

Unveiling data value: A configurational approach to big data and business model design

Fengquan Wang

fq.wang92@foxmail.com

Chongqing Technology and Business University

Article

Keywords: big data, business model design, data value, data-driven business models, data network effects, fsQCA

Posted Date: March 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9062736/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

The rapid growth of big data has attracted academic interest, but extant studies are inadequate to address how to derive data value from the configurations between big data and business model design. To this end, we build a framework that integrates business model design (novelty, efficiency) into big data's 4Vs (volume, variety, velocity, value) and apply a holistic configurational approach to investigate what combinations among volume, variety, velocity, novelty, and efficiency can affect data value. Based on the sample of 41 consumer internet start-ups in China, we use fuzzy-set qualitative comparative analysis (fsQCA) to reveal the configurational results. The findings suggest that there are four configurations leading to high data value, which are respectively labelled as experience-oriented business models (EBM), operation-enhanced business models (OBM), content-centric business models (CBM), and scale-dominant business models (SBM). Thus, we identify different prototypes of data-driven business models from these configurations, the underlying mechanisms are illustrated by data network effects and representative case practices. By contrast, configurations leading to not-high data value are collectively labelled as business models deficient in data-business synergy. Our study not only advances big data research in management and business and data-driven business model research, but also sheds light on the extension of data network effects' theoretical boundary and research context.

1. Introduction

As a game changer in the digital era, big data represents a new technology paradigm for data characterized by volume, variety, velocity, and value (4Vs) (Manyika et al., 2011; Lee, 2017; Cappa, Oriani, Peruffo, & McCarthy, 2021). Among big data's 4Vs, data value is the ultimate goal after a variety of data are generated at high volume and processed at high velocity, which determines the success of firms' digitalization (Grover, Chiang, Liang, & Zhang, 2018). Despite the great importance of big data, many firms seeking digitalization are confronted with a data dilemma, that is, due to the lack of appropriate business model design, implementing or not implementing a big data strategy will result in failure (Ghasemaghaei & Calic, 2020; Olabode, Boso, Hultman, & Leonidou, 2022). Therefore, in order to derive data value from big data, it is necessary to take business model design into account (Ciampi, Demi, Magrini, Marzi, & Papa, 2021; Urbinati, Bogers, Chiesa, & Frattini, 2019), but little is known about the fit between big data and business model design, addressing this issue is critical to understand how firms seeking digitalization can successfully realize data value.

Scholars have already noticed the importance of the fit between technologies and business models, as Chesbrough and Rosenbloom (2002, p. 529) note, "*a successful business model creates a heuristic logic that connects technical potential with the realization of economic value*". In our research context, extant studies are relatively scarce. On the one hand, despite the various definitions of big data, its main characteristics or dimensions have reached a consensus — volume, variety, velocity, and value (4Vs) (Lee, 2017; Grover et al., 2018; Ghasemaghaei & Calic, 2019, 2020). But most of the relevant studies propose big data's dimensions separately, lacking an integrated framework among them, especially when business model design is introduced (Lee, 2017; Wiener, Saunders, & Marabelli, 2020).

On the other hand, as mentioned above, the wide application of big data calls for appropriate business model design (Ciampi et al., 2021; Urbinati et al., 2019), which brings data-driven business models. Big data not only changes the key components of a business model (i.e., value creation, value delivery, and value capture) (Sjödin, Parida, Palmie, & Wincent, 2021; Wiener et al., 2020; Clauss, 2017), but also leads to different themes of business model design (e.g., novelty, efficiency, complementarity, and lock-in) (Amit & Zott, 2001; Zott & Amit, 2007, 2008), while holistic themes are less considered in extant studies, causing a missing in prototypes of data-driven business models (Trischler & Li-Ying, 2022).

In addition, as a new theory put forward recently, data network effects emphasize the boosting role of data-driven learning and improvements in the virtuous cycle between user value and user data generated from product use, it requires the combination of digital technologies – such as big data and artificial intelligence (AI) – and business models (Gregory, Henfridsson, Kaganer, & Kyriakou, 2021, 2022; Haftor, Climent, & Lundstrom, 2021), which is quite in line with our study. Although there are some arguments on data network effects concerning the overlook of value capture and the divergence of data resource ownership (Clough & Wu, 2022), it reflects the fact that the theory of data network effects needs to be examined and advanced in more contextual circumstances (Gregory et al., 2021, 2022). So we adopt the theory of data network effects to uncover the underlying mechanisms behind the fit between big data and business model design and further extend its theoretical boundary and research context.

Therefore, grounded on the extant studies and the identified gaps, the overall research questions are: *How do big data and business model design fit to form different configurations? What are these configurations? And how do these configurations affect data value?*

To answer these questions, we integrate business model design into big data's 4Vs and use fuzzy-set qualitative comparative analysis (fsQCA) to explore their relationships. Specifically, following Zott and Amit (2007, 2008), we consider novelty and efficiency as the two critical themes of business model design. Among big data's 4Vs, we regard volume, variety, and velocity as the antecedents of value, which is also recognized by many scholars (Cappa et al., 2021; Grover et al., 2018; Lee, 2017). Further, referring to Wielgos, Homburg, and Kuehnl (2021), we divide data value into customer performance and firm performance to make it more concrete and parsimonious. Thus, we build a framework to investigate the configurations between big data (volume, variety, velocity) and business model design (novelty, efficiency) as well as their impact on data value (customer performance, firm performance) through fsQCA. Our research sample is 41 consumer internet start-ups in China, the measurement of 4Vs is based on the objective indices of each case's APP from Android Market. Novelty and efficiency are assessed by well-trained researchers based on the primary and secondary data collected from each case. The findings show that velocity is the necessary condition for high data value (both high customer performance and high firm performance), but velocity alone is not enough, it works jointly with other conditions. Taken together, we get three configurations leading to high customer performance, which are respectively labelled as experience-oriented business models (EBM), operation-enhanced business models (OBM), and content-centric business models (CBM); one configuration leading to high firm performance, which is labelled as scale-dominant business models (SBM); and nine configurations

leading to not-high data value, which are collectively labelled as business models deficient in data-business synergy for contrast. Further, the four high-data-value configurations can be seen as different prototypes of data-driven business models, their underlying mechanisms are illustrated by data network effects and representative case practices.

Our study makes three contributions to the literature. First, answering the call for deepening big data research in management and business (Simsek, Vaara, Paruchuri, Nadkarni, & Shaw, 2019), we offer an integrated view of big data by clarifying the relationship among big data's 4Vs when business model design is introduced. Second, we advance data-driven business model research in a holistic and systematic way by identifying different prototypes of data-driven business models from the configurations between big data and business model design (Sorescu, 2017; Park & Mithas, 2020). Third, we illustrate the underlying mechanisms behind the configurations and their impact on data value through the theory of data network effects, which extends its theoretical boundary and research context (Gregory et al., 2021, 2022).

2. Literature review and research framework

2.1 Big data and data value

Big data refers to datasets whose size is beyond the capacity of typical database software tools to capture, store, manage, and analyze data, which is mainly characterized by the four dimensions (also known as 4Vs): volume, variety, velocity, and value (Manyika et al., 2011; Lee, 2017; Simsek et al., 2019). Volume describes the amount of data generated and collected. Variety represents the diversification of data types, including structured, semi-structured, and unstructured data. Velocity relates to the speed at which data are generated, collected, and processed (Lee, 2017; Cappa et al., 2021). Finally, as the most important dimension in 4Vs, value (or data value) is the benefit gained from big data, including better decision-making, cost savings, and higher product/service quality, which highlights the availability of hidden insights derived from big data analysis (Grover et al., 2018; Davenport, 2014; Lavallo, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). In addition, as big data evolves, more dimensions, such as veracity, visualization, and variability, have emerged (Simsek et al., 2019; Cappa et al., 2021). In our study, we focus on the 4Vs mentioned above.

Big data's 4Vs have been proposed for a long time and their connotations are increasingly mature (Lee, 2017; Simsek et al., 2019). However, most of the extant studies propose these dimensions separately, lacking an integrated view of big data (Lee, 2017; Johnson, Friend, & Lee, 2017; Ghasemaghahi & Calic, 2019, 2020). To help fully understand the relationship among the 4Vs, we suggest that volume, variety, and velocity of big data are the antecedents of data value. Specifically, volume, variety, and velocity can lead to data value in that: First, large volumes of data provide a solid foundation for decision-making and act as sufficient evidence to sense market opportunities and threats (Merendino et al., 2018; van Rijmenan, Erekhinskaya, Schweitzer, & Williams, 2019). Second, various types of data indicate that firms have access to rich information to gain more comprehensive views on both internal and external

operations, which contributes to seizing opportunities and avoiding threats (Sorescu, 2017; Ghasemaghaei & Calic, 2020; van Rijmenan et al., 2019). Third, high velocity of data ensures that real-time decisions are made in the changing market, that is, in order to quickly gain new insights and enhance agility, data should be continually generated, collected, and processed in a timely manner (Ghasemaghaei & Calic, 2019, 2020; Mikalef, Pappas, Krogstie, & Pavlou, 2020; Wamba et al., 2017). Moreover, any single dimension of volume, variety, and velocity can hardly bring high data value, they need to interact with each other, and further form configurations with business model design (Olabode et al., 2022; Wiener et al., 2020; Sorescu, 2017), since business models can release the potential value of technical innovation by clarifying the logic of value creation, delivery, and capture (Teece, 2010; Chesbrough & Rosenbloom, 2002), which is also the standpoint of our study.

