2-2019

How to derive causal insights for digital commerce in China? A research commentary on computational social science methods

David C.W. PHANG

Kanliang WANG

Qiu-hong WANG
Singapore Management University, qiuhongwang@smu.edu.sg

Robert John KAUFFMAN
Singapore Management University, rkauffman@smu.edu.sg

Maurizio NALDI

DOI: https://doi.org/10.1016/j.elerap.2019.100837

Electronic Commerce Research and Applications 35 (2019) 100837

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

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

How to derive causal insights for digital commerce in China? A research commentary on computational social science methods

David C.W. Phang[a], Kanliang Wang[b], Qiuhong Wang[c], Robert J. Kauffman[c,*], Maurizio Naldi[d]

[a] University of Nottingham, Ningbo, China
[b] Renmin Business School, Renmin University, China
[c] School of Information Systems, Singapore Management University, Singapore
[d] Dept. of Civil Engineering and Computer Science, University of Rome Tor Vergata, Italy

ARTICLE INFO

Keywords:
Big data
Business insights
Causal inference
Causal methods
Computational social science (CSS)
Consumer behavior
China
Data analytics
Digital economy
E-commerce
Emerging markets
Empirical research
Information systems (IS) research
Machine learning (ML)
M-commerce
Policy analytics
Research design
Secondary data
Sensor data
Streaming data
Social insights
Theory testing

ABSTRACT

The transformation of empirical research due to the arrival of big data analytics and data science, as well as the new availability of methods that emphasize *causal inference*, are moving forward at full speed. In this Research Commentary, we examine the extent to which this has the potential to influence how e-commerce research is conducted. China offers the ultimate in data-at-scale settings, and the construction of real-world natural experiments. Chinese e-commerce includes some of the largest firms involved in e-commerce, mobile commerce, social media and social networks. This article was written to encourage young faculty and doctoral students to engage in research that can be carried out in near real-time, with truly experimental or quasi-experimental research designs, and with the clear intention of establishing causal inferences that relate the precursors and drivers of observable outcomes through various kinds of processes. We discuss: the relevant data sources and research contexts; the methods perspectives that are appropriate which blend Computer Science, Statistics and Econometrics, how the research can be made relevant for China; and what kinds of findings and research directions are available. This article is not a tutorial on big data analytics methods in general though, nor does it cover just those published works that demonstrate big data methods and empirical causality in other disciplines. Instead, the empirical research covered is mostly taken from *Electronic Commerce Research and Applications*, which has published many articles on Chinese e-commerce. This Research Commentary invites researchers in China and the Asia Pacific region to expand their coverage to bring into their empirical work the new methods and philosophy of causal data science.

1. Introduction

The phenomenon of big data is transforming the business landscape, impacting both the input data and the methods adopted for their analysis. The availability of unprecedented volume, velocity, and variety of primary data concerning individual consumers allows the extraction of information about consumers at a much wider and deeper level than before (Erevelles et al., 2016). The growth of available data, often of an unstructured nature, also calls for a new generation of analysis tools, turning investigation mode from deductive to inductive (Kitchin, 2014). A major class of data analysis today is represented by machine learning (ML) algorithms (Witten et al., 2016), and equally important is the use of advanced econometrics methods that support causal inference in natural experimental settings.

At the same time, a relevant geopolitical change is represented by the emergence of China as a major technology and market player. The synergic combination of these two factors – the increased availability of data and data analysis tools and the technological and market size growth – may fuel an extremely powerful research effort. In this article, we propose a view of the current research landscape on big data in China and highlight promising research paths for the near future involving causal inference.

* Corresponding author.
E-mail addresses: cheewei.phang@nottingham.edu.cn (D.C.W. Phang), wangkanliang@rmbs.ruc.edu.cn (K. Wang), qiuhongwang@smu.edu.sg (Q. Wang), rkauffman@smu.edu.sg (R.J. Kauffman), maurizio.naldi@uniroma2.it (M. Naldi).

https://doi.org/10.1016/j.elerap.2019.100837
Received 29 November 2018; Received in revised form 16 February 2019; Accepted 17 February 2019
Available online 18 February 2019
1567-4223/

1.1. Research questions

With this general background in mind, we will explore three different research questions: (1) what are some of the core research opportunities in China and the Asia Pacific region for conducting empirical research on digital commerce that can take advantage of the new paradigm for causal inference that has become available; (2) what specific issues can be explored differently and more effectively than in the past; and (3) generally speaking, how does the relatively new availability of big data analytics methods that blend machine-based and statistical methods in experimental settings change the capabilities of research to explore and understand the inner workings of their research settings in greater depth? This direction in research is much different from prior emphases in data analytics in recent decades that centered on *data mining* (DM) and *online analytical processing* (OLAP), for which there was not specific effort put into the design and leveraging of natural experiments and causal inference approaches.

