



Decision optimization in cooperation innovation: the impact of big data analytics capability and cooperative modes

Guojun Ji¹ · Muhong Yu¹ · Kim Hua Tan² · Ajay Kumar³ · Shivam Gupta⁴

Accepted: 28 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Data-driven innovation enables firms to design products that are more responsive to market needs, which greatly reduces the risk of innovation. Customer data in the same supply chain has certain commonality, but data separation makes it difficult to maximize data value. The selection of an appropriate mode for cooperation innovation should be based on the particular big data analytics capability of the firms. This paper focuses on the influence of big data analytics capability on the choice of cooperation mode, and the influence of their matching relationship on cooperation performance. Specifically, using game-theoretic models, we discuss two cooperation modes, data analytics is implemented individually (i.e., loose cooperation) by either firm, or jointly (tight cooperation) by both firms, and further discuss the addition of coordination contracts under the loose mode. Several important conclusions are obtained. Firstly, both firms' big data capability have positive effects on the selection of tight cooperation mode. Secondly, with the improvement of big data capability, the firms' innovative performance gaps between loose and tight mode will increase significantly. Finally, when the capability meet certain condition, the cost subsidy contract can alleviate the gap between the two cooperative models.

Keywords Big data analytics · Cooperation innovation · Supply chain management

✉ Muhong Yu
9212533@qq.com

Guojun Ji
jiking@xmu.edu.cn

Kim Hua Tan
kim.tan@nottingham.ac.uk

Ajay Kumar
akumar@em-lyon.com

Shivam Gupta
shivam.gupta@neoma-bs.fr

¹ Management School, Xiamen University, Fujian, China

² Department of Operations and Innovation Management, Nottingham University Business School, Nottingham, UK

³ EMLYON Business School, Ecully, France

⁴ NEOMA Business School, Reims, France

1 Introduction

Big data are increasingly driving the changes of decision-making in firms (Brynjolfsson & McElheran, 2016; Fosso Wamba et al., 2018). In the big data environment, management decision-making problems expand from the internal domain to the cross-domain environment and the supplement of cross-domain information makes the measurement of decision factors more complete and reliable, thus improving the accuracy of management decision-making (Davenport et al., 2012; McAfee & Brynjolfsson, 2012). Big data is a complex data set, which needs to realize value through multiple dynamic processes such as data identification, collection, storage and analysis (Lin & Kunnathur, 2019). In these processes, big data analysis is considered to be the most critical link in transforming general knowledge in data into specific knowledge (Xu et al., 2016, Wamba et al., 2017a). By using big data analysis technology, firms can acquire specific knowledge resources needed for product innovation more quickly (Ferraris et al., 2019) to transform business into competitive advantages and help improve business performance (Côte-Real et al., 2017). Compared with firms lagging behind in big data analytics capability, leading firms can capture product development direction in turbulent environment, acquire technical knowledge, develop new products and successfully achieve product innovation (Lin & Kunnathur, 2019). All these information indicate that big data analytics capability have been widely considered a key competitive advantage of marketing and innovation (Feng & Shanthikumar, 2018), also have a positive interaction effect on market performance (Dong & Yang, 2020).

As the amount of data of explosive growth, big data may require a large scale of data centers with huge computing power and resources, which give rise to more consumptions of resources, increasing firm financial pressures (Wu et al., 2016). The effectiveness of decision-making are only as good as the data on which they are based (Hazen et al., 2014). The mature application of big data technology makes it possible for firms to take advantage of consumers' data resources (Bendle & Wang, 2016). Consumers do not need expertise or initiative to generate big data automatically through online behavior that can add value to firms (Xie et al., 2016). Although big data technology improves the degree of digitization of consumer behavior and makes the data generated by it highly accessible and of high commercial value (Erevelles et al., 2016), firms will lose their advantages due to the huge burden in analyzing massive and complicated consumer data (Gruner et al., 2014, Menguc et al., 2014). Data collection and the IT that enables data mining need to focus on data validity and analysis techniques are used (Hazen et al., 2018).

Therefore, some of firms begin to seek external resources. One common method is to outsource big data project, for buying related big data information from the related Data Company (Liu & Yi, 2018a). Purchasing customer information is only applicable to a single project, which is not conducive to firms to build their own data competitive advantage. Another ways is invest in cloud-based big data analytics (Liu et al., 2020a). Cloud-based tools that are low cost with access to sources of consumer data such as social media or other internet retailers, which will help firms overcome some of the main barriers in undertaking new technologies. Due to the different roles of firms in the supply chain, upstream and downstream obtain the data value advantage from different sources, which leads to the different innovation resource. Therefore, it is a feasible and effective big data strategy for firms, especially SMEs, to build cloud-based big data analytics capability and cooperate with supply chain members.

From the perspective of enterprise practice, innovation cooperation based on big data have two modes, separately or in combination. In a loose cooperation, firms can share their data or knowledge but take big data activities independently. In a tight cooperation, firms jointly

establish the big data center, and can perfectly share the outcomes. Tight type focus on co-creation, participants usually invest capital and human resources in a certain proportion to joint establish the data center where data property rights are shared, benefits are shared and risks are shared, so as to maximize data value. Generally, firms are less likely to choose a joint mode when knowledge bases are very different (Sampson, 2004). Loose mode focus on collaboration, based on transaction contract, participants formulate contracts to clearly define the division of tasks and the distribution of benefits between the two parties. The negotiation process is more clear and targeted, and the communication mechanism is easier to establish. Compared with the tight mode, firms in the loose relationship still conduct data analysis within the boundaries of their respective organizations, so the degree of interaction between firms is lower than that in the tight mode, and the relationship of data property rights is simpler. Loose mode provides firms with higher flexibility, tight mode provides with higher data analysis performance.

Customer data in the same supply chain has certain commonality, but the data format, data storage and other data problems make it difficult for firms to directly share data or technology (Ali et al., 2017). Although combination data from various data sources is the major driver for generating additional value, synthesized data have a greater value than the sum of their individual parts (Shollo & Galliers, 2016). Data separation makes it difficult to maximize data value. How to excavate the information of target groups effectively is an important problem to be solved in cooperative innovation of supply chain. In addition, big data analytics capabilities positively influence co-innovation process outcomes (Lozada et al., 2019), also empower participants' collaboration in cooperation innovation, facilitating the creation knowledge (Lozada et al., 2019). In this sense, the big data analytics capability and its matching degree with the partners determine the performance of cooperation innovation.

Under the above conditions, firms need to selecting appropriate mode to balance the relationships between the costs spending on big data and the revenues getting from innovation cooperation. Firms involved in innovation cooperation are more concerned with their individual profits than the channel profits (Ge et al., 2014). For this reason, this paper studies the decision-making of cooperative innovation in supply chain from the perspective of big data analytics capability, and mainly solves three questions: (1) How will the change of big data analytics capability affect firms' choice of cooperative mode? (2) How does a firm's investment in big data analytics capability affect partner cooperation decisions? (3) How will the matching of big data analytics capability and cooperative mode affect innovation performance? Based on a game model with two stages, we study a supply chain with one manufacturer and one retailer and focus on firms' cooperative behavior considering big data analytics capability.

2 Literature review

2.1 Big data with innovation

Big data has received considerable attention from academics and practitioners in recent years (Gandomi & Haider, 2015). Many researchers point out that big data makes significant influence on the firm's innovation (Babu et al., 2021; Fosso Wamba & Akter, 2019; Li et al., 2018). Data-driven innovation is regarded as an emerging approach to enhance innovation by acquiring, analyzing and acting upon consumer data (Babu et al., 2021). In the data-driven

supply chain, data as driving forces and raw materials are getting more and more attentions (Liu & Yi, 2018b).

From the perspective of enterprise operation, Big Data has made a positive contribution for firm to manage the new dynamics trends of consumers as well as an analyses of business survival. For example, Gap sends the appropriate localized information in real-time to particular consumers based on the physical location (The Gap Inc., 2019, <https://www.gapinc.com/en-us/articles/2019/09/banana-republic-and-athleta-launch-buyonline,-pic>, <https://www.gapinc.com/en-us/articles/2019/11/%E2%80%98tis-the-season-for-breakoutholiday-campaigns>, <https://www.gapinc.com/en-us/articles/2019/12/how-gap-inc-brands-bring-holiday-cheeraround-the->). Google to deliver targeted advertising (Davenport & Patil, 2012), Bridgestone America uses supply chain data to alert customers to repair stores in a timely manner (Ransbotham et al., 2017). Leveraging open media data can help firms to quickly find new market opportunities, by taking the search log data as an important data source to mine the competitiveness and intensity of the company's products; firms can identify the competitive brands that reflect the intention and cognition of consumers in the market (Wei et al., 2016). In terms of consumer identification and environmental factors of big data, relevant researches based on personalized recommendation improve the accuracy of recommendation and innovate the business recommendation mode by integrating more consumer behavior information (He & Liu, 2017; He et al., 2019). Based on market oriented and technology oriented, the retailers can develop customer analytics capability from crucial themes of marketing, such as value creation (offering capability and personalization capability), value delivery (distribution capability and communication capability), and value management (data management capability and data protection capability), to engage customers and enhance customer equity (Hossain et al., 2020). From the perspective of enterprises coping with the external environment, Big Data can help identify risks along the supply chain (Belhadi et al., 2021). The improvement of IT capability such as big data analytics could help the SMEs to improve their R&D activities and their supply chain system in an unfavourable situation, i.e. post COVID-19 scenario (Chatterjee et al., 2022). These show that Big Data is a kind of the digital innovation technologies that can provide better support for firms in these complex environment (Piccarozzi & Aquilani, 2022).