In brief, although big data has been discussed broadly on its different dimensions, extant studies are still deficient in building an integrated framework to explore their interrelationships, especially when business model design is introduced. Therefore, our study addresses this gap by combining business model design with big data's 4Vs and investigates how the fit among volume, variety, velocity, and business model design can affect data value.

2.2 Data-driven business models

Data-driven business models refer to firms' new mechanisms of value creation, delivery, and capture in big data environments compared with traditional business models (Sorescu, 2017; Wiener et al., 2020), highlighting the changes in key components of business models as well as the themes of business model design enabled by big data (Trischler & Li-Ying, 2022). On the one hand, big data changes the key components (i.e., value creation, value delivery, and value capture) of business models, which is the mainstream of extant studies. For value creation that centers on the demand side, big data helps firms gain more insights into customers, so as to launch innovative products and services to better meet customer demands and improve their experience (Priem, Butler, & Li, 2013; Holmlund et al., 2020; Erevelles, Fukawa, & Swayne, 2016). For value delivery that centers on the networks and linkages, big data enables the interaction between the demand side and the supply side, promoting a more efficient demand-supply match (Zhang & Xiao, 2020; Xie, Wu, Xiao, & Hu, 2016). For value capture that centers on the supply side, big data facilitates the high-quality operation to reduce costs and increase revenue, thus enhancing firm performance (Priem, Wenzel, & Koch, 2018; Suoniemi, Meyer-Waarden, Munzel, Zablah, & Straub, 2020).

On the other hand, changes in components will ultimately lead to different themes of data-driven business models (Trischler & Li-Ying, 2022; Foss & Saebi, 2017), which is less discussed and the focus of our study. Following Zott and Amit (2007, 2008), we adopt the two critical themes of business model design – novelty and efficiency – to describe data-driven business models. Specifically, novelty represents new ways of transaction among stakeholders, such as new offerings, new linkages, new participants, and new mechanisms (Amit & Zott, 2001; Zott & Amit, 2007, 2008). Big data can connect previously unconnected parties, and all the digital touchpoints and linkages become data sources (Manyika et al., 2011; Simsek et al., 2019). Tremendous data can release tremendous value after being

analyzed into insights in real time, because such insights are crucial for innovating products and services to satisfy customers' various demands (Zhang & Xiao, 2020; Tan & Zhan, 2017; Lavallo et al., 2011). The data resources combined with data analytics capabilities can also be a new business for firms seeking new growth (Davenport, 2014; McAfee & Brynjolfsson, 2012). Therefore, big data can promote the novelty of business model design. Meanwhile, efficiency emphasizes the reduction of transaction costs, including the costs of search, inventory, marketing, and communication (Amit & Zott, 2001; Zott & Amit, 2007, 2008). Big data is capable of eliminating information asymmetry, thus enabling firms as well as customers to make well-informed decisions (van Rijmenan et al., 2019; Merendino et al., 2018; Holmlund et al., 2020). The improvement of decision-making efficiency also comes with the simplified transaction processes and agile reactions to market changes (Hajli, Tajvidi, Gbadamosi, & Nadeem, 2020; Ghezzi & Cavallo, 2020). Besides, based on techniques such as personalized recommendation, big data can enhance the efficiency of demand-supply match (Grigorios, Magrizos, Kostopoulos, Drossos, & Santos, 2022; Erevelles et al., 2016; Gregory et al., 2021), which further contributes to the reduction of transaction costs. Therefore, big data can promote the efficiency of business model design.

As mentioned above, extant studies mainly explore the impact of big data on particular components of business models from a single linear perspective, lacking the consideration for its holistic themes and prototypes (Trischler & Li-Ying, 2022; Foss & Saebi, 2017). Moreover, the essence of business models is an activity system (Amit & Zott, 2001; Zott & Amit, 2010; Foss & Saebi, 2017), big data further blurs its boundary, which highlights the systematic characteristics of data-driven business models (Wiener et al., 2020; Sorescu, 2017). Therefore, in order to reveal data-driven business models from a holistic and systematic view, investigating the configurations between big data and business model design is necessary.

2.3 Data network effects

Data network effects refer to the virtuous cycle that user value and user data promote each other, which derives from the standard network effects (including direct network effects and indirect network effects) for platforms (Gregory et al., 2021; Haftor et al., 2021). Specifically, in the digital era, data have become the new production factor of firms, some scholars try to integrate data-driven learning and improvements realized with AI into standard network effects to explore the connotation and mechanism of data network effects. For example, Gregory et al. (2021, p. 534) claim that *"a platform exhibits data network effects if, the more that the platform learns from the data it collects on users, the more valuable the platform becomes to each user"*, highlighting the importance of AI capabilities (i.e., prediction speed and prediction accuracy) in the process of perceived user value creation, which is moderated by data stewardship (i.e., data quantity and data quality), user-centric design (i.e., performance expectancy and effort expectancy), and platform legitimation (i.e., personal data use and prediction explainability).

As a new theory put forward recently by Gregory et al. (2021), the term "data network effects" is seldom explicitly mentioned and elaborated in other studies, but a few relevant studies touch its core concepts, pointing out that user data are the important source of innovation for firms, by continually interacting

with users, firms can engage users in production activities such as R&D in a digital manner, and data are generated and collected at the same time (Zhang & Xiao, 2020; Xie et al., 2016). Drawing on big data analytics techniques (i.e., AI algorithms), data value can be released (Grover et al., 2018; Mikalef et al., 2020; Lavallo et al., 2011). That is, the iteration of new product development or product innovation is accelerated (Tan & Zhan, 2017), the agility in response to market changes is strengthened (Hajli et al., 2020), and the fit between products and users is enhanced (Holmlund et al., 2020; Erevelles et al., 2016; Xie et al., 2016). Moreover, when firms can satisfy users' diversified demands efficiently and precisely, user scale and corresponding data volume will expand, thus promoting both firm performance and customer performance to start a virtuous cycle (Wielgos et al., 2021; Knudsen, Lien, Timmermans, Belik, & Pandey, 2021). However, this recently proposed theory has also received critical scrutiny in two aspects. One is that the decentralized feature of platforms poses challenges to value capture; the other is that the quantity and quality of platforms' data are loosely associated with the size of platforms' user base (Clough & Wu, 2022). Both seem highly pertinent, and Gregory et al. (2022) make further clarifications to respond to these arguments, which suggests that the theory of data network effects needs to be examined and advanced in more contextual circumstances.

Our study investigates the configurations that affect data value between big data and business model design, highlighting the boosting role of data-driven learning and improvements in this process, which is quite consistent with the connotation of data network effects. Meanwhile, the theory of data network effects is still at an early stage of development, its theoretical boundary and research context are worth being further extended (Gregory et al., 2021, 2022). Therefore, we choose data network effects as the theoretical basis to reveal how the configurations between big data and business model design affect data value.

2.4 Research framework

Based on the discussion above, we build the research framework as shown in Fig. 1. Following Ghasemaghaei and Calic (2019, 2020) and Lee (2017), we adopt big data's 4Vs (i.e., volume, variety, velocity, and value), and referring to Zott and Amit (2007, 2008), we consider novelty and efficiency as the two important themes of business model design. In addition, due to the complexity of data value, we divide it into customer performance and firm performance to make the research framework more parsimonious – customer performance that is derived from value creation describes perceived user value or user experience from the demand side; firm performance that is derived from value capture takes the form of financial performance or competitive advantages on the supply side (Wielgos et al., 2021; Park & Mithas, 2020; Priem et al., 2013, 2018). Thus, we combine big data with business model design to investigate how volume, variety, velocity, novelty, and efficiency fit to form different configurations and the corresponding impact on data value, the underlying mechanisms will be illustrated by data network effects and representative cases.