1.2. Research opportunities in China's emerging markets

A huge and growing population base of 1.417 billion people includes a middle class of 430 million today and is expected to include 780 million in the mid-2020s (Kharas, 2017). The government-supported transportation and Internet infrastructure is further expected to provide capacity and sustainability for the development of the digital economy in China's markets. Along with the penetration of the digital economy, the new technological transition is helping to connect the lives of billions of people and collect the digital traces of their production and consumption activities. This digitalization process generates massive data, and hence enables the scientific investigation into areas where it would be infeasible in the traditional economy.

The attractiveness of the China's emerging markets to IS researchers also is related to the following aspects of China's recent history: geographical imbalances; and rapid technological innovation; and the growth of digital and high-tech entrepreneurship. First, disparities in the economy across broad geographical regions now provide unprecedented opportunities to study the impact of policies and information technologies (ITs) as an enabler in markets, and for improving the education and healthcare environments across regions. This is not limited to the expected disparity between the rural and urban areas. Longitudinal differences also exist, with China's economy in the western area being less developed than in the eastern coastal area. Even within the same geographical region, cities are classified by various media publications into different tiers according to the development of their economy, their commercial market resources, education and healthcare environment, and so on.[1]

For example, Tier 1 and Tier 2 cities historically pioneered Internet penetration and online shopping. The driving forces toward the digital economy are shifting to the Tier 3 and Tier 4 cities, due to the rapid expansion of their Internet and smartphone user bases, and their growing disposable income levels. In fact, in Tier 3 and Tier 4 cities, the business involving digital economy may encounter lower-entry barriers and less intensive competition from their counterparts in the less-developed real economy.

Second, China has created the world's largest middle-income class.[2] This has resulted in forces that have influenced the development of global markets. With the increased demand for goods in overseas markets, cross-border e-commerce made up 7% of overall e-commerce spending in 2018 and is expected to exceed US$140 billion by 2021.[3] The size and diversity of China's middle class and elite have created the basis for the growth of diversified digital business models. They provide new products and services, trigger the formation of new business relationships, and result in interesting business and policy issues.

For example, the ownership of user-generated digital items in online communities and virtual games, the liability of Internet and information-related service providers to their subscribers in the protection of intellectual property and cybersecurity, the sustainability of blockchain-related transactions, and so on. Many of these rising issues have never been addressed in past research or are unable to be explained by established theories, and thus invite extensive research with the combination of in-depth domain knowledge and creative data analytics approaches.

Third, China's emerging markets are not just imitating the business models in the U.S. and other established markets. They are moving from imitation to high-tech innovation through technology advances and transformation. Thus, the digital economy in China provides the promising prototypes for researchers to study the potential impact of mechanisms pushed by the next-generations of technology and platforms on markets and industries. These pioneering studies will be able to draw lessons, identify best practices and generate new knowledge for the interests of the other markets in the world. For example, the mobile payment (m-payment) platforms in China (namely AliPay and WeChat Pay) have leapfrogged the traditional credit card-based payment platforms. Following the huge success in China, m-payments have become more and more important in other emerging markets, for example, in India and the Philippines. As the next strategic transformation in China, artificial intelligence (AI)-based information systems (IS) are replacing the role of human beings in warehouse management, payment authentication, and transportation monitoring. It is foreseeable that the interaction and integration between AI-driven systems and humans will generate many research scenarios that deserve investigation by data scientists and IS researchers.

1.3. A proposed research paradigm for data analytics in Chinese e-commerce

In recent years, Computational Social Science (CSS) on the one hand, and DM, ML and Data Science on the other hand, have all risen rapidly as the new research methods for empirical research. For example, Chang et al. (2014, p. 67) commented in *Decision Support Systems*:

> "*The era of big data has created new opportunities for researchers to achieve high relevance and impact amid changes and transformations in how we study social science phenomena. With the emergence of new data collection technologies, advanced [DM] and analytics support, there seems to be fundamental changes that are occurring with the research questions we can ask, and the research methods we can apply. … The changing costs of data collection and the new capabilities that researchers have to conduct research … suggest the possibility of a scientific paradigm shift toward [CSS]. The new thinking related to empirical regularities analysis, experimental design, and longitudinal empirical research further suggests that these approaches can be tailored for rapid acquisition of big data sets.*"

The combination of new perspectives on empirical research, along with dramatic changes in the philosophy of science for the discovery of new knowledge, play a special role in the broad area of e-commerce. Many potential benefits can be reaped in the collection and analysis of data by combining ML methods from Computer Science (CS), Statistics and Econometrics. Nowadays, data include both structured and an ever-growing portion of unstructured data (textual, image and audio), as well as secondary data in large archives and primary data from streaming digital sources. Those data can be used to establish the basis in overall research design from Social Science, so that causal inferences and original conclusions and insights can be acquired from very large datasets. We believe that learning will occur in the process of collecting

[1] For a classification of Chinese cities into four tiers, see www.chinacheckup.com/blogs/articles/china-city-tiers.