When firms consider big data technologies to accelerate innovation, it is important to intentionally integrate big data into their business and organizational structures to adapt it to the values and needs that emerge from time to time, rather than focusing on the technologies themselves. On the one hand, data refinement has great importance in utilizing the advantages of big data to produce results that include successful innovation (Boiten, 2016). Big data characteristics (data veracity, data velocity, and data variety) have positive impacts on enhancing data-driven insight generation, which consequently impacts firm innovation competency (i.e., exploitation competency and exploration competency), while data volume does not significantly impact (Ghasemaghahi & Calic, 2019). In the process of Data-driven innovation, sources of algorithmic bias (data bias, method bias and societal bias) can produce detrimental impacts on the outcomes of the data products, which may result in unjust and unfair outcomes, so the decision making autonomy of both humans and machines should both be augmented (Aker et al., 2021). On the other hand, Understanding the effect of data analytics on innovation and how organizational practices may moderate these relationships are especially important (Lynn Wu, 2020). The value created by big data is reflected in the effective transformation of data information into knowledge in the feature database. Compared with the process of big data identification, collection and storage, big data analytics can better reflect the technical tool and resource transformation process of big data generating

value (Akter et al., 2016; Pigni et al., 2016). Therefore, researchers began to use big data analytics capability to indicate the proficiency of firms in using big data to achieve goals and acquire new knowledge (Gupta & George, 2016a). Big data analytics capabilities can lead to enhanced incremental and radical innovation capabilities by affecting the underlying processes of a firm's dynamic capabilities, both the technical and managerial skills are core elements for firms to realize big data success (Patrick Mikalef et al., 2019). In terms of data-driven innovation capability, market orientation capability and innovation talent capability are the two most significant capabilities, followed by infrastructure capability, which means that firms should consider these three key aspects in the innovation process to ensure the whole system performs effectively (Sultana et al., 2022). Among them, talent capability is regarded as a significant distinguishing factor of data-oriented innovation capability, and firms must retain the unique resources (i.e., data, technology) to build unique competencies for innovation (Sultana et al., 2022b). In a data-rich environment, firms can improve marketing analytics capability with the adoption of artificial intelligence, which can help to sense the market, identify market changes and understand customers' expectations, enhances the holistic marketing decision-making, thus improves firms' competitive marketing performance (Rahman et al., 2021). Likewise, service firms enrich their marketing information system management capability could improve their service innovation processes and guide service managers toward innovations that are more in accordance with merging consumer needs (Rahman et al., 2020). In addition, big data analytics capability have a positive effects on business model innovation, including direct impact and indirectly by stimulating firms to proactively take innovative and risky decisions (Ciampi et al., 2021).

2.2 Big data analytics capability

Based on this significant influence, some scholars paid attention to the management and incorporation of big data into innovation, known as big data analytics capability (Lynn Wu, 2020).

The big data analytics capability refers to the ability to gain strategic and operational insights from big data (Akter et al., 2016), that is, the continuous use and deployment of big data resources with the strategic goal of creating value and developing a competitive advantage for the firm (Gupta & George, 2016b, Wamba et al., 2017a). Being enabled by the big data capabilities, firms strive to identify an appropriate and competitive data product to be developed (Sultana et al., 2021).

In general terms, the contribution of big data technology to performance depends on the ability of big data analysis (Yasmin et al., 2020). The important purpose of big data analytics capability is to extract the knowledge that can serve the enterprise product innovation, market demand and gain competitive advantage from the massive and complicated data (Mikalef et al., 2019), which places more emphasis on the data basis for specific knowledge (Xu et al., 2016). Existing studies have analyzed the internal mechanism of big data analytics capability mainly from the resource-based perspective (Yang & Zhou, 2015). Big data analytics capability includes tangible resources and intangible resources. Tangible resources refer to basic resources, technology and data, this requires the firms undertake the necessary investments to advance big data initiatives (Wamba et al., 2017b). Efficient data management requires adequate infrastructure (George et al., 2016), which requires a large initial investment. Intangible resources are indicators to human skills, drive culture and organizational learning, which is often also referred to as knowledge based capital (Chen et al., 2016a, b). Firms cope with

the uncertainty of product innovation activities by continuously acquiring, creating and integrating knowledge to expand knowledge base (Antonelli & Fassio, 2016). New products, services and processes of different forms are generated through the process of knowledge fusion (Ferraris et al., 2019). There is a complementarity effects between investments in tangible and intangible capital (Corrado et al., 2017). In order to improve the big data analytics capability, firms need to coordinate the input of the two resources to achieve the optimal investment returns.

However, analyzing so large and complex data is a huge challenge for most firms. It is difficult for the IT department of the traditional enterprise to use Big Data well because of the “volume” nature of Big Data (Liu et al., 2020). The main challenge related to the use of Big Data, specifically the skills for handling it, has been identified as being of particular concern, as not only are the skills difficult to find but they are, most importantly, expensive to acquire (Del Vecchio et al., 2018). Considering Big Data as an important approach to help firms to maximize their innovation, efficiency (Babu et al., 2021), it is important to seek external cooperation to improve their big data capabilities as well as contribute to their success in promoting collaborative innovation.

2.3 Data collaboration

Generally speaking, innovation cooperation have two different modes: share resource through contract to coordinate their decisions; jointly venture which two firms perfectly share knowledge considered useful to innovation (Ge et al., 2014). The problem of data-based cooperative innovation can also refer to this two modes. According to data problem, the majority of the related studies has focused on data sharing or information sharing. Supply chain management research has long recognized the importance of information sharing between multiple parties (Ghoshal et al., 2020; Stefansson, 2002).

There are several studies from different perspectives. From the perspective of cooperative performance, many researches focus on shared incentives. For example, (Chu et al., 2017) discusses the incentive for information sharing for manufacturers to make simultaneous decisions on capacity and wholesale prices. (Ha et al., 2017) analyzed the incentive problem of vertical information sharing for retailers when manufacturers have cost saving efforts. (Taylor & Xiao, 2009) compared the incentive effects of sales rebate and residual compensation contracts on retailers’ information acquisition behavior. Based on the fixed information acquisition cost. (Tang & Girotra, 2017) studies how suppliers use advance purchase discount contract to motivate retailers’ demand information acquisition and sharing behavior. From the perspective of data value, many studies focus on the behavior of information acquisition. Shin and Tunca (2010) provides a coordination contract based on market value index for retailers’ information investment considering the convexity information acquisition cost in the case of competition among retailers. Guo (2009) studies the impact of information investment and sharing on supply chain performance under the condition of information acquisition costs in two constant states. Chen et al. (2016a, b) studies the contract mechanism of coordinating information investment and sales efforts of retailers simultaneously under the consideration of convexity information acquisition cost. From the perspective of data investment, Liu and Yi (2018c) compared the investment decision and coordination of supply chain on BDI in the case of information symmetry and asymmetry, and adopted the revenue sharing contract to coordinate supply chain. Considering the rise and rapid growth of Data Company, they discussed the investment decision-making problems in a three-stage supply chain with taking Data Company as a member (Liu and Yi, 2018a). These studies all focus on coordination in a

single mode of cooperation, this paper will discuss the cooperation of supply chain members under different cooperation mode.

2.4 Cooperative innovation in supply chain

Cooperative innovation is a form of contractual to obtain external resources for joint R&D (Fusfeld & Haklisch, 1985), to achieve tacit knowledge transfer between firms; reduce innovation costs and avoid innovation risks (Williams, 2005). The research content of supply chain cooperative innovation mainly focuses on cooperative innovation mechanism, cooperative performance and influence factors.

In terms of influencing factors of supply chain cooperative innovation, Hsueh et al., (2010) found that the participation of other cooperative organizations would have a significant impact on innovation performance, and innovation performance would become higher with stronger network embeddedness. Trigo and Vence (2012) shows that innovation level is positively correlated with cooperation level through empirical research, and cooperation can also promote enterprise innovation level. Wu (2014) believes that with the fierce competition, more and more enterprises take the initiative to cooperate with other organizations, even competitors, to form an innovation network system. Skippari et al. (2017) studies the factors of cognitive barriers that supply chain members will face in the process of cooperative innovation, and puts forward that the generation of cooperative innovation will be affected by the different views of the relationship between supply chain members.

In terms of the innovation performance of supply chain cooperation, Bellantuono et al. (2009) analyzed the profit distribution problem of the two-level supply chain cooperation and found that the retailer's profit when cooperating with suppliers was greater than when acting alone. Bai and Sarkis (2016) studies different developments of suppliers, and the results show that cooperative and non-cooperative decisions between manufacturers and suppliers have a direct impact on supplier investment. Hu et al. (2016) believes that the internal and external integration of the supply chain can promote the exploration and development of innovative knowledge, and the supply chain thus obtains complementary resources and technologies, which improves its competitive advantages and drives the research and development of new products. Friedl and Wagner (2016) designed a contract model of supply chain innovation composed of suppliers and manufacturers, and found that the optimal value of supply chain can be obtained when cooperation stimulates supplier innovation.