3. Method and data

3.1 Analytical approach

We used fuzzy-set qualitative comparative analysis (fsQCA) to discover the configurations between big data and business model design that affect data value. FsQCA emphasizes causal complexity and combines inductive case study and deductive empirical study (Ragin, 2008; Misangyi et al., 2017), which has several important features. First, fsQCA explains conjunctural causation, that is, outcomes are usually results of more than two conditions. Second, fsQCA reports equifinality, that is, different configurations may lead to the same result. Third, fsQCA supports asymmetry, that is, causally relevant conditions may account for the presence of an outcome but not for its absence. And fourth, fsQCA can be conducted using a small, medium, or large sample, which shows a good prospect for wider application (Fiss, 2011; Misangyi et al., 2017; Kumar, Sahoo, Lim, Kraus, & Bamel, 2022).

Therefore, fsQCA is particularly suitable for our study to generate new insights regarding what configurations between big data and business model design are, and how these configurations affect data value.

3.2 Sample selection

Since consumption is highly digitalized and China is globally leading in this domain (Guo, Wang, Su, & Wang, 2020), we set consumer internet start-ups in China as the research sample, which originates from ITJUZI.COM (known as a venture capital database and business information service provider focusing on Chinese internet start-ups) that covers the basic firm information, such as foundation, financing, mergers and acquisitions.

Specifically, under the overall guidance of “theoretical sampling” — that is, the sample is chosen based on research questions and theory construction, which requires the sample to have similarities in general as well as differences in individual (Ragin, 2008; Fiss, 2011; Misangyi et al., 2017), we used three criteria to screen the sample. First, we selected those start-ups founded after 2010 and have/had operated for more than three years, for the reason that China’s rapid development of mobile internet started in 2010 and the entrepreneurial practice boomed since then (many famous start-ups showed up, such as Xiaomi, Meituan, and Pinduoduo), and more than three years of duration time ensures that the start-up is less speculative and more viable in business (Velu, 2015). We got 128 cases in this step. Second, we further selected the start-ups whose online operation is mainly based on their own APPs, which makes it possible for firms to gain more comprehensive data to make more informed and timely decisions. In addition, we selected start-ups varying in industries (e.g., life services, shopping e-business, travelling, etc.) and performance (both high and not-high), leaving 75 cases. Third, we drew on multiple sources to collect data and excluded those cases lacking information on the key constructs, the final sample decreased to 41 cases.

3.3 Measurement and calibration

To avoid common variance, we measured constructs through multi-source heterogeneous data, including objective indices (for 4Vs) and subjective assessments (for novelty and efficiency). After

assigning value to each construct, calibration – a process of assigning constructs with set membership scores – was conducted for fsQCA. Following Fiss (2011) and Greckhamer (2016), we adopted the adjusted distribution method to calibrate constructs. That is, arranging in numerical order from small to large, we set fully in membership anchor at the 90th percentile, crossover point at the 50th percentile, and fully out membership anchor at the 10th percentile. Table 1 shows the measurement and calibration of constructs. The details of measurement are given below.

Table 1
Measurement and calibration of constructs

Constructs	Measurement	Calibration anchors
Data value		
Customer performance	APP user rating (1 to 5) in Android Market	4.8, 4.3, 3.7
Firm performance	APP average annual growth rate (%) of active users in Android Market	41.65, 0.99, -0.13
Big data		
Volume	The number of APP downloads (0.1 billion) in Android Market	102.11, 6.05, 0.13
Variety	The number of types of user information requested by APP in Android Market	23, 19, 14
Velocity	The number of average annual updates of APP after its launch in Android Market	48, 29, 15.33
Business model design		
Novelty	Five items from the developed scale of Zott and Amit (2007, 2008) (5-point assessments)	4.67, 3.6, 2.67
Efficiency	Five items from the developed scale of Zott and Amit (2007, 2008) (5-point assessments)	4.6, 3.53, 2.53

Data value. Two components – customer performance and firm performance – constitute data value. With reference to Cappa et al. (2021), Wielgos et al. (2021), Priem et al. (2013, 2018), and Johnson et al. (2017), we respectively measured each case’s customer performance and firm performance by its APP user rating (ranging from 1 to 5, which represents user satisfaction level) and APP average annual growth rate (%) of active users (for firms, the more active users, the more potential to monetize). Relevant data were recorded from Android Market by October 2023.

Big data. We focused on the three dimensions of big data’s 4Vs – volume, variety, and velocity, which are the antecedents of data value. Following Cappa et al. (2021), we measured each case’s volume by the number of APP downloads (0.1 billion), variety by the number of types of user information requested by

APP, and velocity by the number of average annual updates of APP after its launch. Relevant data were recorded from Android Market by October 2023.

Business model design. In accordance with Zott and Amit (2007, 2008), we considered novelty and efficiency as the two critical themes of business model design. Following their developed scales and adopting 5-point assessments (“5” to “1” represent “strongly agree” to “strongly disagree”), we respectively used five items to measure novelty and efficiency, these items are listed in Table 2. Specifically, we collected materials of each case from multiple sources, such as official websites, media reports, internet forums, semi-structured interviews (if available), IPO prospectuses and annual reports (if listed). Material collection lasted until October 2023. Three researchers (they are well-trained and have a good knowledge of business models) separately read each case’s materials and scored the items – the final score of each item took the average score of three researchers, and the assigned value of novelty and efficiency took the average score of corresponding items. To ensure reliability, we used SPSS 21.0 to examine Cronbach’s alpha of novelty and efficiency, the results show that they both exceed 0.7, so the reliability is satisfactory.

Table 2
Scales and items

Constructs	Items	Ratings	Cronbach's alpha
Novelty			0.96
	N1: The business model involves new participants.	5; 4; 3; 2; 1	
	N2: The business model links participants in novel ways.	5; 4; 3; 2; 1	
	N3: The business model's incentive mechanisms are novel.	5; 4; 3; 2; 1	
	N4: The business model offers new products and services.	5; 4; 3; 2; 1	
	N5: The business model contains novel resources and capabilities.	5; 4; 3; 2; 1	
Efficiency			0.94
	E1: The business model reduces costs of search, inventory, etc.	5; 4; 3; 2; 1	
	E2: The business model eliminates information asymmetry.	5; 4; 3; 2; 1	
	E3: The business model enables simple and fast transactions.	5; 4; 3; 2; 1	
	E4: The business model matches demand and supply precisely.	5; 4; 3; 2; 1	
	E5: The business model makes quick responses to market changes.	5; 4; 3; 2; 1	

3.4 Analytical procedures

We used fsQCA 3.0 – a software for configurational analysis – to conduct necessity analysis, sufficiency analysis, and robustness tests in sequence.

Necessity analysis. We started the analysis by examining if there is any necessary condition for high data value (including high customer performance and high firm performance) and its negation (including not-high customer performance and not-high firm performance). The necessary condition in fsQCA indicates the degree to which the condition accounts for the outcome (Ragin, 2008). Following Greckhamer (2016), we set a consistency score of 0.9 as the necessity cut-off threshold, that is, a condition with a consistency score greater than or equal to 0.9 can be considered as a necessary condition. Table 3 shows the consistency scores of necessity analysis, which suggests that velocity is the necessary condition for high data value (both high customer performance and high firm

performance), it also means that firms achieving high data value must have high velocity of big data, but velocity alone is not enough to result in high data value, some other conditions (i.e., volume, variety, novelty, and efficiency) need to be taken into account as a whole. In addition, no necessary condition exists in the negation of data value.

Table 3
Consistency scores of necessity analysis

Conditions	High customer performance	High firm performance	Not-high customer performance	Not-high firm performance
Volume	0.62	0.74	0.48	0.45
~Volume	0.67	0.65	0.77	0.79
Variety	0.74	0.77	0.6	0.6
~Variety	0.55	0.58	0.65	0.62
Velocity	0.9	0.94	0.4	0.43
~Velocity	0.42	0.43	0.87	0.8
Novelty	0.83	0.89	0.5	0.49
~Novelty	0.5	0.47	0.77	0.74
Efficiency	0.78	0.88	0.51	0.51
~Efficiency	0.53	0.53	0.75	0.75

The negate sign (~) indicates the absence of a condition.

Sufficiency analysis. We discovered configurations acting as sufficient solutions to data value (both high and not-high) by the Truth Table algorithm. In line with Fiss (2011), we set the consistency threshold at 0.8 and PRI consistency threshold at 0.75, and we also set the frequency threshold at 1 to ensure at least one representative case exists in each identified configuration. Table 4a and Table 4b respectively show the configurational results of high data value and not-high data value.

Two key indicators — solution consistency and coverage — are reported in Table 4a and Table 4b. The consistency score represents how well the configuration corresponds to the data, which is generally suggested to achieve 0.8 to be acceptable (Fiss, 2011; Ragin, 2008; Greckhamer, 2016). In this study, overall consistency scores are 0.93, 0.94, 0.95, and 0.96 for high customer performance, high firm performance, not-high customer performance, and not-high firm performance, respectively, and the individual consistency scores of all the configurations are also greater than 0.8. Besides, the coverage score represents how important the configuration is to the outcome. Although there is no consensus on the acceptable level of overall coverage that indicates the empirical importance of the whole solutions, prior studies usually achieve greater than 0.3 for this indicator (Fiss, 2011; Greckhamer, 2016). We obtained overall coverage scores of 0.71, 0.63, 0.83, and 0.86 for high customer performance, high firm

performance, not-high customer performance, and not-high firm performance, respectively. In addition, raw coverage (an individual configuration's explanatory power) and unique coverage (each configuration's unique contribution) are also reported.