[2] The relevant website is en.people.cn/n3/2018/0314/c90000-9437028.html.

[3] See www.emarketer.com/Chart/Cross-Border-Retail-Ecommerce-Sales-China-2016–2021-billions-change/212844.

data via machine-based methods so that research questions for a given setting can be evaluated with advanced models and computational methods, while preserving and strengthening key elements of theory development, hypothesis testing, and managerial interpretation.

Our hope is that this research commentary article will guide empirical researchers in China and elsewhere in Asia and around the world, to recognize the opportunities to go beyond the currently established methods of survey research of behavioral intentions in various e-commerce contexts and traditional statistical models that only produce association and correlation results (Efron and Hastie, 2016). We encourage them to favor research designs that include experimental and quasi-experimental approaches, as well as the use of ML methods combined with explanatory statistics and econometrics, so that it will be possible to achieve retrospective and predictive insights based on causal inferences and the new ways to work with data. This will permit our *philosophy of science for e-commerce empiricism* to change in step with the times (Chang et al., 2014).

2. Data sources for electronic, mobile and social commerce

We advocate the collection of large-scale datasets to explore the variety of issues that are present in the areas of electronic, mobile and social commerce (e-, m-, and s-commerce). This is natural for the study of these areas in China, where many different kinds of business, consumer and social activities operate on an enormous scale, and create tremendous amounts of data. At the outset of our discussion in this article, we wish to distinguish among the different levels of granularity of large-scale datasets, recalling the *micro–meso–macro data spectrum* introduced in Chang et al. (2014). They noted that:

- *Micro-data* refers to the least aggregated level of data in very large data sets, resulting from the technology-mediated human, social, machine and physical settings. They are essentially atomic data (hence with very fine granularity), which represent the level at which primary data are captured. These include things such as the digital traces of people's movements based on mobile phone geolocation, their social media posts, their channel switching when they are watching TV, digital sensing done via Internet of Things (IoT) devices, monitoring of the movement of sharing economy bicycles, and clickstream data collection during consumer shopping navigation.
- *Meso-data* represent an intermediate level of data aggregation in big data sets. They include tracking the extent of user tweeting and retweeting, in response to issues or the tweets of other users, and the data may still be society-wide in their coverage.
- *Macro-data* provide a view of the phenomenon under study, via: market, regional and geographic areas; industry or economy sectors; and national and international scope. Macro-data include patterns of electricity use, and phone calling patterns of people across different regions. In the context of China, researchers have tracked macro-data related to Chinese New Year season train and other public transportation travel (Prabu, 2014).

We next discuss the different kinds of data sources for the study of issues and problems in China that are available for data analytics research. We adopt both a platform-oriented view (as for social media and mobile platforms) and an industry-oriented one (the case of fintech innovations).

2.1. Electronic markets, Internet-based selling and e-commerce

The past decades have witnessed phenomenal growth of electronic commerce in China. In 2013, the country racked up US$314 billion in online sales, surpassing that (US$255 billion) in the U.S. (Morgan Stanley Research, 2015). This figure has since reached US$1 trillion in 2017, accounting for nearly half of worldwide e-commerce sales, according to eMarketer (Long, 2017).

The rapid growth of e-market and e-commerce activities has allowed to collect massive data concerning consumers' online behavior, which provide unprecedented opportunities to explore the latest theory and practices for e-commerce. Several examples can be mentioned for the micro-level. Huang et al. (2009) employed data from 70 Chinese retailers to develop a performance assessment model for e-commerce. Using clickstream data from a large e-commerce retailer in China, Ding et al. (2015) investigated the relationship between online product page viewing and product returns. Their study shows that, interestingly, the more product pages a consumer views, the greater their likelihood of returning the purchased product to the retailer. Yao et al. (2017) analyzed sales data from a leading e-commerce retailer in China to show that the offering of free product samples to customers on the e-commerce platform can have both immediate and delayed sales effects for the brands offering the samples. In this special issue as well, Tian et al. (2018) unveiled the relationships between weather and consumers' variety-seeking behavior, which have implications for how e-commerce retailers may strategically promote their products based on the weather conditions.

Many insights may be obtained by analyzing data at the macro-level. For example, the aggregated transaction data from Taobao's Tmall.com across different provinces in China may inform how consumers' spending preferences and habits vary between the more developed and less developed areas. China has a large geographical area (9.6 million square kilometers, the world's fourth largest country) and varying states of economic development across the regions. Apart from the disparity between the urban and rural areas, the eastern, coastal regions generally have had better economic development than the western regions (Fan et al., 2011). The emergence of e-commerce has facilitated the flow of goods across the regions, from urban to rural areas and from more developed to less developed regions, and vice versa. Indeed, an important driver of e-commerce growth in China is rural consumers' huge demand for branded and latest consumer goods sold in the megacities, which are difficult to access (or only obtainable at much higher prices) for these consumers prior to e-commerce (Morgan Stanley Research, 2015). At the same time, it also becomes possible for consumers in the urban and more-developed areas to buy native produce and local specialities directly from rural and less-developed areas.

