, N., Usman, M., 2018. Researching Open Innovation in SMEs. World Scientific.
Vega-Jurado, J., Gutiérrez-Gracia, A., Fernández-de-Lucio, I., 2009. Does external knowledge sourcing matter for innovation? Evidence from the Spanish manufacturing industry. Ind. Corp. Chang. 18, 637–670.
Veugelers, R., Cassiman, B., 1999. Make and buy in innovation strategies: evidence from Belgian manufacturing firms. Res. Pol. 28, 63–80.
Wadhwa, A., Kotha, S., 2006. Knowledge creation through external venturing: evidence from the telecommunications equipment manufacturing industry. Acad. Manag. J. 49, 819–835.
Wang, T., Libaers, D., Park, H.D., 2017. The paradox of openness: how product and patenting experience affect R&D sourcing in C hina? J. Prod. Innovat. Manag. 34, 250–268.
Wang, Y., Zhou, Z., Li-Ying, J., 2013. The impact of licensed-knowledge attributes on the innovation performance of licensee firms: evidence from the Chinese electronic industry. J. Technol. Transf. 38, 699–715.
West, J., Kuk, G., 2016. The complementarity of openness: how MakerBot leveraged Thingiverse in 3D printing. Technol. Forecast. Soc. Change 102, 169–181.
Westerlund, M., Leminen, S., Rajahonka, M., 2018. A topic modelling analysis of living labs research. Technol. Innov. Manag. Rev. 8.
Xia, F., Yetisgen-Yildiz, M., 2012. Clinical corpus annotation: challenges and strategies. In: Proceedings of Third Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM'2012) of the International Conference on Language Resources and Evaluation. LREC).
Xie, X., Wang, H., 2020. How can open innovation ecosystem modes push product innovation forward? An fsQCA analysis. J. Bus. Res. 108, 29–41.
Zacharias, N.A., Daldere, D., Winter, C.G.H., 2020. Variety is the spice of life: How much partner alignment is preferable in open innovation activities to enhance firms' adaptiveness and innovation success? J. Bus. Res. 117, 290–301.
Zobel, A.-K., Balsmeier, B., Chesbrough, H., 2016. Does patenting help or hinder open innovation? Evidence from new entrants in the solar industry. Ind. Corp. Change 25, 307–331.