Following Ragin (2008), only intermediate solutions generated by fsQCA 3.0 are reported in Table 4a and Table 4b. Thus, we found four configurations for high data value (three for high customer performance, one for high firm performance) and nine configurations for not-high data value (three for not-high customer performance, six for not-high firm performance).

Robustness tests. We performed robustness tests to examine the stability of the results. Following prior studies, we increased the consistency threshold and PRI consistency threshold to 0.85 and 0.8, respectively, the results remained unchanged. Besides, we changed the frequency threshold from 1 to 2, which resulted in fewer configurations and lower overall coverage scores, but also more parsimonious solutions, and the patterns are similar. So our key findings are robust.

Table 4a. Configurational results of sufficiency analysis (high data value)

	High customer performance			High firm performance
	Configuration 1a	Configuration 2a	Configuration 3a	Configuration 4a
Volume		⊗	●	●
Variety		⊗	⊗	
Velocity	●	●	●	●
Novelty	●		●	●
Efficiency	●	●		●
Consistency	0.94	0.94	0.91	0.94
Raw coverage	0.67	0.36	0.31	0.63
Unique coverage	0.29	0.02	0.01	0.63
Configuration label	Experience-oriented business models (EBM)	Operation-enhanced business models (OBM)	Content-centric business models (CBM)	Scale-dominant business models (SBM)
Representative case	Meituan	Convenient Bee	Xiaohongshu	Pinduoduo
Overall solution consistency	0.93			0.94
Overall solution coverage	0.71			0.63

● indicates the presence of a condition, ⊗ indicates the absence of a condition. Large characters indicate the core conditions, small characters indicate the peripheral conditions. Blank spaces indicate the presence or absence “does not matter”.

Representative cases are selected from the cases with membership greater than 0.5.

Table 4b. Configurational results of sufficiency analysis (not-high data value)

	Not-high customer performance			Not-high firm performance					
	Configuration 1b	Configuration 2b	Configuration 3b	Configuration 4b	Configuration 5b	Configuration 6b	Configuration 7b	Configuration 8b	Configuration 9b
Volume		⊗	●		⊗	⊗	⊗		●
Variety		⊗			⊗	⊗	⊗	●	●
Velocity	⊗	⊗	⊗	⊗			⊗		⊗
Novelty	⊗		●	⊗		⊗		⊗	●
Efficiency	⊗		●	⊗	⊗			⊗	●
Consistency	0.95	0.95	0.98	0.99	0.97	0.97	0.99	0.96	0.98
Raw coverage	0.64	0.54	0.29	0.58	0.46	0.45	0.5	0.45	0.22
Unique coverage	0.19	0.11	0.05	0.04	0.02	0.01	0.06	0.04	0.02
Configuration label	Business models deficient in data-business synergy								
Representative case	of								
Overall solution consistency	0.95			0.96					
Overall solution coverage	0.83			0.86					

● indicates the presence of a condition, ⊗ indicates the absence of a condition. Large characters indicate the core conditions, small characters indicate the peripheral conditions. Blank spaces indicate the presence or absence “does not matter”.

All the configurations are labelled collectively, and the most representative case is selected from the cases with membership greater than 0.5.

4. Findings and proposition development

Based on the configurational results reported above, we highlight the configurations leading to high data value. Referring to Wiener et al. (2020) and Weill and Woerner (2013), we consider these configurations as different prototypes of data-driven business models and respectively label them as experience-oriented business models (EBM), operation-enhanced business models (OBM), content-centric business models (CBM), and scale-dominant business models (SBM). By contrast, for the configurations leading to not-high data value, we collectively label them as business models deficient in data-business synergy. Next, combined with the theoretical analysis of data network effects and representative case practices, we interpret our findings and develop corresponding propositions.

4.1 Experience-oriented business models (EBM)

Configuration 1a (*Velocity*Novelty*Efficiency*) is labelled as experience-oriented business models (EBM), which indicate that high customer performance can be realized when big data is processed rapidly and business model design is featured with both novelty and efficiency, thus optimizing customer experience. According to data network effects, “*the value each user perceives depends on the scale of data-driven learning and improvements realized with AI*” (Gregory et al., 2021, p. 535), indicating that data-driven learning and improvements enable the formation of a virtuous cycle between user value and user data (Gregory et al., 2021; Knudsen et al., 2021; Haftor et al., 2021). As one of the crucial elements in data network effects, prediction speed emphasizes the importance of big data velocity. That is, in order to improve perceived user value, AI algorithms based on the learning of user data should continuously bring users with products or real-time personalized experience in line with their preferences (Gregory et al., 2021; Haftor et al., 2021). In this process, business model design should also adapt to big data environments. As for novelty, new participants (e.g., ecosystem partners), linkages (e.g., online-to-

offline), mechanisms (e.g., freemium), offerings (e.g., digital content), resources (e.g., data), and capabilities (e.g., big data analytics capabilities) are introduced to the existing activity system (Wang, 2021; Tidhar & Eisenhardt, 2020; Mikalef et al., 2020), thus promoting performance expectancy in data network effects. As for efficiency, the free flow of big data largely eliminates information asymmetry, which can not only reduce transaction costs and complexity, but also make quick responses to market changes and reach a precise demand-supply match (Grover et al., 2018; Davenport, 2014; Hajli et al., 2020), thus bringing effort expectancy in data network effects.

A representative case of EBM is Meituan. Founded in March 2010, Meituan's positioning is to be an online-to-offline e-business platform in life services (similar to Groupon), and its mission is *"we help people eat better, live better"*. After more than 10 years of development, Meituan is now the largest and most successful life-service platform in China, with a market value of more than 180 billion dollars, ranking third among Chinese internet companies. Meituan is excellent at utilizing data to improve user experience, based on its advanced AI algorithms, the products or services displayed in user interfaces are adjusted in real time to meet users' preferences according to their geographical locations and browsing history. Besides, the data-driven AI algorithms can choose an optimal delivery route for takeaway services to minimize users' waiting time. Moreover, Meituan integrates such data capabilities into digital solutions provided for industrial partners, thus not only enabling firms' operational activities, such as marketing, delivering, financing, IT system, and supply chain management, but also enhancing user experience comprehensively from the business ecosystem. Specifically, Meituan adjusted its organizational structure after its listing, by setting up two business groups – "To Store" and "To Home", establishing a "Quick Burro" business unit that provides supply chain services for platform-embedded firms, and adjusting the "Small Elephant" business unit to "Meituan Selection" that covers the fresh retail market, Meituan focuses its strategy on "Food+Platform". Meanwhile, Meituan set up a user platform responsible for user acquisition and growth and a LBS (Location-Based Services) platform responsible for the development of businesses related to LBS. Under the new organization structure, Meituan integrates the industrial chain through technological enablement to provide better services for consumers and improve the efficiency of the supply side as well. With the vision of satisfying users' demands in different contexts, Meituan has developed a wide range of life-service businesses (e.g., food, accommodation, entertainment, community group buying, etc.). Just as Wang Xing – the founder and CEO of Meituan – said, *"'Eating' will be our core business and the most important category, we will continue to invest in 'eating', continue to do more in going upstream of the industrial chain and connecting the value chain... Meituan wants to be another Amazon in the field of life services"*. Therefore, we suggest:

Proposition 1

Experience-oriented business models (EBM), which are associated with high velocity of big data, high novelty and high efficiency of business model design, can lead to high customer performance in data value. The underlying mechanism is that EBM triggers speed of prediction, performance expectancy, effort expectancy, and perceived user value of data network effects.

4.2 Operation-enhanced business models (OBM)

Configuration 2a ($\sim \text{Volume} * \sim \text{Variety} * \text{Velocity} * \text{Efficiency}$) is labelled as operation-enhanced business models (OBM), which indicate that high customer performance can be realized when big data is processed rapidly and business model design is featured with efficiency, even though the volume and variety of big data are not high, operation is put in the first place. According to data network effects, prediction speed positively affects perceived user value (Gregory et al., 2021), which underlines the interaction of big data velocity and business model efficiency to offset the shortage of data resources. Specifically, constructing data value chain is important to realize data value. That is, data need to be collected, cleaned, stored, analyzed, and applied in a closed loop, the faster its cycling speed is, the more value this process generates (Manyika et al., 2011; Mikalef et al., 2020; McAfee & Brynjolfsson, 2012). Meanwhile, the rapid flow of data can also strengthen business model efficiency as well as effort expectancy in data network effects. On the one hand, it will be easier for different departments in a firm to coordinate with each other due to data linkages (Davenport, 2014; Merendino et al., 2018; Manyika et al., 2011), and firms are able to cope with market changes more flexibly after building agile organizations (Hajli et al., 2020; Ghezzi & Cavallo, 2020). On the other hand, transaction costs (e.g., search, inventory, marketing, communication, etc.) will be reduced and transaction processes will be simplified when data are easily available and timely processed, thus facilitating the use of products or services and promoting customer experience (Holmlund et al., 2020; Grover et al., 2018; Davenport, 2014).