Again at a macro-level, it will be interesting to investigate how e-commerce has impacted the economic growth of the more- and less-developed regions, in particular considering the job market. On one hand, the rise of e-commerce may bring with it job opportunities in the less-developed areas. For instance, it has been reported that there are now 28 million workers in rural areas employed by e-commerce businesses, which is likely to be associated with the expansion of warehouses and pick-up facilities by major Chinese e-commerce retailers, such as Alibaba and JD.com, as they compete to provide fast delivery even to remote areas of China (Tong, 2017). On the other hand, the rise of e-commerce is also argued to lead to the shutdown of many offline shopping malls and retail stores in both the developed and less-developed areas (Xinhua, 2018). Using macro-level transaction data from major e-commerce retailers, insights may thus be obtained with regard to the impacts of e-commerce and the related consumer purchase-making behavior, and inform policy-making for e-commerce development in the developed versus less-developed areas. (Representative e-commerce empirical articles from China and the Asia Pacific region from ECRA are reported in Appendix Table A1.)

2.2. Mobile sensing and m-commerce

Mobile commerce, conducted primarily through smartphones and tablets today, is an extension of e-commerce. Rapid growth has occurred in China in recent years and mobile has become the dominant channel through which consumers buy things online. For instance, KPMG's research (Tong, 2017) has shown that more than 90% of Chinese online consumers have made purchases using their mobile phone, compared with just under 70% globally. A statistic from eMarketer suggests that at least 60% of online sales occur through mobile phones versus PC desktops.

The interest associated with m-commerce is based on the capabilities of mobile technologies to "sense" the environment, that is, context-aware capabilities (Chen and Kotz, 2000). Such capabilities enable the provision of information content based on consumer contexts, or mobile targeting (Andrews et al., 2016; Chen and Yin, 2017; Luo et al., 2013), such as recommending a product that fits consumer needs when they are in a particular environmental context. Two fundamental contextual dimensions have been noted to be key for mobile context-aware sensing capabilities to enable better targeting, in terms of location and time (Balasubramanian et al., 2002). Indeed, being able to receive product information any place, any time is a distinguishing attribute of m-commerce versus e-commerce for its users.

There also has been active research on how mobile capabilities can be better leveraged, involving not only academics but also industry practitioners. For instance, a number of field experiments were conducted with the collaboration of mobile operators in China to explore innovative applications of mobile targeting. Among these studies, Luo et al. (2013) highlighted the need to consider the effect of targeting based on location together with time. Their study shows that, while giving a short lead time can generate higher purchase rates when targeting nearby consumers. For consumers who are farther away (e.g., 1 km), giving a moderate lead time (1 day) can lead to optimal performance compared to giving a short lead time (2 h). Fang et al. (2015), in contrast, showed that mobile targeting is mainly effective in inducing spontaneous, impulse-driven purchases, though it can motivate delayed purchases as well. Fong et al. (2015) tested the idea of sending mobile coupons to consumers close to a competitor's store in a shopping area. They showed that this poaching strategy could be effective in generating incremental sales and not cannibalizing existing margins. This is in contrast to targeting of a firm's own customers.

With the mobile-first consumer behavior in China, m-commerce applications have become increasingly innovative in the country. Data generated on mobile usage can be employed to discover how consumers use mobile devices to navigate apps, consider making purchases, and complete transactions. It may also be interesting to investigate how cultural factors play a part in influencing mobile usage in China, possibly causing differences in the usage behaviors compared to other countries. For instance, Tan et al. (2014) showed that a *high context culture*[4] may explain why consumers in China have a relative preference to receive promotional message via mobile technology (SMS) compared to e-mail. With the availability of objective data on individuals' mobile usage behaviors, issues related to cultural influences may be investigated in a more accurate manner, extending beyond one-time purchase-making to each and every activity along the consumer decision making journey and also across different consumption scenarios. (Representative articles in ECRA on m-commerce in China and the Asia Pacific region are reported in Appendix Table A2.)

2.3. Social media and s-commerce

Social media has enjoyed tremendous growth in China, serving as an avenue for interactions as well as product information. According to the Boston Consulting Group (Walters et al., 2011), more than 40% of Chinese consumers have posted or read reviews on social media websites, a figure close to twice that of the U.S. consumers. Chinese consumers are among the most social shoppers in the world. Other research also corroborates this view. Vuylsteke et al. (2010) showed that, compared to their Western counterparts, Chinese consumers show a particularly high preference for using social media to seek opinions from other consumers rather than experts.