In terms of cooperation mechanism, it is usually to integrate with external resources, but in the process of cooperation, difficulties in knowledge integration and risks such as knowledge leakage will inevitably be encountered. Therefore, how to cooperate and what degree of openness become key decisions in this process (Laursen & Salter, 2014). Baldwin et al. (2006) compared the costs and expected benefits of various product innovations and found that the mode of cooperative innovation in which manufacturing enterprises play the leading role and other related enterprises cooperate can not only improve market satisfaction, but also significantly improve the benefits obtained from product innovation. Knight et al. (2016) proposed that cooperative innovation with upstream and downstream enterprises of the supply chain can enhance the internal and external expansion effect of innovation and stimulate the amount of enterprise resources input in the process of innovation, so as to maintain the competitive advantage of enterprises.

In summary, different cooperation modes will affect the final innovation performance, and the big data analytics capability will affect the choice of cooperation. The loose mode pays more attention to collaboration, the data risk of the firms is relatively lower, the contribution

and benefit distribution of both parties are clearly defined by contract, and the data analysis activities are still within the organizational boundaries. Data property rights are clear, each still has its own data advantage, but data separation makes it difficult to maximize the value of data. The tight mode is more creative, and the degree of interaction of data is much higher than the loose mode, which can maximize the value of data. However, in the tight mode, the data risk of both parties is the highest, and it is easy to lose the unique advantage of data resources. Previous studies have not discussed the correlation between big data analytics capability and cooperation mode decisions. Therefore, this paper will explore the relationship between big data analytics capability and cooperation mode, and then analyze the influence mechanism of their interaction relationship on innovation performance.

3 Problem description and model establishment

In the big data environment, enterprise designs and produces product based on the accurate and timely consumer preference information, to meet consumer demand. In order to focus on the different cooperation mode, we consider a simple supply chain consisting of an upstream manufacturer (denoted as “m”) and a downstream retailer (denoted as “r”). The actual activities of developing, designing and producing new products are carried out by the manufacturer. The retailer collect customer information in the process of selling products. Cooperative innovation means that the upstream firm invests in the big data analytics capability and the downstream firm invests in the big data information collection, the two sides cooperate through big data. Given that our research focuses on data analytics capabilities, we assume that manufacturers already have a certain amount of consumer data, have sufficient capacity to utilize the acquired data, and do not invest in additional data acquisition. The retailer is closer to the consumer end, so it is the responsibility of the retailer to obtain and provide the additional data needed for innovation activities. Data quality is used to measure retailers’ big data contribution to the collaboration process. Before presenting our model, we introduce some notations as Table 1.

Table 1 Parameters and their definitions

Parameter	Definition
p	The product retail price
w	The wholesale price of the product
c	The average unit cost of production
c_t	The average unit cost of big data analytics capability
c_d	The average unit cost of big data collection
α	The value discount factor which influenced by the precision of consumer preference information
β	The conversion coefficient of preference information extracted from consumer data
π_i	The profits of the firms, $i = m, r$
T	Index of cooperation type, $T = L, T$

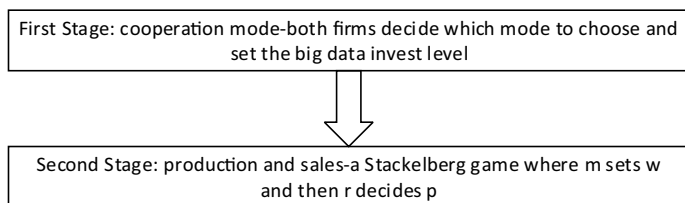


Fig. 1 The sequence of actions

Based on the utility function theory, let $U = \alpha v - p$, $v \in [0, 1]$. We get the market demand formula to the following one (Liu and Yi, 2018a):

$$Q = 1 - p/\alpha \quad (1)$$

Assume that supply chain can get the total number of consumer information is D . Q/D represents the degree of consumer preference information conversion. We can get $D = \beta Q$, $\beta > 1$, which means that the data quality improves with the decrease of β . α is determined by the precision of consumer preference information, stand for the level of big data analytics capability.

Without loss of generality, we assume that the manufacturer has enough production capacity. Supply chain members are completely rational and risk neutral.

In the two cooperation models, the manufacturer acts as the leader and the retailer acts as the follower in a Stackelberg game. This power structure is common in manufacture industry (Ge et al., 2014). The decision sequence is as Fig. 1. In stage 1, the manufacturer with retailer negotiate to decide the cooperation mode. This will influence the total big data invest in innovation activities. In stage 2, the manufacturer sets the wholesale price. And then, the retailer decides the product retail price. Let's first discuss the optimal decision under different modes.

3.1 Tight cooperation model (T)

We first explore firms' decisions and corresponding profits in tight mode, which is a benchmark. In this model, manufacturers and retailers build data centers to support innovation and share the cost of big data proportionately (k). Data resources and inputs of both parties are shared to the greatest extent. New products developed under this scenario can meet the needs of consumers' better than other situation. The optimization problem of the manufacturer and retailer are, respectively,

$$\max_w \pi_m^T = \left(w - c - k \times \left(c_t + \beta^T \times c_d^T \right) \right) \times Q \quad (2)$$

$$\max_p \pi_r^T = \left(p - w - (1 - k) \times \left(c_t + \beta^T \times c_d^T \right) \right) \times Q \quad (3)$$

We can use backward induction to solve it, get the optimal decision:

$$\left(w^T, p^T \right) = \left(\frac{1}{2} \left(c + \alpha + (2k - 1) \left(\beta^T c_d^T + c_t \right) \right), \frac{1}{4} \left(c + 3\alpha + \beta^T c_d^T + c_t \right) \right) \quad (4)$$

And the supply chain members' profits are:

$$(\pi_m^T, \pi_r^T) = \left(\frac{(\alpha - c - \beta^T c_d^T - c_t)^2}{8\alpha}, \frac{(\alpha - c - \beta^T c_d^T - c_t)^2}{16\alpha} \right) \quad (5)$$

Proposition 1 *In the tight model, (w^T, p^T) is the equilibrium solution. (π_m^T, π_r^T) is the optimal profits of supply chain members.*

It is worth noting that both firms should consider the data quality (β) and big data analytics capability (α) when making optimal decisions. In addition, the cost sharing ratio of the data center has no effect on the optimal profit of both parties. In other words, costs are transferred internally through wholesale prices in this model, the two parties obtain a certain profit ratio in the cooperation, namely $\pi_m^T = 2\pi_r^T$.

Corollary 1 *For the equilibrium solution, we have:*

- (1) $\alpha > c + \beta^T c_d^T + c_t$
- (2) $\frac{\partial \pi^T}{\partial \alpha} > 0$ and $\frac{\partial \pi^T}{\partial \beta^T} < 0$

Corollary 1 shows that the value discount factor should be greater than the total cost per unit product. Intuitively, with an increased big data analytics capability and data quality, both the supply chain members could get more benefits.

3.2 Loose cooperation model (L)

Under the loose model, members take big data activities separately. The retailer will use big data technology to collect consumer data, and share the data and customer information extracted during the collection process with the manufacturer. The manufacturer mainly invest in big data analytics capability, and conduct customer information mining based on the existing customer database and the data shared by retailers. Obviously, the accuracy of consumer preference information is lower than it in the tight model. The firms' optimization problems are:

$$\max_w \pi_m^L = (w - c - c_t - \beta^L \times c_{d1}^L) \times Q \quad (6)$$

$$\max_p \pi_r^L = (p - w - \beta^L \times c_{d2}^L) \times Q \quad (7)$$

Considering the data format mismatch, security and so on in the process of sharing, data quality in T model should be higher than L model, which means more data and higher costs for the same amount of sales quantity ($\beta^T (\beta^L; c_{d1}^L + c_{d2}^L) c_d^T$). Similar to Model T, we solve (6) and (7) using backward induction. And we get the equilibrium solution:

$$(w^L, p^L) = \left(\frac{1}{2} (c + \alpha + \beta^L (c_{d1}^L - c_{d2}^L) + c_t), \frac{1}{4} (c + 3\alpha + \beta^L (c_{d1}^L + c_{d2}^L) + c_t) \right) \quad (8)$$

And the supply chain members' profits are:

$$(\pi_m^L, \pi_r^L) = \left(\frac{(\alpha - c - \beta^L (c_{d1}^L + c_{d2}^L) - c_t)^2}{8\alpha}, \frac{(\alpha - c - \beta^L (c_{d1}^L + c_{d2}^L) - c_t)^2}{16\alpha} \right) \quad (9)$$

Proposition 2 *In the loose model, (w^L, p^L) is the equilibrium solution. (π_m^L, π_r^L) is the optimal profits of supply chain members.*

Note that in this model, manufacturers have difficulty controlling the quality of data shared by retailers, and this is also the initial stage of data-based collaboration. In order to motivate retailers to improve the quality of shared data and enhance their willingness to cooperate, manufacturers usually use contracts to motivate retailers. Next, we discuss the two incentive contracts under this model.