A representative case of OBM is Convenient Bee. Founded in December 2016, Convenient Bee is a new retail firm as well as a new convenience store with the driving force of digital technologies, providing consumers with high-quality products and efficient services. With the core competitiveness of “*lower labor cost, faster expansion speed, and more standardized operation*”, Convenient Bee uses “digital management” to solve the problem of high labor costs and high rent costs in traditional convenience stores. Specifically, Convenient Bee adopts SOP (Standard Operating Procedure) in all businesses, which leads to automatic management that relies more on the system than on the human, thus bringing lots of advantages. On the one hand, through intelligent ordering system, big data selection system, self-service cashier system, dynamic pricing system, and other algorithm-driven programs, Convenient Bee frees employees from ordering, selecting, cashier, and other trivial tasks, which leads to higher performance with less staff. Meanwhile, Convenient Bee completes the training for a store assistant in five days, and a store manager can be basically competent after six months of training. By contrast, Japanese convenience stores, represented by 7-Eleven, FamilyMart, and Lawson, rely more on humans for management, and their training costs are high, which cannot support rapid expansion. On the other hand, all decisions of Convenient Bee are made by data and algorithms, all employees in the business chain should follow the guidance of the system and rely on big data or algorithms to avoid the error caused by traditional information transmission, thus making the operation of Convenient Bee more standardized. Therefore, data and algorithms have become the basis for store management and expansion, relevant data show that it took three Japanese convenience store giants — 7-Eleven, FamilyMart, and Lawson — 15, 16, and 23 years, respectively, to open 2,000 stores in Chinese Mainland, while Convenient Bee only

took three years. As Zhuang Chenchao — the founder of Convenient Bee — mentioned, “*On the surface, Convenient Bee is a chain convenience store, but in essence, we are a data technology company. In all aspects of production, transportation, sales, service, etc., ‘digitalization’ has gone deep into the ‘gene’ of Convenient Bee’s development.*” Therefore, we suggest:

Proposition 2

Operation-enhanced business models (OBM), which are associated with high velocity of big data and high efficiency of business model design, even though not high in volume and variety of big data, can still lead to high customer performance in data value. The underlying mechanism is that OBM triggers speed of prediction, effort expectancy, and perceived user value of data network effects.

4.3 Content-centric business models (CBM)

Configuration 3a ($Volume \sim Variety \sim Velocity \sim Novelty$) is labelled as content-centric business models (CBM), which indicate that high customer performance can be realized when large volumes of data are processed rapidly and business model design is featured with novelty, even though the variety of big data is not high, content plays a vital role in this process. According to data network effects, data quantity positively moderates the relationship between prediction speed and perceived user value (Gregory et al., 2021), suggesting that big data’s volume and velocity should be combined to enhance data value as well as prediction accuracy, such a relationship is usually mediated by business model novelty (Olabode et al., 2022; Wiener et al., 2020), often in the form of digital content (Subramanian, Mitra, & Ransbotham, 2021). Specifically, based on the data value chain, tremendous data can be turned into tremendous value only when they are timely refined from data (information) to wisdom (knowledge) (Davenport, 2014; Simsek et al., 2019). The high-value data are embedded into business models to strengthen novelty and performance expectancy by generating content associated with products or services (Huotari & Ritala, 2021; Subramanian et al., 2021). Content is crucial for firms in that: First, content is more available to users and determines their willingness-to-pay (Subramanian et al., 2021). Second, content provides a platform for users to interact or communicate, which improves user retention and loyalty (Grigorios et al., 2022). Third, users usually have easy access to the content in line with their preferences based on personalized recommendations, thus enhancing user experience (Erevelles et al., 2016; Holmlund et al., 2020).

A representative case of CBM is Xiaohongshu. Founded in June 2013, Xiaohongshu — with the mission of “*inspire lives*” — is a content platform recording life, sharing lifestyles, and interacting with each other through short videos, pictures, and texts. As of October 2023, the number of MAU (Monthly Active Users) of Xiaohongshu has exceeded 200 million, of which 72% are the post-90s generation, and it has continued to grow rapidly. Specifically, Xiaohongshu has two main modules. One is the content community, users can share beauty, sports, tourism, hotels, restaurants, personal care, home furnishing, and other information about consumption experience and lifestyle in Xiaohongshu. Through big data and AI, Xiaohongshu accurately matches the content in the community to users interested in it, thus improving user experience. The other is product e-business, Xiaohongshu analyzes the most popular

products and global shopping trends based on the accumulated overseas shopping data as well as community data, and provides users with the best products from all over the world efficiently and precisely, which is featured with word-of-mouth marketing and big data selection. Content is the most important business element of Xiaohongshu, which brings a lot of structured, semi-structured, and unstructured user data. With tremendous content generated every day, data-driven AI algorithms are adopted to precisely match content and users, which not only enhances user decision efficiency as well as the digitalization of consumption, but also enables brands to grow rapidly. Content also plays a vital role in user conversion and retention, taking live-streaming sales as an example, its conversion rate for KOL (Key Opinion Leader) was 21.4% in April 2020, much higher than that of Douyin and Microblog in the same period (about 10%). As Qu Fang – the founder of Xiaohongshu – said, *“Xiaohongshu is using AI and big data analytics to provide more new scenes for brands that are experienced and shared online, digital technologies have been applied in content recommendation, search, and understanding...More brands will be co-created with users in the future, by connecting users of similar demands, brand value is delivered, which also increases users’ willingness-to-pay and reaches an all-win situation”*. Therefore, we suggest:

Proposition 3

Content-centric business models (CBM), which are associated with high volume and high velocity of big data, and high novelty of business model design, even though not high in variety of big data, can still lead to high customer performance in data value. The underlying mechanism is that CBM triggers data quantity, speed of prediction, accuracy of prediction, performance expectancy, and perceived user value of data network effects.

4.4 Scale-dominant business models (SBM)

Configuration 4a (*Volume*Velocity*Novelty*Efficiency*) is labelled as scale-dominant business models (SBM), which indicate that high firm performance can be realized when large volumes of data are processed rapidly and business model design is featured with both novelty and efficiency, this process highlights the scale advantage. Compared with EBM, SBM owns more data resources (or data quantity, which contributes to a higher level of prediction accuracy in data network effects), indicating that SBM attracts more users and brings more frequent use. It is a unique competitive advantage for internet firms in the digital era, because growth (especially user scale growth) is very important for these internet firms to occupy the market and obtain financing in the early development (Guo et al., 2020; Ghezzi & Cavallo, 2020). Once the scale advantage is formed, firms can seek various ways to monetize (e.g., advertising, freemium, subscription, etc.) and reduce cost through scale economics on both supply and demand sides as well (Tidhar & Eisenhardt, 2020; Johnson et al., 2017; Huotari & Ritala, 2021; Subramanian et al., 2021), which not only calls for the big data analytics capabilities to process tremendous data timely, but also requires the business model to be novel and efficient (Haftor et al., 2021). On the one hand, novelty ensures the offerings (i.e., products, services, and content) to users are new – making a difference in users’ mind, and so are the linkages (e.g., customer relationships, key partnerships, channels, etc.) among stakeholders – breaking existing bottlenecks of transaction (Zott & Amit, 2007, 2008;

Ghasemaghaei & Calic, 2019, 2020). On the other hand, efficiency facilitates value capture by increasing revenue streams (including the main business revenue and diversified business revenue) and optimizing cost structure (including fixed costs and variable costs), thus ensuring the economic sustainability of business models (Zott & Amit, 2007, 2008; Knudsen et al., 2021).