The popularity of social media and its centrality in Chinese consumers' everyday life have spun off innovative social commerce models. One notable instance is that on WeChat, where individual users have been conducting the buying and selling of products with their friends (called "micro merchants" or "wei shang"). A related scenario occurs when a consumer promotes a product the person finds good in her social circle (e.g., on WeChat); if friends show interest by inquiring about the product, the user may act as the seller and sell the product directly to them, especially if the product is not readily accessible in the market. Companies have tapped into this consumer behavior by engaging their customers as direct sellers of their products, with incentives such as receving a commission for successful sales. Such commerce relies on social connections and the social capital possessed by the users.

With data accumulated on the transactions among the users and their social network, interesting research questions can be explored, such as how firms may better leverage their customers' social network to generate sales, what kind of incentives (apart from financial ones) may be used to motivate the users and also their friends to participate in promoting their products, and whether individuals' use of their social network to conduct commercial activities may have negative repercussions such as hurting their social relationships (e.g., causing annoyance among friends). It may also be possible that this s-commerce phenomenon is culturally-related. In particular, China is known to be high on the *collectivist cultural dimension*, which emphasizes interpersonal relationships (Hofstede, 2006). Research ought to investigate how cultural factors may influence this kind of social commerce practice, and if such a practice can be applied in other cultural contexts, fully or partially. (Representative empirical articles in ECRA concerning s-commerce in China and the Asia Pacific region are reported in Appendix Table A3.)

2.4. Sharing economy

Another area of growing interest concerning platforms in China is that of the sharing economy, whose expansion over a number of industries, especially in transportation and lodging (e.g., Uber, Airbnb, Grab Taxi and many other new sharing economy firms), has been described by Clemons et al. (2017). The authors note with respect to these new companies that "[*t*]*hey allow consumers to adjust their consumption to their actual needs, and increase the economic usefulness of their assets*" (p. 430).

China's role in the global sharing economy developments relative to global capitalization in the sector is enormous. *Foreign Affairs* magazine, for example, has estimated the total value of its sharing economy at an astonishing US$520 billion, representing a 103% increase in value compared to the prior year (Yan, 2017). In addition, the *South China Morning Post* reported in March 2018 that there are now 31 *technology unicorns*, or firms with more than US$1 billion in capitalization. One of the most well-known ones includes Didi Chuxing for ride-hailing and taxi services in the sharing economy, Ofo and Mobike for bicycle sharing, Meituan-Dianping for group-buying services like Groupon and for social opinions like Yelp, and the news aggregator Toutiao, among others. Sharing platforms are a natural generator of huge amounts of data concerning both consumer and seller behavior (e.g., the case of Uber, the availability of drivers, the service times, the buyer-seller matching conditions), so this sector can be framed in the overall big data picture too.

[4] This describes an environment in which people value intimate interrelationships, and where information is typically communicated through messages that may be simple but entail deep meanings, making communication more personal.

The opportunities for conducting empirical research in this sector in China are especially interesting because of its rapid growth (which seems to have peaked now), extraordinary tech innovation, and the extent to which the problems to be studied are so interdisciplinary. Research opportunities will come from testing some of the interesting theoretical and model-based assertions by different researchers related to e-commernce that we have seen over the past few years.

For example, research has been conducted to examine what damage happens in the sharing of automobiles and rental properties related to asset use and renter care when a digital intermediary offers optimal insurance contracts and appropriate incentives to consumers (Weber, 2014). Other studies involve the theorized presence in the market of sharing premia on new goods that will be used as sharable assets, largely due to the higher extent of their after-purchase usage, and the need for embedded intelligence in smart sharing products to reduce the coordination costs of after-sales market sharing (Weber, 2017).

In addition, Razeghian and Weber (2019) have studied what happens with bundling, pricing and the profitability of a durable-goods monopolist, as well as consumers' peer-trading propensity (i.e., their willingness to rent as opposed to purchase) increases in an economy. This is especially interesting for empirical research in the China context, where 600 million people made about US$500 billion worth of sharing economy purchases (loosely defined). These numbers contrast with the U.S., where only 55 million people were estimated to use a sharing service in 2017 (Larmer, 2017).

It will be as interesting for empirical researchers who focus on e-commerce, economics, regulation and the law, strategy and management, and operations and information technology (IT) to examine the variety of issues that have arisen around the new business practices of the sharing economy, especially in the operations of its platforms. In addition, we also have to consider that such platforms generate a huge amount of data, and that there are other aspects to consider.

2.5. Fintech innovation

After examining the opportunities concerning the use of platforms (communications platforms, such as the mobile network and social media, and service platform, such as those involved in the sharing economy) we now turn to a specific industry. Gomber et al. (2017) have suggested that the developments in the financial services technology arena – popularly referred to today as *fintech innovations* – are essentially the product of technological innovation-based process and market disruptions, as well as financial services transformation. There has been an exceptionally large amount of capital flowing into this new industry sector around the world. For example, VentureScanner (2017) reported that fintech start-ups, as recently as the fourth quarter in 2017, had cumulatively raised venture capital for a grand total of US$80.4 billion, involving 1537 companies in 64 countries. Opening up the umbrella of fintech coverage to include related sectors, there was an additional US $4.5 billion for cryptocurrency ventures (including Bitcoin and others), related to another 291 start-ups and 74 other countries, and also US $19.5 billion flowing into insurance technology firms, representing 61 countries and 449 new firms. More recently, Consultancy.uk (2017) has reported that the total amount of venture capital in fintech start-ups exceeded the level of US$100 billion worldwide, with US$27.5 billion flowing into the sector in 2017 alone.