3.3 Loose model with data-subsidy contract (LS)

Generally speaking, the data cost subsidy comes in two forms, one is based on the amount of data shared; another one is based on the effective amount of information converted by data. To incentivize retailers to improve the quality of their data, manufacturers will subsidize the data per unit of new product sales. Under the subsidy contract, the retailer improve data quality and recoup data costs through manufacturer subsidies. Set the cost subsidy rate is h , $h \in (0, 1)$. The firms' optimization problems are:

$$\max_w \pi_m^{LS} = (w - c - c_t - \beta^{LS} \times c_{d1}^L - h \times c_{d2}^L) \times Q \quad (10)$$

$$\max_p \pi_r^{LS} = (p - w - \beta^{LS} \times c_{d2}^L + h \times c_{d2}^L) \times Q \quad (11)$$

Considering retailers have improved the quality of their data, and then $\beta^T < \beta^{LS} < \beta^L$. Similar to the solution process of the above model, we can get the equilibrium solution:

$$(w^{LS}, p^{LS}) = \left(\frac{\frac{1}{2}(c + \alpha + \beta^{LS}c_{d1}^L - (\beta - h)c_{d2}^L) + c_t}{\frac{1}{4}(c + 3\alpha + \beta^{LS}c_{d1}^L + (\beta - h)c_{d2}^L + c_t)}, \right) \quad (12)$$

And the supply chain members' profits are:

$$(\pi_m^{LS}, \pi_r^{LS}) = \left(\frac{(\alpha - c - \beta^{LS}(c_{d1}^L + c_{d2}^L) - c_t)^2 - h^2(c_{d2}^L)^2}{\frac{8\alpha}{(\alpha - c - \beta^{LS}(c_{d1}^L + c_{d2}^L) - c_t + hc_{d2}^L)^2}}, \right) \quad (13)$$

Proposition 3 In the LS model, (w^{LS}, p^{LS}) is the equilibrium solution. (π_m^{LS}, π_r^{LS}) is the optimal profits of supply chain members.

This is a special case between mode (L) and model (T) with different h . According to formulas (13), we get the subsidy rate needs to meet the following condition— $h \in (0, \frac{\alpha - c - \beta^{LS}(c_{d1}^L + c_{d2}^L) - c_t}{c_{d2}^L})$, to ensure that both profits are non-negative.

The equilibrium decisions and profits of mode (T), (L) and (LS) are listed in Table 2.

As can be seen from Table 2, the price (wholesale price and unit price) is gradually reduced, and the profit of each enterprise is gradually increased, with the cooperation mode shifts from loose to close. This result indicates that big data-based cooperation of supply chain is conducive to improving innovation performance, which is consistent with most relevant research results (Tan et al., 2015, Liu & Yi, 2018c). In addition, it can be found that the cooperation performance of LS mode between tight and loose is not necessarily between the two, depending on data quality, data cost and contract ratio. This result is different from previous studies, the intermediate mode is not necessarily a transitional mode between tight and loose, and cooperation may directly change from loose to tight. In contrast to data quality, the change of big data analytics capability does not have a continuous impact on the choice of cooperation mode, but there may be a critical point. Based on the optimal decision of each mode, we discuss each factor further in the next section.

Table 2 The optimal decisions and corresponding profits in different modes

Mode	T	L	LS
w	$\frac{(c+\alpha+(2k-1)(\beta^T c_d^T - c_l))}{2}$	$\frac{(c+\alpha+\beta^L(c_{d1}^L - c_{d2}^L) + c_l)}{2}$	$\frac{(c+\alpha+\beta^{LS}c_{d1}^L - (\beta-h)c_{d2}^L) + c_l}{2}$
p	$\frac{(c+3\alpha+\beta^T c_d^T - c_l)}{4}$	$\frac{(c+3\alpha+\beta^L(c_{d1}^L + c_{d2}^L) + c_l)}{4}$	$\frac{(c+3\alpha+\beta^{LS}c_{d1}^L + (\beta-h)c_{d2}^L) + c_l}{4}$
π_m	$\frac{(\alpha - c - \beta^T c_d^T - c_l)^2}{8\alpha}$	$\frac{(\alpha - c - \beta^L(c_{d1}^L + c_{d2}^L) - c_l)^2}{8\alpha}$	$\frac{(\alpha - c - \beta^{LS}(c_{d1}^L + c_{d2}^L) - c_l)^2 - h^2(c_{d2}^L)^2}{8\alpha}$
π_r	$\frac{(\alpha - c - \beta^T c_d^T - c_l)^2}{16\alpha}$	$\frac{(\alpha - c - \beta^L(c_{d1}^L + c_{d2}^L) - c_l)^2}{16\alpha}$	$\frac{(\alpha - c - \beta^{LS}(c_{d1}^L + c_{d2}^L) - c_l + hc_{d2}^L)^2}{16\alpha}$
π_{sc}	$\frac{3(\alpha - c - \beta^T c_d^T - c_l)^2}{16\alpha}$	$\frac{3(\alpha - c - \beta^L(c_{d1}^L + c_{d2}^L) - c_l)^2}{16\alpha}$	$\frac{(c - \alpha + \beta c_{d1}^L + (\beta - h)c_{d2}^L + c_l)(3\beta c_{d1}^L + (h+3\beta)c_{d2}^L + 3(c - \alpha + c_l))}{16\alpha}$

4 Effects of ways of cooperation on profits

In this section, we address whether firms have incentives to get tight mode and whether they have intentions to improve their data capability.

4.1 Effects of modes

Knowing whether tight mode can improve firms' profits can help managers make better cooperation mode decisions. This subsection discusses whether or not optimal cooperation mode is exist, and how to get this mode.

Proposition 4 *For the chains and two members' profit, $\pi^T \geq \pi^L$ always holds.*

The proposition above shows that mode (T) is always better than mode (L) and mode (LS). This is similar to the process of innovation cooperation in reality. Generally, the tight mode appears among firms who have a long history of cooperation or a close vertical relationship (Ge et al., 2014). The cooperation will gradually change from mode (L) to mode (T) through the running-in of mode (LS). In order to discuss the transformation of the cooperation mode, we divided the discussion into two situations.

(1) Mode (L) transfer to Mode (LS)

Assume that $\Delta\pi_i$ stands for the profits differences between modes for firm i , $i = m, r$. $\Delta\pi_{i1} = \pi_i^{LS} - \pi_i^L > 0$, for given $\beta^{LS} = \beta^L = \beta$, and according to formulas (9) and (13), we get the $\Delta\pi_{i1} > 0$ always hold for $h \in (0, \frac{\alpha - c - \beta(c_{d1}^L + c_{d2}^L) - c_t}{c_{d2}^L})$. It expresses that when $0 < h < \frac{\alpha - c - \beta^{LS}(c_{d1}^L + c_{d2}^L) - c_t}{c_{d2}^L}$ can be met, both firms are more active to participate in the cooperation mode (LS). We also get $\frac{\partial \Delta\pi_{m1}}{\partial h} = -2hc_{d2}^{L2} < 0$; $\frac{\partial \Delta\pi_{r1}}{\partial h} = 2c_{d2}^L(\alpha - c - \beta(c_{d1}^L + c_{d2}^L) - c_t + hc_{d2}^L) > 0$.

Corollary 2 *When the cost subsidy rate h less than $\frac{\alpha - c - \beta^{LS}(c_{d1}^L + c_{d2}^L) - c_t}{c_{d2}^L}$, the manufacturer and the retailer can get more benefits from mode (LS). Moreover, the retailer can get more benefits from the transition mode (L) to mode (LS) than the manufacturer.*

(2) Mode (LS) transfer to Mode (T)

For given $\beta^T = \beta^{LS} = \beta$, $\Delta\pi_{i2} = \pi_i^T - \pi_i^{LS} > 0$, according to formulas (5) and (13), we get $\Delta\pi_{i2} > 0$ always hold when $0 < h < \min(\frac{\sqrt{(2(\alpha - c - c_t) - \beta^{LS}(c_{d1}^L + c_{d2}^L) - \beta c_d^T)(c_{d1}^L + c_{d2}^L - c_d^T)}}{c_{d2}^L}, \frac{\alpha - c - \beta^{LS}(c_{d1}^L + c_{d2}^L) - c_t}{c_{d2}^L})$ can be met. And we also get $\frac{\partial \Delta\pi_{m2}}{\partial h} = 2hc_{d2}^{L2} > 0$; $\frac{\partial \Delta\pi_{r2}}{\partial h} = -2c_{d2}^L(\alpha - c - \beta(c_{d1}^L + c_{d2}^L) - c_t + hc_{d2}^L) < 0$.

Corollary 3 *When the $0 < h < \min(\frac{\sqrt{(2(\alpha - c - c_t) - \beta^{LS}(c_{d1}^L + c_{d2}^L) - \beta c_d^T)(c_{d1}^L + c_{d2}^L - c_d^T)}}{c_{d2}^L}, \frac{\alpha - c - \beta^{LS}(c_{d1}^L + c_{d2}^L) - c_t}{c_{d2}^L})$ can be met, both the manufacturer and the retailer can get more benefits from mode (T). Moreover, the manufacturer can get more benefits from the transition mode (LS) to mode (T) than the retailer.*

The above analysis indicates that both firms can achieve larger profits through tight cooperation mode. Loose mode performed the worst relative to others. For the manufacturer,

mode (T) brings more revenue growth than mode (LS). This is because the manufacturer spends more data cost under mode (LS), so the marginal rate of return of big data decreases relatively. For retailers, mode (LS) brings more revenue growth than mode (T). Under mode T, retailers share more data costs for manufacturers, and their marginal rate of return on big data decreases relatively. Therefore, manufacturers and retailers can negotiate the appropriate subsidy ratio to reach the optimal cooperation mode. It can be found that the most possible compromise between the supply chain members is to form the CONTRACT mode of LS first, and then change from LS to T or L. Consistent with the conclusion of previous cooperation innovation studies, both parties usually carry out cooperation in contract mode to ensure the smooth progress of cooperation. However, in the case of big data cooperation, the different result is that LS mode may go further form a tight mode, or may fall back to a loose mode, even if the partnership is formed through long-term contractual cooperation.