A representative case of SBM is Pinduoduo. Founded in September 2015, Pinduoduo – with the branding “*cheaper when buying together*” – is a new e-business platform dedicated to providing affordable products and interesting shopping experiences for users and is characterized by socialization. After seven years of development, Pinduoduo has acquired more than 860 million annual active buyers and 9 million active merchants, with an annual GMV (Gross Merchandise Volume) of 375 billion dollars and a net profit of 1.2 billion dollars, ranking the second among Chinese e-business platforms (as for active buyers) and the fifth among Chinese internet firms (as for market value). Behind the rapid growth of Pinduoduo, AI, big data, and other digital technologies have contributed a lot. As the pioneer of new e-business, Pinduoduo is committed to creating sustainable value and innovative experiences for most users by adding “more benefits” and “more fun”. Specifically, driven by the distributed intelligent agent network, Pinduoduo integrates fun into every shopping link while providing cost-effective products through the combination of “Costco” and “Disney”. By cooperating with many scientific research institutions worldwide to jointly promote the development of distributed AI technology, Pinduoduo has become a data/AI-driven e-business platform and turned the searching logic of e-business from “customers searching for products” into “products searching for customers”. It is also the innovative searching logic that becomes the key momentum for Pinduoduo’s rapid growth, which explains why the search bar is not apparent in Pinduoduo’s APP, because big data and AI algorithms are used to continuously understand users’ contextual demands so that personalized matching and recommendation can be conducted efficiently and precisely. Just as Huang Zheng – the founder and former CEO of Pinduoduo – mentioned, “*Based on distributed AI technologies, Pinduoduo creates a path of differentiation and individuation in e-business industry. That is, Pinduoduo is able to change the searching logic of e-business completely. With the increasing number of active users, Pinduoduo will keep improving AI engines to match various products and users’ diversified demands precisely*”. Therefore, we suggest:

Proposition 4

Scale-dominant business models (SBM), which are associated with high volume and high velocity of big data, high novelty and high efficiency of business model design, can lead to high firm performance in data value. The underlying mechanism is that SBM triggers data quantity, speed of prediction, accuracy of prediction, performance expectancy, effort expectancy, and perceived user value of data network effects.

4.5 Business models deficient in data-business synergy

By contrast, configuration 1b-9b are collectively labelled as business models deficient in data-business synergy, which indicate that big data and business model design must be combined to realize data value,

the absence of either party will cause failure (Haftor et al., 2021; Knudsen et al., 2021; Davenport, 2014). Prior studies have emphasized the importance of business models for technical innovation. For example, Chesbrough and Rosenbloom (2002) argue that the business model unlocks latent value from technology. Moreover, some scholars point out that data network effects work well only when combined with business models, because the insights gained from big data need to be implemented through business models, and the implementation of business models generates data to nurture data network effects (Gregory et al., 2021; Knudsen et al., 2021; Sjödin et al., 2021). In our research context, big data should also achieve a synergy with business model design, configuration 1a-4a confirm this point. Configuration 1b-9b, by contrast, are basically absent in either big data (especially velocity) or business model design, under which data network effects can hardly be triggered, causing not-high data value.

The most representative case for business models deficient in data-business synergy is ofo, a bike-sharing platform that is now bankrupt. Founded in March 2014, ofo started its bike-sharing business on the campus of Peking University and later expanded to more universities, which proved a success – by November 2016, ofo had acquired more than 4 million weekly active users, and its user satisfaction ranked the first among Chinese bike-sharing platforms. Then, ofo sought to further expand to cities at home and abroad. Its founder, Dai Wei, said, “*We hope to mobilize idle bicycle resources in cities, promote green and low-carbon travel, improve urban congestion, and make cities more beautiful*”. However, ofo deviated from the original aspiration in the following operation and was totally involved in the market competition through blind expansion and price wars, ignoring the digitalization of bikes, improvement of user experience, and exploration of commercialization. When the market calmed down, ofo was confronted with severe financial pressure and had to scale back business, it conducted a series of radical commercialization to save itself, but these measures were only a drop in the ocean compared with the huge debt. According to the survey of ITJUZI.COM, ofo went bankrupt in October 2020, leaving more than 16 million users’ deposits unrefunded. Therefore, we suggest:

Proposition 5

Business models deficient in data-business synergy, which are absent in either big data (especially velocity) or business model design, will cause not-high data value, because data network effects are not triggered.

5. Discussion

We adopt a configurational approach and the theory of data network effects to investigate how the configurations between big data and business model design affect data value. Table 5 shows the summary of configurations.

Table 5
Summary of configurations

Labels	Conditions	Data value	Data network effects elements
Experience-oriented business models (EBM)	Velocity*Novelty*Efficiency	High customer performance	Speed of prediction, performance expectancy, effort expectancy, perceived user value
Operation-enhanced business models (OBM)	~Volume*~Variety*Velocity*Efficiency	High customer performance	Speed of prediction, effort expectancy, perceived user value
Content-centric business models (CBM)	Volume*~Variety*Velocity*Novelty	High customer performance	Data quantity, speed of prediction, accuracy of prediction, performance expectancy, perceived user value
Scale-dominant business models (SBM)	Volume*Velocity*Novelty*Efficiency	High firm performance	Data quantity, speed of prediction, accuracy of prediction, performance expectancy, effort expectancy, perceived user value
Business models deficient in data-business synergy	Absent in either big data (especially velocity) or business model design	Not-high data value	N/A

The negate sign (~) indicates the absence of a condition.

Specifically, there are four configurations leading to high data value (seen as prototypes of data-driven business models), including three configurations of high customer performance, which are respectively labelled as experience-oriented business models (EBM), operation-enhanced business models (OBM), and content-centric business models (CBM), and one configuration of high firm performance, which is labelled as scale-dominant business models (SBM). Extant studies usually examine how the dimensions of big data or the themes of business model design affect data value in a separate and linear manner (Olabode et al., 2022; Cappa et al., 2021; Ghasemaghaei & Calic, 2019, 2020). Our study reveals the configurational results of big data and business model design, highlighting that any single condition is not enough for creating data value, their joint effects matter. According to the theoretical framework of data network effects put forward by Gregory et al. (2021), these four configurations can realize data value because they include some of the key elements of data network effects triggered by data-business

synergy. By contrast, those configurations leading to not-high data value (including not-high customer performance and not-high firm performance) are collectively labelled as business models deficient in data-business synergy, which indicates that the absence in either big data (especially velocity) or business model design will not bring data network effects (Haftor et al., 2021; Knudsen et al., 2021), thus causing failure in realizing data value (Wiener et al., 2020; Sorescu, 2017; Olabode et al., 2022).

Based on these findings, we next present theoretical implications, practical implications, and future research directions.

5.1 Theoretical implications

Our study adds to the existing literature on big data, data-driven business models, and data network effects in several ways. First, we offer an integrated view of big data by clarifying the relationship among big data's 4Vs (i.e., volume, variety, velocity, and value) and business model design, which responds to the call for deepening big data research in management and business (Simsek et al., 2019). Specifically, extant studies usually propose the dimensions of big data separately and are abundant in technical discussions (e.g., algorithms), a comprehensive framework of big data is absent in management and business (Lee, 2017; Ghasemaghahi & Calic, 2019, 2020). We suggest that, among big data's 4Vs, volume, variety, and velocity are the antecedents of value, and we further build a framework integrating business model design into big data's 4Vs to provide a more comprehensive view of big data. Moreover, our study indicates that, as the only necessary condition, velocity of big data is the crucial determinant of high data value. Many scholars have also found that big data analytics, whose core concept is in line with velocity, positively affect performance through the linear regression approach (Wamba et al., 2017; Olabode et al., 2022). Nevertheless, our study argues that velocity alone is not enough, and the holistic configurational approach reveals its joint effects, which contributes to the research on the relationship between big data analytics and performance. In addition, we measure big data's 4Vs by objective indices of each case's APP from Android Market, which also makes a difference in the mainstream studies relying on subjective survey-based measures (Cappa et al., 2021).

Second, we identify different prototypes of data-driven business models from the configurations between big data and business model design, which responds to the call for advancing data-driven business model research in a holistic and systematic way (Sorescu, 2017; Park & Mithas, 2020). Specifically, scholars conduct a lot of research on how digital technologies (especially big data) change the components of business models to generate data-driven business models (Wiener et al., 2020; Trischler & Li-Ying, 2022), most of which are based on the framework of value creation, value delivery, and value capture (Sorescu, 2017; Sjödin et al., 2021). However, business models are defined as an activity system (Amit & Zott, 2001; Zott & Amit, 2010), any single componential change may not fully explain data-driven business models, such deficiency can be compensated by holistic themes as well as prototypes. Therefore, our study adopts novelty and efficiency as the two critical themes of business model design, after combining with the dimensions of big data (i.e., volume, variety, and velocity) through fsQCA, 13 configurations affecting data value appear — four for high data value (three for high customer

performance, one for high firm performance), and nine for not-high data value. Among them, the four configurations leading to high data value are seen as different prototypes of data-driven business models, which are respectively labelled as experience-oriented business models (EBM), operation-enhanced business models (OBM), content-centric business models (CBM), and scale-dominant business models (SBM). Thus, we address the gap of systematically researching data-driven business models through the typological lens and holistic configurational approach.