Meanwhile China's fintech ventures reaped US$10 billion in 2016, and then a much more modest level of US$2.3 billion in venture capital in 2017. The *Nikkei Asian Review* (Tse and Mellor, 2016) reported that the 2016 venture capital volume was due to multi-billion dollar deals involving Ant Financial and the Shanghai Lujiazui International Financial Asset Exchange (LUFAX). The consultancy firm, Wharton Fintech, has reported that the fundamentals for rapid development of China's fintech sector are already in place (Han, 2018). The list of major indicators include: (1) more than 3.4 billion third-party payment accounts in the country; (2) growth in Internet-based lending balances by 36 times from 2013 to 2016, and total third-party payment value growth even higher at 74 times from 2010 to 2016; (3) presence of five Chinese companies in the Top ten global fintech firms in 2017, according to a KPMG/H2 ranking; (4) high valuations of new fintech firms, such as Ant Financial at between US$100 and US$150 billion, similar to U.S.-based Goldman Sachs; and (5) compound annual fintech venture capital growth for the China fintechs on the order of 300% from 2014 to 2017.

Further predisposing the country for high fintech growth is the high number of Internet users in China at 772 million people with 98% of them (753 million people) who use mobile phones. Also important is the phenomenal growth in mobile payment transaction volumes to an estimate of nearly RMB 200 trillion in 2017, according to the Brookings Institution, based on data from the China Internet Network Information Center (Wang and Dollar, 2018). So it stands to reason that the fintech sector and related financial services activities will be of especially high interest, as researchers in China gear up to understand the emerging innovations in the business and consumer economy in more depth.

3. A methods perspective for causal empirical research with big data

We next discuss choosing appropriate methods for data analytics research for large datasets from China. In our view, an essential property of analysis methods must be their flexibility in several respects. First, the methods need to accommodate different kinds of data. In some cases the data are already available, possibly from public or corporate sources, but need to be pre-processed prior to their analysis. Or they may become available only when the research analyst identifies ways to capture them in their primary form, for example, as with streaming data in a platform market, audio content via radio, TV, or the Internet, and textual data via digital newspapers and publications, or query-capable large-scale databases.

Second, the methods also need to be in tune with the research issues they allow to be studied. For example, will the emphasis be on identifying patterns of online consumer transactions? Or will it be related to extracting causal inferences for what drives the sequences that consumers go through with digital entertainment services, from free sampling, to their initial conversion to purchasing, to engaging in repeated purchases over time? The methods must be matched to the discovery of empirical causality.

Third, the methods also have to be flexible with respect to the type of study. At one end of the spectrum are theory-based studies that aim to explain consumer behavior in e-commerce, m-commerce and s-commerce. Such studies can benefit from the use of data analytics to create the basis for how different processes can be assessed in theoretical terms. At the other end of the spectrum are observational studies that are mostly focused on the discovery of empirical regularities in the data for a specific setting, and the facts as apparent *ground truth.* Such data can be used to characterize: new technology applications and consumer, firm and marketing adoption and usage behavior; and how different marketing, sales and operational processes work at the micro-, meso-, and macro-levels of data, industry and sector performance, among other things.

3.1. Relevant methods from Computer Science

It is important in a Research Commentary on data analytics for empirical research involving causal inference related to China that we should also take some space to mention some of the key areas of contribution from Computer Science (CS). (See Table 1 for an overview.)

Table 1
Articles in Literature That Take Advantage of CS Methods for Social Science Problems.

Authors (Year)	Research Context	CS Methods	Connection to CSS
Aggarwal et al. (2012)	Social media and venture capital financing	Text analytics, natural language processing, binary classifiers, valence analysis	Demonstrates combination of methods with advanced econometrics for causal inference, and enhances pattern recognition
Tosun et al. (2010), Rana et al. (2014)	System bug detection in application software	Machine learning (ML), artificial intelligence (AI), naïve Bayes classifier	Supports causal inference through the identification of richer bug patterns and their association with object classes and bug frequency
Yang et al. (2018)	Misclassification of variables in data mining leading to analyst bias	Multi-stage econometric modeling w/ SIMEX and MC-SIMEX; data mining performance metrics: confusion matrix, error variance	Supports discovery of bias problems from mismeasurement of variables and their correction to improve forecasting with models
Gu et al. (2007)	Virtual investment communities	Signal, neutral and noise messages analytics; network queries, language processing; classifiers for reading, lexicon base, distance, and weights	Supports assessment of investment community's services as a basis for makingparticipation in it worthwhile for financial investors
LeCun et al. (2015)	Review article covering numerous contexts, including speech and text recognition, visual object recognition, autonomous car navigation, selection of relevant search results, etc.	Deep learning with multilayer neural networks and back-propagation; convolutional and deep feed-forward neural networks; supervised learning	Methods match machine learn-ing to human and animal unsupervised learning styles to support simple decisions. Useful for chat bots, vision, etc. but need complex reasoning capabilities to support deep learning for content / structure

Note: Other CS methods have been offered as well. They include: topic modeling and latent Dirichlet analysis; social networks, friends-of-friends, network and graph analytics; and visualization and graph modeling.