4.2 Effects of big data

(1) The effect of big data analytics capability

Given the $\beta^T = \beta^{LS} = \beta^L = \beta = 1.2$. For the manufacturer, we can get $\frac{\partial \Delta \pi_{m1}}{\partial \alpha} = \frac{\partial (\pi_m^{LS} - \pi_m^L)}{\partial \alpha} = \frac{h^2 c_{d2}^L}{8\alpha^2} > 0$; $\frac{\partial \Delta \pi_{m2}}{\partial \alpha} = \frac{\partial (\pi_m^T - \pi_m^{LS})}{\partial \alpha} = \frac{(-h^2 + \beta^2)c_{d2}^L + 2\beta c_{d2}^L(c + \beta c_{d1}^L + c_i) + \beta(-c_d^T + c_{d1}^L)(\beta(c_d^T + c_{d1}^L) + 2(c + c_i))}{8\alpha^2} < 0$; and $\frac{\partial (\pi_m^T - \pi_m^L)}{\partial \alpha} = \frac{-\beta(c_d^T - c_{d1}^L - c_{d2}^L)(\beta(c_d^T + c_{d1}^L + c_{d2}^L) + 2(c + c_i))}{8\alpha^2} < 0$.

The results show that under the condition of fixed data quality, the increase of big data analytics capability input can improve the profit of the manufacturer, but the profit growth rate from mode (L) or mode (LS) to mode (T) gradually decreases. Only increasing the analytical capability from mode (L) to (LS) can increase the rate of profit growth. This shows that when the manufacturer encourage retailers to participate in the mode (LS), improving the analytical ability is conducive to improving the profit growth.

For the retailer, we can get $\frac{\partial \Delta \pi_{r1}}{\partial \alpha} = \frac{\partial (\pi_r^{LS} - \pi_{mr}^L)}{\partial \alpha} = \frac{hc_{d2}^L((2\beta - h)c_{d2}^L + 2(c + \beta c_{d1}^L + c_i))}{16\alpha^2} > 0$; $\frac{\partial \Delta \pi_{r2}}{\partial \alpha} = \frac{\partial (\pi_r^T - \pi_r^{LS})}{\partial \alpha} = \frac{-(\beta(c_d^T - c_{d1}^L - c_{d2}^L) + hc_{d2}^L)(\beta c_d^T + \beta c_{d1}^L - (\beta - h)c_{d2}^L + 2(c + c_i))}{16\alpha^2} < 0$; and $\frac{\partial (\pi_r^T - \pi_r^L)}{\partial \alpha} = \frac{-\beta(c_d^T - c_{d1}^L - c_{d2}^L)(\beta(c_d^T + c_{d1}^L + c_{d2}^L) + 2(c + c_i))}{16\alpha^2} < 0$. The same results as the manufacturer.

Proposition 5 With the increase of big data analytics capability, the enthusiasm of firm's participation mode (LS) is relatively increased, while the enthusiasm of participation mode (L) and mode (T) is relatively decreased.

This proposition indicates that big data analytics capability has a greater influence on the transition from loose mode to tight mode. This is because the ability is invested by the manufacturer independently. Under the loose mode, the manufacturer increase investment in analytical capability to improve overall innovation performance and motivate the retailer. When the cooperation between the two parties reaches a certain level (such as mode LS), the influence of the big data analytics capability on the decision-making of the cooperation mode is gradually weakened. That is, it is most effective for manufacturers to improve their data analytical capabilities as they transfer from Mode (L) to Mode (LS).

(2) The effect of data quality

In the situation form mode (L) transfer to mode (LS), given the α , set $\beta^L = \beta^{LS}$. For the manufacturer, we can get $\frac{\partial \Delta \pi_{m1}}{\partial \beta^L} = \frac{\partial (\pi_m^{LS} - \pi_m^L)}{\partial \beta^L} = 0$. For the retailer, we have

$\frac{\partial \Delta \pi_{r1}}{\partial \beta^L} = \frac{\partial (\pi_r^{LS} - \pi_{mr}^L)}{\partial \beta^L} = -\frac{hc_{d2}^L(c_{d1}^L + c_{d2}^L)}{8\alpha} < 0$. At this time, data quality has little influence on the manufacturer's decision-making, but has a positive influence on the retailer's cooperation enthusiasm, that is, with the improvement of data quality, the retailer are more inclined to shift from mode (L) to mode (LS).

In the situation form mode (LS) transfer to mode (T), set $\beta^T = \beta^{LS}$, we can get $\frac{\partial \Delta \pi_{m2}}{\partial \beta^T} = \frac{\partial (\pi_m^T - \pi_m^{LS})}{\partial \beta^T} = -\frac{(c_d^T - c_{d1}^L - c_{d2}^L)(\alpha - c - \beta^T(c_d^T + c_{d1}^L + c_{d2}^L) - c_t)}{4\alpha} 0$, $\frac{\partial \Delta \pi_{r2}}{\partial \beta^T} = \frac{\partial (\pi_r^T - \pi_r^{LS})}{\partial \beta^T} = \frac{\beta^T c_d^{2T} + c_d^T(c - \alpha + c_t) - (c_{d1}^L + c_{d2}^L)(c - \alpha + \beta^T c_{d1}^L + (\beta^T - h)c_{d2}^L + c_t)}{8\alpha} 0$. For the manufacturer, $\Delta \pi_{m2}$ is a convex function of data quality. When the $\beta^T = \frac{c_d^{2T} - (c_{d1}^L + c_{d2}^L)^2}{4\alpha}$, the profit increment brought by improving data quality is the highest. For the retailer, profit growth is influenced by both data quality and subsidy rate.

Proposition 6 *For data quality, there is an optimal quality level that allows manufacturers to have extreme points in the mode decisions. Under this quality level, there is a certain threshold of subsidy rate, which makes the retailer more inclined to choose mode (T).*

This proposition indicates that the manufacturer can motivate the retailer to reach the optimal quality level through appropriate subsidy contract, and both parties have the strongest willingness to participate in mode (T) at this time. It can be seen that externally shared data quality is more important for manufacturers, which is consistent with the conclusion that “enterprise innovation requires new external knowledge, thus weakening the need for internal knowledge combination (Cheng et al., 2016, Xu et al., 2016)”. For retailers, manufacturers' big data analytics capability has a greater impact on their choice of contractual cooperation and a smaller impact on their choice of further tight mode. Retailers' collaborative decisions are more focused on the cost of sharing data, so the balance between the two firms is how manufacturers compensate retailers for the loss of shared data.

5 Numerical simulation

In this section, a numerical example is presented to show the results' effectiveness. To ensure more realistic, we select the following parameter settings according to (Liu and Yi, 2018a) which are summarized in Table 3.

In order to demonstrate the influence of the big data capability on supply chain cooperation decision-making, we vary the value of α and β . Based on Propositions 1 to 3, we get $(\pi_m^T, \pi_r^T, \pi_m^L, \pi_r^L) = (4.79; 2.39; 3.34; 1.67)$. Based on the above analyses, we get $\Delta \pi_m > 0$ and $\Delta \pi_r > 0$. Thus, proposition 4 is verified.

The effects of the big data analytics capability on the profits differentials among different mode are shown in Fig. 2. From Fig. 2, we can get that with the increase of big data analytics capability (α), the increase of data analysis input can improve the profit of the members. It indicates that if the manufacturer want to gain more benefits, they need to improve the

Table 3 Parameter settings for numerical analysis

Parameters	Data selection
α	$\alpha^T = 50; \alpha^L = 40$
β	$\beta^T = 1.2; \beta^L = 1.5$
C	$c_t = 0.5; c_d = 1; c_{d1}^L = c_{d2}^L = 0.6; c = 5$

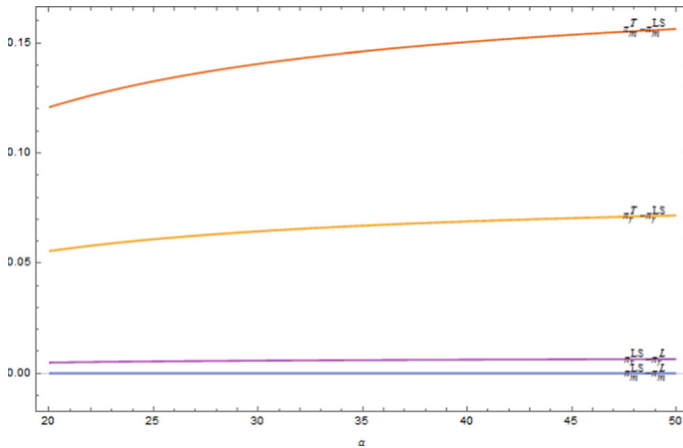


Fig. 2 Effects of α on the profits difference of members under modes

big data analytics capability. In addition, manufacturers' profit growth shift from easy to tight mode is high, while retailers' profit growth value is low. The profit increment obtained from transform mode LS to T is higher than that obtained by transform mode L to LS. It indicates that the higher the level of big data analytics capability, the easier it is to form tight cooperation.