Third, we illustrate the underlying mechanisms behind the configurations and their impact on data value by data network effects, which responds to the call for extending the theoretical boundary and research context of data network effects (Gregory et al., 2021, 2022). Specifically, as a quite new theory, data network effects have attracted scholars' interest in recent years with the rapid development and wide application of digital technologies such as big data and AI (Haftor et al., 2021). Building on the standard network effects, data network effects emphasize the boosting role of data-driven learning and improvements realized with AI, arguing that the more data generated, the more value provided. Gregory et al. (2021) take the lead in comprehensively introducing the theory of data network effects by demonstrating AI capabilities' impact on perceived user value, such a relationship is moderated by data stewardship, user-centric design, and platform legitimation. Despite the critical scrutiny raised by Clough and Wu (2022), the theory of data network effects still shows a huge application potential, some studies have already followed its core concepts to investigate how users and products are digitalized in the digital era (Zhang & Xiao, 2020; Xie et al., 2016), but it is far from enough. Therefore, we adopt data network effects to reveal the underlying mechanisms behind different configurations and the corresponding impact on data value, making advancements in two aspects. On the one hand, we broaden the theoretical boundary of data network effects by considering not only value creation (i.e., customer performance) but also value capture (i.e., firm performance). On the other hand, we enrich the research context of data network effects by combining business model design and big data's 4Vs in an integrated framework and empirically testing it.

5.2 Practical implications

Based on the findings of our study, there are several practical implications particularly conducive to firms seeking digitalization in the digital era. First of all, the essence of firms' digitalization is to coordinate the relationship between big data and business model design, its success depends on the data-business synergy that leads to high data value (including high customer performance and high firm performance), otherwise not-high data value will be incurred, causing failure for firms seeking digitalization (see configuration 1b-9b, which are collectively labelled as business models deficient in data-business synergy). Then, building on the data-business synergy, firms should give priority to satisfying the diversified demands of customers. That is, customer performance is prior to firm performance, and promoting customer performance is the prerequisite for realizing high data value. By focusing on experience, operation, and content, firms can adopt multiple business models to increase customer performance (see configuration 1a-3a, which are respectively labelled as experience-oriented business models, operation-enhanced business models, and content-centric business models). Finally, firm

performance can be improved when customer performance is high (i.e., firm performance is based on customer performance), user scale acts as the primary determinant in this process, which indicates that firms have a variety of ways to monetize (e.g., advertising, freemium, subscription, etc.) once they acquire enough users as well as user data, thus fully realizing data value (see configuration 4a, which is labelled as scale-dominant business models).

5.3 Limitations and future research

Despite the contributions of our study, there are still several limitations for future research to address. First, model complexity increases exponentially by each new condition in fsQCA (the number of combinations/configurations is 2^n if the number of antecedent conditions is n) (Ragin, 2008). Due to the limited sample size, we only chose five antecedent conditions – volume, variety, and velocity for big data, novelty and efficiency for business model design – to keep the model parsimonious, which might omit some other important conditions (such as veracity, visualization, and variability for big data, complementarity and lock-in for business model design). Also, we divided the outcome – data value – into customer performance and firm performance, but when considering sustainability concerning grand challenges, such division is not enough, social and environmental performance can be taken into account (El-Haddadeh, Osmani, Hindi, & Fadlalla, 2021). Future studies could extend the antecedent conditions as well as the outcome.

Second, compared with the theoretical framework put forward by Gregory et al. (2021), data network effects are only partially elaborated and tested in our study, because we mainly focused on the variables such as AI capabilities, data stewardship, user-centric design, and perceived user value, leaving platform legitimation unconsidered. Meanwhile, although we integrated business model design into big data to examine not only value creation (i.e., customer performance) but also value capture (i.e., firm performance), which extends its theoretical boundary and research context to a certain degree, it is still inadequate to deliver solid answers to the arguments made by Clough and Wu (2022). Future studies could further advance the theory of data network effects in more contextual circumstances.

Third, we assessed novelty and efficiency of business model design by researchers' ratings, although considerable efforts were undertaken to ensure reliability and validity, the potential subjective biases might still exist. Future studies could draw on machine learning techniques, such as content analysis and natural language processing, to analyze case materials, which allows to prevent potential biases regarding manual ratings (van Rijmenan et al., 2019).

Finally, we did not include the notion of time in our study, which is also one of the deficiencies of the fsQCA method (Ragin, 2008; Fiss, 2011). Future studies could use longitudinal case studies to uncover the process of how different configurations between big data and business model design come into being and evolve.

6. Conclusion

Given the importance of big data for firms' digitalization, a comprehensive understanding of how big data configures business model design is necessary for both scholars and practitioners. Based on the existing literature, we create a framework that integrates business model design (novelty, efficiency) into big data's 4Vs (volume, variety, velocity, value) and use fsQCA to investigate how the configurations among volume, variety, velocity, novelty, and efficiency affect data value (customer performance, firm performance). Our findings indicate that velocity is the necessary condition for high data value, but velocity alone is not enough, it only works in combination with other conditions. Overall, we identify different prototypes of data-driven business models from the configurations leading to high data value, which are respectively labelled as experience-oriented business models (EBM), operation-enhanced business models (OBM), content-centric business models (CBM), and scale-dominant business models (SBM). By contrast, configurations leading to not-high data value are collectively labelled as business models deficient in data-business synergy. Our study advances the interaction between big data and business model design as well as the extension of data network effects' theoretical boundary and research context. We suggest future research deepen this study by adopting a larger sample and taking more conditions into account, thus providing more insights for scholars and practitioners.

Declarations

Author Contribution

The author confirms sole responsibility for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Specifically, the author was responsible for the conceptualization of the study, design of the research methodology, collection and analysis of data, drafting of the original manuscript, and final review and approval of the manuscript for submission.

Data Availability

The dataset and QCA software used in this study are uploaded as the related files.

References

1. Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal*, 22(6-7), 493-520.
2. Cappa, F., Oriani, R., Peruffo, E., & McCarthy, I. (2021). Big data for creating and capturing value in the digitalized environment: Unpacking the effects of volume, variety, and veracity on firm performance. *Journal of Product Innovation Management*, 38(1), 49-67.
3. Chesbrough, H., & Rosenbloom, R. S. (2002). The role of the business model in capturing value from innovation: Evidence from Xerox corporation's technology spin-off companies. *Industrial and Corporate Change*, 11(3), 529-555.

4. 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.
5. Clauss, T. (2017). Measuring business model innovation: Conceptualization, scale development, and proof of performance. *R & D Management*, 47(3), 385-403.
6. Clough, D. R., & Wu, A. (2022). Artificial intelligence, data-driven learning, and the decentralized structure of platform ecosystems. *Academy of Management Review*, 47(1), 184-189.
7. Davenport, T. H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Boston: Harvard Business Review Press.
8. El-Haddadeh, R., Osmani, M., Hindi, N., & Fadlalla, A. (2021). Value creation for realising the sustainable development goals: Fostering organisational adoption of big data analytics. *Journal of Business Research*, 131, 402-410.
9. Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897-904.
10. Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal*, 54(2), 393-420.
11. Foss, N. J., & Saebi, T. (2017). Fifteen years of research on business model innovation: How far have we come, and where should we go? *Journal of Management*, 43(1), 200-227.
12. Ghasemaghaei, 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.
13. Ghasemaghaei, M., & Calic, G. (2020). Assessing the impact of big data on firm innovation performance: Big data is not always better data. *Journal of Business Research*, 108, 147-162.
14. Ghezzi, A., & Cavallo, A. (2020). Agile business model innovation in digital entrepreneurship: Lean startup approaches. *Journal of Business Research*, 110, 519-537.
15. Greckhamer, T. (2016). CEO compensation in relation to worker compensation across countries: The configurational impact of country-level institutions. *Strategic Management Journal*, 37(4), 793-815.
16. Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, S. H. (2021). The role of artificial intelligence and data network effects for creating user value. *Academy of Management Review*, 46(3), 534-551.
17. Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2022). Data network effects: Key conditions, shared data, and the data value duality. *Academy of Management Review*, 47(1), 189-192.
18. Grigorios, L., Magrizos, S., Kostopoulos, I., Drossos, D., & Santos, D. (2022). Overt and covert customer data collection in online personalized advertising: The role of user emotions. *Journal of Business Research*, 141, 308-320.
19. Grover, V., Chiang, R. H. L., Liang, T. P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), 388-423.