This supports our view that it is important to try to integrate new big data analytics techniques in research to transform empirical research.[5] With this in mind, we next offer some comments on some key technical approaches to data analytics.

New methods from CS that have become increasingly useful in combination with explanatory methods in Social Science. They include: text analytics, natural language processing, opinion mining, and sentiment analysis; topic modeling and latent Dirichlet analysis; social network, friends of friends, network and graph analytics; buzz, word-of-mouth and viral marketing analytics; and visualization and graph modeling. These open up new ways to study digital and physical products, relationships between people and companies, as well as mobile phones, bicycle sharing and transport services patterns. In short, whether for business, consumer or social applications, pursuing the related research opportunities that such methods have made available is likely to create the basis for new scientific approaches to assessing causal inference involving the combination of CS and Social Science methods in China and elsewhere.

A major area of research innovation that has been supported by CS methods is *word-of-mouth* (WOM) effects on a variety of phenomena in e-commerce. One such area involves connecting *electronic word-of-mouth* (EWOM) to how business ventures are able to obtain venture capital financing. In this research, Aggarwal et al. (2012) studied the area of social media sentiment EWOM, and applied valence analytics. They drew upon past work by Koppel and Schler (2006), who proposed the use of neutral examples for representing learning sentiment, and Savicky and Fürnkranz (2003), who trained meta-level classifier to resolve disputes that occurred with the conflicting predictions of the binary classifiers they used to classify sentiment valence. The authors, Aggarwal et al. (2012), used these methods with advance regression and statistics methods for robust findings in their research context.

Another important area in IS Management and CS research is software bug detection, and the kinds of advances that have been possible with ML that go beyond traditional knowledge of how bugs can be predicted and discovered. The traditional approach has involved making reasonable assumptions about how bugs are produced in software via the development process. They include the memoryless property of their production, and that fixing specific bugs does not lead to the production of additional correlated bugs (Goel and Okumoto, 1979). More recent work has brought together attempts to explain the production of software bugs with ML and AI methods to explain their occurrence, patterns, and connections. Machine learning methods support the discovery of deeper patterns than an analyst can discover on her own, and make it possible to connect a variety of patterns that a machine can learn and cross-associate with other descriptive characteristics of a software application, leading to more effective bug detection and improved software reliability engineering. Tosun et al. (2010) used an AI technique, the naïve Bayes classifier, to compute the posterior probability of a software module being free of any defects or defective, based on its attributes. It turns out that such an approach is not difficult for practitioners to implement, and others have shown that this approach offers high predictive accuracy.

Moreover, Rana et al. (2014) have shown by conducting multiple case validation with two European firms, Ericsson and the Volvo Car Group, that adopting ML algorithms for software defect prediction enabled the identification of two key benefits and two perceived barriers. The benefis of such data analytics are prediction accuracy and ability to discover new insights, while the barriers include the inability to recognize novel patterns in the bug observationsm and problems with the generalizability of the context-specific ML algorithms. Malhotra et al. (2017) have generalized these kinds of context-specific findings, so the range of ML and AI methods – 14 in all for this application area – have been extended to include object-oriented software, bugs specific to object classes, detection methods for code and object repositories, open source code, and tests of the performance of single-level perceptrons for learning.

Although software development is still subject to errors in bug prediction and software reliability due to mismeasurement, CS methods are now being more broadly applied to a range of issues that can be evaluated using similar blended methods involving machine-based algorithms and statistical analysis. For example, Yang et al. (2018) studied the occurrence of measurement errors and the misclassification of variables in the use of DM, which would have a significant impact on the efficacy of using such methods in big data analytics, especially in second-stage econometric analysis, where the problems with biased estimated can be severe. The authors showed that, as the functional form of the measurement error or the specification of the econometric model grows more complex, anticipating the severity of bias becomes more challenging. They further offered some methodological insights and specific solutions, including error correction methods that

[5] We benefited from the suggestions of the anonymous reviewers on incorporating this content in this essay, and what the general coverage and main messages should be.

implement performance metrics from DM models involving the error variance and confusion matrix, that help to improve performance. The techniques were demonstrated based on the authors' applications to travel, social networks, and crowdfunding campaign websites. The authors' methods will be of interest to digital commerce researchers in China.