The effect of data quality on the profits differentials among mode are shown in Fig. 3. From Fig. 3, we can get that data quality have a positive relationship with the choice of cooperation mode. Moreover, with the rise of big data capability, the gaps between firms' innovative performance under loose and tight mode will increase. Thus, Propositions 4–5 are confirmed.

The effects of the subsidy rate on the profits differentials among different mode are shown in Fig. 4. We set the big data capability as then should be in each mode. From Fig. 4, we can get that the data cost subsidy ratio has no effect on manufacturers, but it has a big impact on retailers' decision-making. When the cost subsidy rate is at the extreme value (approaching

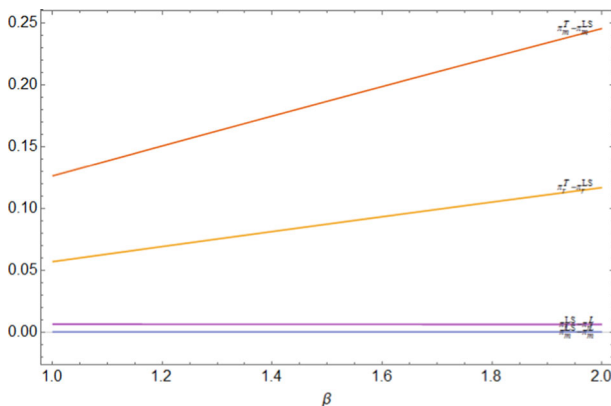


Fig. 3 Effects of β on the profits difference of members under modes

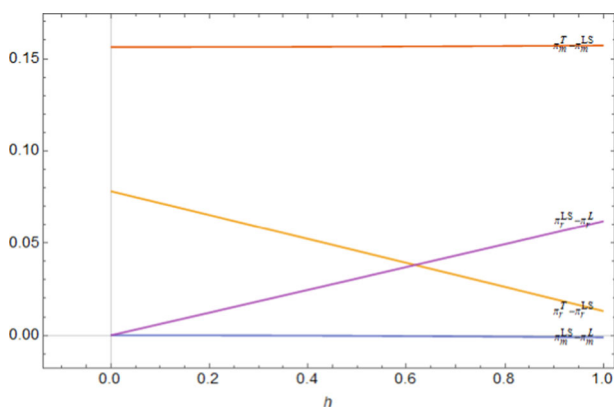


Fig. 4 Effects of h on the profits difference of members under modes

0 or 1), the retailer's mode profit difference is greatest. When the cost subsidy rate is in a low region, retailers are more willing to participate in close cooperation to obtain more cooperation benefits. This is because the subsidy alone is not enough to make up for the retailer's total data input, and the tight model can help him gain more from the partnership. When the cost subsidy rate is in a high region, the cost subsidy is enough to make up for the big data input of retailers, and their enthusiasm to participate in the tight mode decreases accordingly. Proposition 6 are confirmed.

6 Discussion

Massive data utilization makes significant influence on the firm's innovation. Due to the different roles of firms in the supply chain, upstream and downstream obtain the data value advantage from different sources, which leads to the different innovation resource. Customer data in the same supply chain has certain commonality, but the data format, data storage and other data problems make it difficult for firms to directly share data or technology. The high time requirement of big data processing and the small proportion of valuable data, make firms have to cooperate on data to improve the innovation performance. However, data sharing loses its unique right to data, and the mode of data cooperation between firms also affects the efficiency of data analysis. Data risk, security unguaranteed, data control and validation issues that make most firms tend to choose loose cooperation type, which is usually represented as contract to selectively share data or knowledge. Therefore, the selection of an appropriate mode for cooperation innovation should be based on the particular big data capability of the firms.

- (1) Collaborative innovation based on big data significantly improves innovation performance, regardless of the cooperation mode. Previous studies have shown that cooperation innovation is conducive to improving innovation performance (Trigo & Vence, 2012, Hu et al., 2016), and the big data is also beneficial to the firm's innovation (Babu et al., 2021; Fosso Wamba & Akter, 2019; Li et al., 2018). This study confirms that data cooperation is mutually beneficial for supply chain members, and that no matter what kind of collaboration mode can benefit from data-driven innovation. As the

core component of data-driven innovation, big data analytics capability have a significant impact on cooperation. With the improvement of big data analytics capability, each supply chain member can gain more benefits.

- (2) Tight partnership are always better than others, and the efficiency of big data on innovation performance is the lowest under loose mode. This may be due to the different data values obtained from the same batch of data in different modes. Combined with the research questions, this paper draws on the the research hypothesis of Ge et al. (2014) and Ha et al. (2017) on cooperation, divides the cooperation mode into two extreme cases, and discusses the impact of cost subsidy contract on the transition of cooperation mode. On this basis, it is further found that there is a critical value of the big data analytics capability which influence on cooperative mode decision-making. Innovation returns can be significantly improved by proper contracts, but a high proportion of cost subsidy will inhibit the enthusiasm of retailers to participate in tight cooperation. Big data analysis capability is the index to measure the value output of data, and this key threshold is the balance point for both partners to choose a tight partnership.
- (3) The big data analytics capability and data quality have positive effects on the optimal decision. This is consistent with previous studies on big data and decision-making process. High levels of learning capacity enable the combination and validation of knowledge extracted from big data, rendering informed decision-making process (Ghasemaghaei, 2019). This paper draws on the the research results of Yasmin et al. (2020) and Xu et al. (2016) on big data, focuses on the impact of data analytics capability and data quality on cooperative mode decision making. The results show that the enhancement of big data analytics capability will exacerbate the benefits gap between different modes. This is because big data analytics capability not only provide the knowledge base for the innovation process (Mikalef et al., 2019), but also improve the enterprise's knowledge integration (Xu et al., 2016) and management capabilities (Ferraris et al., 2019). This difference in performance is reflected in the efficiency of transforming big data into valuable knowledge. For the supply chain members, trying their best to extract the value of Big Data will help them increase marginal rate of return on big data. For the manufacturer, improving the big data analytics capability will help them gain more benefits, in addition, keeping proper data cost subsidy can better encourage retailers to participate in tight cooperation.
- (4) Data quality has an even bigger impact on manufacturers. For retailers, it is more advantageous to maintain the contract mode if the big data analysis capability of the partner is below the threshold. However, the benefits of tight mode are better than other modes, so retailers can motivate partners to improve their big data analysis capability by improving data quality. Manufacturers improve their data analytics capability can promote cooperation transfer from loose to tight. Retailers provide high-quality data can promote manufacturers to improve their analytical capabilities. The improvement of big data analytics capability and data quality will be mutually reinforcing.

6.1 Theoretical contribution

The main theoretical contribution of this study is that: (1) Previous studies have emphasized that big data analytics capability can help firms improve innovation performance (Babu et al., 2021; Fosso Wamba & Akter, 2019; Li et al., 2018), and analyzed the process of big data analytics capability assisted decision-making (He & Liu, 2017; He et al., 2019; Mikalef et al., 2019, Yasmin et al., 2020). However, existing studies still limit the commercial value of big

data analytics capability to the firm's operation process, while ignoring the impact of big data analytics capability on the cooperation process. This study combines big data analytics capability with cooperative innovation mode. It not only effectively supplements the current research on the aftereffects of big data analytics capability, but also deepens the interpretation of the current innovation management theory on the formation of cooperative innovation process from the perspective of big data analytics capability. (2) This finding enriches the data-driven innovation literature by showing that cooperation to expand data resources and gain unique data advantages can help enterprises to carry out innovation activities efficiently, so as to achieving longer-lasting innovation advantages, especially in unfavourable environment. This set of unique data resources makes it possible to build unique competencies for innovation, and it is the combination of these resources that will enable firms to develop big data analytics capability and realize value gains (Patrick Mikalef et al., 2019). It needs the firms to nurture big data analytics capability by specifically investing in the basic capabilities, i.e. talent capability (Sultana et al., 2022b), infrastructure capability (Sultana et al., 2022a), data resource (Ghasemaghaei and Calic, 2019) and so on. In particular, this study find that cooperation can promote the improvement of big data analytics capabilities of both sides, which will be mutually reinforcing each other. (3) In the theory of cooperative innovation, previous studies have discussed the influence of different factors on the cooperation from the perspectives of cooperation object (Hsueh et al., 2010, Skippari et al., 2017), cooperation mode (Ge et al., 2014) and cooperation performance (Hu et al., 2016, Bai and Sarkis, 2016). A few studies have discussed the impact of big data analytics capability on cooperative performance (2018a; Liu & Yi, 2018b), but the impact of big data analytics capability on the formation process of cooperation is still blank. This study introduces the cooperation mode in the relationship between big data analytics capability and cooperative innovation, further refines the influence mechanism of big data analytics capability on supply chain cooperation from the perspective of the relationship between supply chain members in the formation of cooperation, and enriches the research on the role of big data analytics capability and cooperative innovation.