20. Guo, H., Wang, C., Su, Z., & Wang, D. (2020). Technology push or market pull? Strategic orientation in business model design and digital start-up performance. *Journal of Product Innovation Management*, 37(4), 352-372.
21. Haftor, D. M., Climent, R. C., & Lundstrom, J. E. (2021). How machine learning activates data network effects in business models: Theory advancement through an industrial case of promoting ecological sustainability. *Journal of Business Research*, 131, 196-205.
22. Hajli, N., Tajvidi, M., Gbadamosi, A., & Nadeem, W. (2020). Understanding market agility for new product success with big data analytics. *Industrial Marketing Management*, 86, 135-143.
23. Holmlund, M., Van Vaerenbergh, Y., Ciuchita, R., Raval, A., Sarantopoulos, P., Ordenes, F. V., & Zaki, M. (2020). Customer experience management in the age of big data analytics: A strategic framework. *Journal of Business Research*, 116, 356-365.
24. Huotari, P., & Ritala, P. (2021). When to switch between subscription-based and ad-sponsored business models: Strategic implications of decreasing content novelty. *Journal of Business Research*, 129, 14-28.
25. Johnson, J. S., Friend, S. B., & Lee, H. S. (2017). Big data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process. *Journal of Product Innovation Management*, 34(5), 640-658.
26. Knudsen, E. S., Lien, L. B., Timmermans, B., Belik, I., & Pandey, S. (2021). Stability in turbulent times? The effect of digitalization on the sustainability of competitive advantage. *Journal of Business Research*, 128, 360-369.
27. Kumar, S., Sahoo, S., Lim, W. M., Kraus, S., & Bamel, U. (2022). Fuzzy-set qualitative comparative analysis (fsQCA) in business and management research: A contemporary overview. *Technological Forecasting and Social Change*, 1-23. <https://doi.org/10.1016/j.techfore.2022.121599>.
28. Laval, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *Mit Sloan Management Review*, 52(2), 21-32.
29. Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293-303.
30. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*.
31. McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 1-9.
32. Merendino, A., Dibb, S., Meadows, M., Quinn, L., Wilson, D., Simkin, L., & Canhoto, A. (2018). Big data, big decisions: The impact of big data on board level decision-making. *Journal of Business Research*, 93, 67-78.
33. Mikalef, P., Pappas, I. O., Krogstie, J., & Pavlou, P. A. (2020). Big data and business analytics: A research agenda for realizing business value. *Information & Management*, 57(1), 1-6.
34. Misangyi, V. F., Greckhamer, T., Furnari, S., Fiss, P. C., Crilly, D., & Aguilera, R. (2017). Embracing causal complexity: The emergence of a neo-configurational perspective. *Journal of Management*,

- 43(1), 255-282.
35. Olabode, O. E., Boso, N., Hultman, M., & Leonidou, C. N. (2022). Big data analytics capability and market performance: The roles of disruptive business models and competitive intensity. *Journal of Business Research*, 139, 1218-1230.
 36. Park, Y., & Mithas, S. (2020). Organized complexity of digital business strategy: A configurational perspective. *Mis Quarterly*, 44(1), 85-127.
 37. Priem, R. L., Butler, J. E., & Li, S. (2013). Toward reimagining strategy research: Retrospection and prospection on the 2011 AMR decade award article. *Academy of Management Review*, 38(4), 471-489.
 38. Priem, R. L., Wenzel, M., & Koch, J. (2018). Demand-side strategy and business models: Putting value creation for consumers center stage. *Long Range Planning*, 51(1), 22-31.
 39. Ragin, C. C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago, IL: University of Chicago Press.
 40. Simsek, Z., Vaara, E., Paruchuri, S., Nadkarni, S., & Shaw, J. D. (2019). New ways of seeing big data. *Academy of Management Journal*, 62(4), 971-978.
 41. Sjödin, D., Parida, V., Palmie, M., & Wincent, J. (2021). How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. *Journal of Business Research*, 134, 574-587.
 42. Sorescu, A. (2017). Data-driven business model innovation. *Journal of Product Innovation Management*, 34(5), 691-696.
 43. Subramanian, H., Mitra, S., & Ransbotham, S. (2021). Capturing value in platform business models that rely on user-generated content. *Organization Science*, 32(3), 804-823.
 44. Suoniemi, S., Meyer-Waarden, L., Munzel, A., Zablah, A. R., & Straub, D. (2020). Big data and firm performance: The roles of market-directed capabilities and business strategy. *Information & Management*, 57(7), 1-17.
 45. Tan, K. H., & Zhan, Y. (2017). Improving new product development using big data: A case study of an electronics company. *R & D Management*, 47(4), 570-582.
 46. Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning*, 43(2-3), 172-194.
 47. Tidhar, R., & Eisenhardt, K. M. (2020). Get rich or die trying... finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal*, 41(7), 1245-1273.
 48. Trischler, M. F. G., & Li-Ying, J. (2022). Digital business model innovation: Toward construct clarity and future research directions. *Review of Managerial Science*, 1-30.
<https://doi.org/10.1007/s11846-021-00508-2>.
 49. Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from big data: A multiple-case study analysis of provider companies. *Technovation*, 84-85, 21-36.

50. van Rijmenan, M., Erekhinskaya, T., Schweitzer, J., & Williams, M.-A. (2019). Avoid being the Turkey: How big data analytics changes the game of strategy in times of ambiguity and uncertainty. *Long Range Planning*, 52(5), 1-21.
51. Velu, C. (2015). Business model innovation and third-party alliance on the survival of new firms. *Technovation*, 35, 1-11.
52. Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, 70, 356-365.
53. Wang, P. (2021). Connecting the parts with the whole: Toward an information ecology theory of digital innovation ecosystems. *Mis Quarterly*, 45(1), 397-422.
54. Weill, P., & Woerner, S. L. (2013). Optimizing your digital business model. *MIT Sloan Management Review*, 54(3), 71-78.
55. Wielgos, D. M., Homburg, C., & Kuehn, C. (2021). Digital business capability: Its impact on firm and customer performance. *Journal of the Academy of Marketing Science*, 49(4), 762-789.
56. Wiener, M., Saunders, C., & Marabelli, M. (2020). Big-data business models: A critical literature review and multi-perspective research framework. *Journal of Information Technology*, 35(1), 66-91.
57. 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.
58. Zhang, H., & Xiao, Y. (2020). Customer involvement in big data analytics and its impact on B2B innovation. *Industrial Marketing Management*, 86, 99-108.
59. Zott, C., & Amit, R. (2007). Business model design and the performance of entrepreneurial firms. *Organization Science*, 18(2), 181-199.
60. Zott, C., & Amit, R. (2008). The fit between product market strategy and business model: Implications for firm performance. *Strategic Management Journal*, 29(1), 1-26.
61. Zott, C., & Amit, R. (2010). Business model design: An activity system perspective. *Long Range Planning*, 43(2-3), 216-226.

Figures

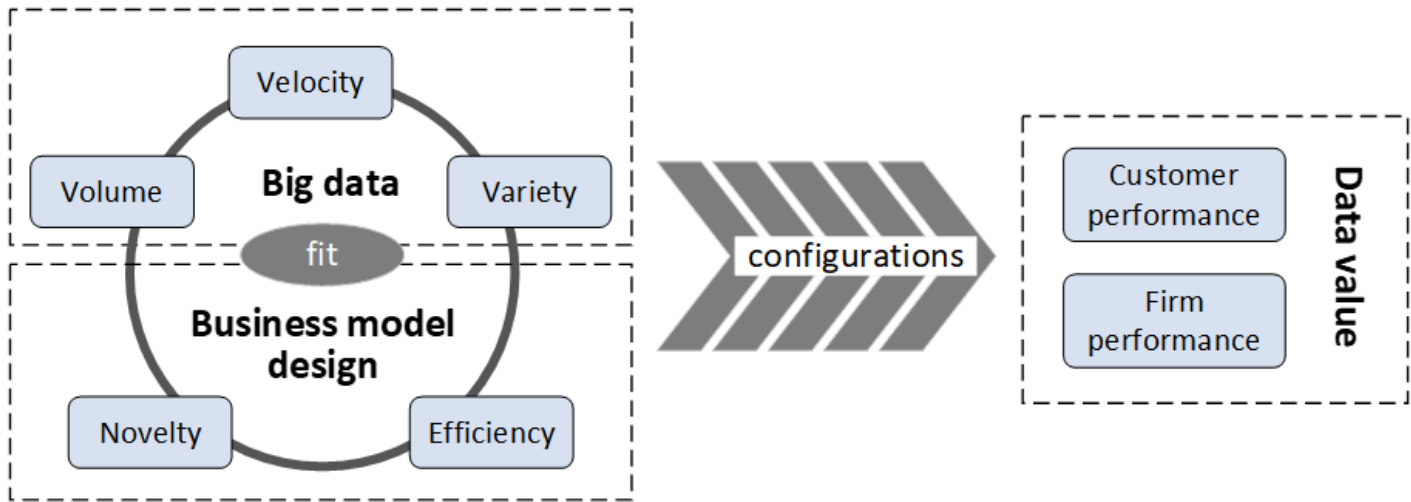


Figure 1

Research framework