6.2 Practical contribution

According to the conclusions, this study has some practical implications. (1) Firms should focus on fostering big data analytics capability. On the one hand, firms should strengthen the construction of big data infrastructure and recruit analysis technical personnel to improve the value conversion rate of big data, so as to maintain the advantageous position of firms in the cooperative relationships. On the other hand, firms should actively accumulate data in each cooperation process, so as to increase their own data reserves, which will be an important resource for firms to make independent innovation decisions. (2) Firms should give priority to partners' analytics capability when choosing cooperation modes. On the one hand, the firm needs to consider whether the big data analytics capability of the partner is strong enough, to determine whether to lead the innovation activities. On the other hand, the firm needs to consider their own big data analytics capability, to determine the cooperation mode. it is easiest to form a tight partnership when the supply chain member's big data analytics capability are matched and coordinated. (3) Firms can choose contract mode to start the cooperation. The subsidy contract can still coordinate the supply chain. When the capability is not matched in the cooperation, both parties can be encouraged to increase the investment in big data analytics capability through incentive contract, so as to improve the overall big data analytics capability and then shift to the tight mode. It can also reduce incentives, reduce

costs, maintain loose cooperation to exchange data resources, and break down cooperation when resource exchange is saturated.

7 Conclusion, limitations and future directions

7.1 Conclusion

In this paper, we constructs a supply chain consisting of an upstream manufacturer and a downstream retailer, and establishes three analytical models: data analytics is implemented individually (the loose mode) by either firm, or jointly (the tight mode) by both firms, and the addition of coordination contracts under the loose mode, to studied the impacts of the big data analytics capability on the decision-making of cooperation innovation mode. We also identify the optimal cooperation mode for the manufacturer and the retailer. Based on our analytical results and numerical examples, we have the following conclusions.

- (1) How will the change of big data analytics capability affect firms' choice of cooperative mode? Our results show that both firms' big data analytics capability and data quality both have positive effects on the selection of tight cooperation mode. With the enhancement of big data analytics capability, the firms' innovative performance gaps between loose and tight mode will increase, and the enthusiasm of firm's participation in tight mode is relatively increased. Data quality has an even bigger impact on manufacturers than retailer, a higher proportion of cost subsidy will inhibit the enthusiasm of retailers to participate in tight cooperation. For manufacturers, incentivizing retailers to improve the quality of their data could yield better returns.
- (2) How does a firm's investment in big data analytics capability affect partner cooperation decisions? Our results show that firms can achieve larger innovation profits through big data cooperation, whether loose or tight mode. Members need to start with loose cooperation. For manufacturers, appropriate data cost subsidies can motivate retailers to improve data quality, thus strengthening the data cooperation between the two sides. Keeping the ratio of data to the cost of the subsidy can better encourage retailers to choose the tight cooperation mode. For retailers, they can maintain benefits by negotiating higher cost subsidies when manufacturer's big data analysis capability is low. Providing high-quality data can promote manufacturers to improve their own data analysis ability. With the improvement of the overall big data capability of the cooperation, it is easier for both parties to form a close cooperation mode. Supply chain members can establish a data-driven innovation cooperation to achieve win-win outcomes.
- (3) How will the matching of big data analytics capability and cooperative mode affect innovation performance? Our results show that firms that engage in actual innovation activities are more willingly to promote tight collaboration. When supply chain members only cooperate in a loose mode, the efficiency of big data on innovation performance is the lowest. Innovation returns can be significantly improved by proper contracts. Meanwhile, the manufacturer will get more benefits from tight mode, for the retailer, its benefits will also be increase, namely, there is "win-win relationship". In this way, a good cycle is formed jointly promoted by both parties to maximize the value of big data.

7.2 Limitations and future directions

There are several potential directions worthy for further research. In this paper, we have demonstrated how the big data analytics capability affect the decision-making in cooperative innovation process among supply chain members. This paper discusses two extreme modes of cooperation and one form of contract, but there are many kinds of contracts in supply chain cooperative, and different contracts have their applicable situations. Therefore, future research can consider the boundary conditions of the impact of big data analytics capability on different contractual cooperation modes. This study draws on the existing research on the benefits of big data analytics capability, focusing on the benefits of cooperative. Cooperation innovation is also analyzed from the perspective of cost reduction, such as Ge et al. (2014). Future research can consider the big data analytics capability both affects innovation revenue and cost. The problem in this paper is set in the context of manufacturers are leaders. When the power structure of supply chain is different, it occupies different dominant positions. Whether the conclusion is still universal needs to be further verified under different power structures.

Acknowledgements This research was supported by the National Natural Science Foundation of China (NSFC) (Project Nos. 71871197 and 71872158).

Data availability The model verification data used to support the findings of this study are included within the article.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Akter, S., Mccarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'ambra, J., & Shen, K. N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, 60, 102387.
- Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, 182, 113–131.
- Ali, M. M., Babai, M. Z., Boylan, J. E., & Syntetos, A. A. (2017). Supply chain forecasting when information is not shared. *European Journal of Operational Research*, 260(3), 984–994.
- Antonelli, C., & Fassio, C. (2016). The role of external knowledge(s) in the introduction of product and process innovations. *R&D Management*, 46(S3), 979–991.
- Babu, M. M., Rahman, M., Alam, A. & Dey, B. L. (2021). Exploring big data-driven innovation in the manufacturing sector: Evidence from UK firms. *Annals of Operations Research*.
- Bai, C., & Sarkis, J. (2016). Supplier development investment strategies: A game theoretic evaluation. *Annals of Operations Research*, 240, 583–615.
- Baldwin, C., Hiennerth, C., & von Hippel, E. (2006). How user innovations become commercial products: A theoretical investigation and case study. *Research Policy*, 35(9), 1291–1313.
- Belhadi, A., Kamble, S., Jabbour, C. J. C., Gunasekaran, A., Ndubisi, N. O., & Venkatesh, M. (2021). Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries. *Technological Forecasting and Social Change*, 163, 120447.
- Bellantuono, N., Giannoccaro, I., Pontrandolfo, P., & Tang, C. S. (2009). The implications of joint adoption of revenue sharing and advance booking discount programs. *International Journal of Production Economics*, 121(2), 383–394.
- Bendle, N. T., & Wang, X. S. (2016). Uncovering the message from the mess of big data. *Business Horizons*, 59(1), 115–124.

- Boiten, E. A. (2016). *Big Data Refinement*. [arXiv:1606.02017](https://arxiv.org/abs/1606.02017).
- Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5), 133–139.
- Chatterjee, S., Chaudhuri, R., Shah, M., & Maheshwari, P. (2022). Big data driven innovation for sustaining SME supply chain operation in post COVID-19 scenario: Moderating role of SME technology leadership. *Computers & Industrial Engineering*, 168, 108058.
- Chen, F., Lai, G., & Xiao, W. (2016a). Provision of incentives for information acquisition: Forecast-based contracts vs. menus of linear contracts. *Management Science*, 62(7), 1843–2147.
- Chen, W., Niebel, T., & Saam, M. (2016b). Are intangibles more productive in ICT-intensive industries? Evidence from EU countries. *Telecommunications Policy*, 40(5), 471–484.
- Cheng, C. C. J., Yang, C., & Sheu, C. (2016). Effects of open innovation and knowledge-based dynamic capabilities on radical innovation: An empirical study. *Journal of Engineering and Technology Management*, 41, 79–91.
- Chu, L. Y., Shamir, N., & Shin, H. (2017). Strategic communication for capacity alignment with pricing in a supply chain. *Management Science*, 63(12), 4366–4388.
- Ciampi, F., Demi, S., Magrini, A., Marzi, G., & Papa, A. (2021). Exploring the impact of big data analytics capabilities on business model innovation: The mediating role of entrepreneurial orientation. *Journal of Business Research*, 123, 1–13.
- Corrado, C., Haskel, J., & Lasinio, C. J. (2017). Knowledge spillovers, ICT and productivity growth. *Oxford Bulletin of Economics and Statistics*, 79(4), 592–618.
- Côrte-Real, N., Oliveira, T., & Ruivo, P. (2017). Assessing business value of Big Data Analytics in European firms. *Journal of Business Research*, 70, 379–390.
- Davenport, T. H. & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 90(10), 70–76, 128.
- Davenport, T. H., Barth, P., & Bean, R. (2012). How “Big Data” is different. *MIT Sloan Management Review*, 54(1), 43–46.
- Del Vecchio, P., Di Minin, A., Petruzzelli, A. M., Panniello, U., & Pirri, S. (2018). Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges. *Creativity and Innovation Management*, 27(1), 6–22.
- Dong, J. Q., & Yang, C.-H. (2020). Business value of big data analytics: A systems-theoretic approach and empirical test. *Information & Management*, 57(1), 103124.
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904.
- Feng, Q., & Shanthikumar, J. G. (2018). How research in production and operations management may evolve in the era of big data. *Production and Operations Management*, 27(9), 1670–1684.
- Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2019). Big data analytics capabilities and knowledge management: Impact on firm performance. *Management Decision*, 57(8), 1923–1936.
- Fosso Wamba, S., & Akter, S. (2019). Understanding supply chain analytics capabilities and agility for data-rich environments. *International Journal of Operations & Production Management*, 39(6/7/8), 887–912.
- Fosso Wamba, S., Gunasekaran, A., Dubey, R., & Ngai, E. W. T. (2018). Big data analytics in operations and supply chain management. *Annals of Operations Research*, 270(1–2), 1–4.
- Friedl, G., & Wagner, S. M. (2016). Supplier development investments in a triadic setting. *IEEE Transactions on Engineering Management*, 63(2), 136–150.
- Fusfeld, H. I., & Haklisch, C. S. (1985). Cooperative R&D for competitors. *Harvard Business Review*, 63(6), 60–76.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.
- Ge, Z., Hu, Q., & Xia, Y. (2014). Firms’ R&D cooperation behavior in a supply chain. *Production and Operations Management*, 23(4), 599–609.
- George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, 59(5), 1493–1507.
- Ghasemaghaci, M. (2019). Does data analytics use improve firm decision making quality? The role of knowledge sharing and data analytics competency. *Decision Support Systems*, 120, 14–24.
- Ghasemaghaci, M., & Calic, G. (2019). Does big data enhance firm innovation competency? The mediating role of data-driven insights. *Journal of Business Research*, 104, 69–84.
- Ghoshal, A., Kumar, S., & Mookerjee, V. (2020). Dilemma of data sharing alliance: when do competing personalizing and non-personalizing firms share data. *Production and Operations Management*, 29(8), 1918–1936.
- Gruner, R. L., Homburg, C., & Lukas, B. A. (2014). Firm-hosted online brand communities and new product success. *Journal of the Academy of Marketing Science*, 42, 29–48.

- Guo, L. (2009). The benefits of downstream information acquisition. *Marketing Science*, 28(3), 403–615.
- Gupta, M., & George, J. F. (2016a). Toward the development of a big data analytics capability. *Information & Management*, 53(8), 1049–1064.
- Gupta, M., & George, J. F. (2016b). Toward the development of a big data analytics capability. *Information & Management*, 53(8), 1049–1064.
- Ha, A. Y., Tian, Q., & Tong, S. (2017). Information sharing in competing supply chains with production cost reduction. *Manufacturing & Service Operations Management*, 19(2), 246–262.
- Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72–80.
- Hazen, B. T., Skipper, J. B., Boone, C. A., & Hill, R. R. (2018). Back in business: Operations research in support of big data analytics for operations and supply chain management. *Annals of Operations Research*, 270(1–2), 201–211.
- He, J., & Liu, H. (2017). Mining exploratory behavior to improve mobile app recommendations. *ACM Transactions on Information Systems*, 35(4), 1–37.
- He, J., Fang, X., Liu, H., & Li, X. (2019). Mobile app recommendation: An involvement-enhanced approach. *MIS Quarterly*, 43(3), 827–849.
- Hossain, M. A., Akter, S., & Yanamandram, V. (2020). Revisiting customer analytics capability for data-driven retailing. *Journal of Retailing and Consumer Services*, 56, 102187.
- Hsueh, J., Lin, N., & Li, H. (2010). The effects of network embeddedness on service innovation performance. *Service Industries Journal*, 30(10), 1723–1736.
- Hu, B., Meng, C., Xu, D., & Son, Y. (2016d). Three-echelon supply chain coordination with a loss-averse retailer and revenue sharing contracts. *International Journal of Production Economics*, 179, 192–202.
- Knight, L., Tate, W. L., Matopoulos, A., Meehan, J., & Salmi, A. (2016). Breaking the mold: Research process innovations in purchasing and supply management. *Journal of Purchasing and Supply Management*, 22(4), 239–243.
- Laursen, K., & Salter, A. J. (2014). The paradox of openness: Appropriability, external search and collaboration. *Research Policy*, 43(5), 867–878.
- Li, L., Chi, T., Hao, T., & Yu, T. (2018). Customer demand analysis of the electronic commerce supply chain using Big Data. *Annals of Operations Research*, 268(1–2), 113–128.
- Lin, C., & Kunnathur, A. (2019). Strategic orientations, developmental culture, and big data capability. *Journal of Business Research*, 105, 49–60.
- Liu, P., & Yi, S. (2018a). Investment decision-making and coordination of a three-stage supply chain considering Data Company in the Big Data era. *Annals of Operations Research*, 270(1–2), 255–271.
- Liu, P., & Yi, S. (2018b). A study on supply chain investment decision-making and coordination in the Big Data environment. *Annals of Operations Research*, 270(1–2), 235–253.
- Liu, Y., Soroka, A., Han, L., Jian, J., & Tang, M. (2020). Cloud-based big data analytics for customer insight-driven design innovation in SMEs. *International Journal of Information Management*, 51, 102034.
- Lozada, N., Arias-Pérez, J., & Perdomo-Charry, G. (2019). Big data analytics capability and co-innovation: An empirical study. *Heliyon*, 5(10), e02541.
- Lynn Wu, L. H. B. L. (2020). Data analytics, innovation, and firm productivity. *Management Science*, 66(5), 2017–2039.
- Mcafee, A. & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60–68, 128.
- Menguc, B., Auh, S., & Yannopoulos, P. (2014). Customer and supplier involvement in design: The moderating role of incremental and radical innovation capability. *Journal of Product Innovation Management*, 31(2), 313–328.
- Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management*, 30(2), 272–298.
- Patrick Mikalef, M. B., Lekakos, G., & Krogstie, J. (2019). Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *British Journal of Management*, 30, 272–298.
- Piccarozzi, M., & Aquilani, B. (2022). The role of Big Data in the business challenge of Covid-19: A systematic literature review in managerial studies. *Procedia Computer Science*, 200, 1746–1755.
- Pigni, F., Piccoli, G., & Watson, R. (2016). Digital data streams: Creating value from the real-time flow of big data. *California Management Review*, 58(3), 5–25.
- Rahman, M. S., Hossain, M. A., & Abdel Fattah, F. A. M. (2021). Does marketing analytics capability boost firms' competitive marketing performance in data-rich business environment? *Journal of Enterprise Information Management*, 35(2), 455–480.

- Rahman, M. S., Hossain, M. A., Fattah, F. A. M. A., & Akter, S. (2020). Optimizing competitive performance of service firms in data-rich environment. *Journal of Service Theory and Practice*, 30(6), 681–706.
- Ransbotham, S., Kiron, D., Gerbert, P., & Reeves, M. (2017). *Reshaping business with artificial intelligence: Closing the gap between ambition and action*. Massachusetts Institute of Technology.
- Sampson, R. C. (2004). Organizational choice in R&D alliances: Knowledge based and transaction cost perspectives. *Managerial & Decision Economics*, 25(6), 421–436.
- Shin, H., & Tunca, T. (2010). Do firms invest in forecasting efficiently? The effect of competition on demand forecast investments and supply chain coordination. *Operations Research*, 58(61), 1592–1610.
- Shollo, A., & Galliers, R. D. (2016). Towards an understanding of the role of business intelligence systems in organisational knowing. *Information Systems Journal*, 26(4), 339–367.
- Skippari, M., Laukkanen, M., & Salo, J. (2017). Cognitive barriers to collaborative innovation generation in supply chain relationships. *Industrial Marketing Management*, 62, 108–117.
- Stefansson, G. (2002). Business-to-business data sharing: A source for integration of supply chains. *International Journal of Production Economics*, 75(1), 135–146.
- Sultana, S., Akter, S., & Kyriazis, E. (2022a). How data-driven innovation capability is shaping the future of market agility and competitive performance? *Technological Forecasting & Social Change*, 174, 121260.
- Sultana, S., Akter, S., & Kyriazis, E. (2022b). Theorising data-driven innovation capabilities to survive and thrive in the digital economy. *Journal of Strategic Marketing*.
- Sultana, S., Akter, S., Kyriazis, E., & Wamba, S. F. (2021). Architecting and developing big data-driven innovation (DDI) in the digital economy. *Journal of Global Information Management*, 29(3), 165–187.
- Tan, K. H., Zhan, Y., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities An analytic infra. *International Journal of Production Economics*, 165, 223–233.
- Tang, W., & Girotra, K. (2017). Using advance purchase discount contracts under uncertain information acquisition cost. *Production and Operations Management*, 26(8), 1553–1567.
- Taylor, T. A., & Xiao, W. (2009). Incentives for retailer forecasting: Rebates vs. returns. *Management Science*, 55(10), iv–1753.
- Trigo, A., & Vence, X. (2012). Scope and patterns of innovation cooperation in Spanish service enterprises. *Research Policy*, 41(3), 602–613.
- Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356–365.
- Wei, Q., Qiao, D., Zhang, J., Chen, G., & Guo, X. (2016). A novel bipartite graph based competitiveness degree analysis from query logs. *ACM Transactions on Knowledge Discovery from Data*, 11(2), 1–25.
- Williams, T. (2005). Cooperation by design: Structure and cooperation in interorganizational networks. *Journal of Business Research*, 58(2), 223–231.
- Wu, J. (2014). Cooperation with competitors and product innovation: Moderating effects of technological capability and alliances with universities. *Industrial Marketing Management*, 43(2), 199–209.
- Wu, J., Guo, S., Li, J., & Zeng, D. (2016). Big Data Meet Green Challenges: Greening Big Data. *IEEE Systems Journal*, 10(3), 873–887.
- Xie, K., Wu, Y., Xiao, J., & Hu, Q. (2016). Value co-creation between firms and customers: The role of big data-based cooperative assets. *Information & Management*, 53(8), 1034–1048.
- Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, 69(5), 1562–1566.
- Yang, S., & Zhou, K. (2015). Management issues in big data : The resource-based view of big data. *Journal of Management Sciences in China*, 18(5), 1–8.
- Yasmin, M., Tatoglu, E., Kilic, H. S., Zaim, S., & Delen, D. (2020). Big data analytics capabilities and firm performance: An integrated MCDM approach. *Journal of Business Research*, 114, 1–15.