The final area of study that we point out which uses methods from CS, is for the study of competition in investing-related virtual communities and how participants' asset valuations are influenced by other participants. Gu et al. (2007) used 500,000 textual postings from three large *virtual investment communities* (VICs) for 14 different stocks over four years. The authors were interested in assessing the extent to which virtual communities are a source of information for investors. They assessed the trade-offs between information quantity and quality, and the sources of positive and negative externalities in virtual communities.

To assess the posting quality of investment discussion boards, they applied a *network query language* (NQL) to identify posts for downloading, and multiple classification algorithms to identifdy the posts as providing *signal, neutral* or *noise* messages. The algorithms included: a *lexicon-based classifier* (LBC); a *readability-based classifier* (RBC); a *weighted lexicon classifier* (WLC), a *vector distance classifier* (VDC), and a *differential weights lexicon classifier* (DWLC). They used three stages, with the first stage involving the application of the algorithms we just noted. The second stage involves application of the algorithms to a holdout sample of 800 posts, and the creation of a statistical accuracy measure to compute *corrected classification rates* for the three kinds of messages. The last step was to apply the classification approach to the entire dataset, and then to generate posting quality as a percent of relevant postings. The authors completed their data analytics by assessing multiple hypotheses for their context in terms of the patterns that their methods produce from the pre-processed textual data.

This research is interesting, not because it is specificially focused on causal inference, but because of the novel manner that it used to work with text data, as a basis for testing three sets of hypotheses. They relate to the marginal value and impact of a virtual investment community's: the number of useful postings, the size of the useful postings, and the investors' cost of using the community; the effect of community size on service value and posting quality; and finally, higher-valuation investors' preference for higher quality communities, and lower information-processing cost investors' preference for higher quality communities.

Finally, a review article on deep learning from *Nature* (Lecun et al., 2015) shown in Table 1 is especially interesting in this regard, it offers useful insights about current methods that will expand the power of AI and ML solutions in ways that are likely to improve the technologies for much strong business, consumer and social insights solutions in the future. Overall, the research we have discussed offers many useful interdisciplinary methods ideas, but its special focus on CS, ML and AI algorithms is suggestive of the technical innovations that can be applied by researchers to problems in digital commerce in China.

3.2. Computational Social Science thinking

The approaches to empirical research that we discuss involving computational social science thinking are able to support a variety of new research directions for data analytics related to large datasets in China. They include: assessing contextual awareness related to e-commerce transaction-making and evaluating the business value of personalization; using patterns and explanations to create the basis for predictions; and extending our current reach in e-commerce research to sensing analytics.

Quasi-experimental research designs and causal inference approaches associated with CSS research make it possible to conduct studies that seek to understand the impact of location, channel, and time on sales. This is useful for the study of e-commerce settings in which sellers and service providers invest in a single channel (among several alternatives) or multiple channels to connect with their consumers. In the latter case, a relevant issue is channel attribution for the returns on ad campaigns. It also creates a basis for comparative studies of firm and process outcomes involving m-commerce and s-commerce, in contrast to traditional e-commerce and Internet-based selling. Another area of benefit is research on social media and social networks, and our capability to explain the power of consumer opinions and shared sentiments, in the presence of friends-of-friends' influences (Natali et al., 2017). Contextual information is crucial for clues as to why different outcomes materialize. This is useful for developing new traditions involving causal empiricism in China's big data research.

Such methods can also be applied when individual-level data and information are available, even in an anonymized form, to explore what drives consumer decision-making—and ultimately, through technologies like deep neural networks, will support the development of sufficient machine intelligence to make it possible to produce and effectively operate autonomous vehicles, such as for supply chain delivery in repeated routes. Also, though firms and government agencies have different access to information about consumers and citizens, more revealing empirical research can be conducted to identify consumers' product and service choice behaviors, their willingness-to-pay, and propensity to participate in different kinds of activities. These things can be observed rather than predicted based on the capture of their intentions in survey questionnaires. In different kinds of marketing activities, this capability has the potential to support sales of hyperdifferentiated "segment of one" pricing and services (Clemons et al., 2006, 2017).[6] We expect that deep learning-based prediction will also enhance these kinds of activities.

3.3. Causal inference process

The conduct of empirical research that is most likely to produce causal inferences is based on fully-experimental and natural or quasi-experimental methods. Experiments have been designed over many years in the Natural Sciences, Psychology, Sociology and Economics. And more recently, we have evaluated the extent to which real-world natural experiments make it possible to assess causality in meaningful ways (e.g., Banker et al., 1990; Kauffman et al., 2009; Li et al., 2014). Three key principles have been recognized that are associated with assurance of the production of causal explanations (Pearl, 2009). They include the following principles:

- *Correlation-based association.* The variable that captures information about the cause must be statistically correlated with the variable that represents information about the outcome.
- *Temporal precedence.* The cause must be present in advance of the observation of the effect.
- *No non-spurious associations.* The research analyst also must be able to rule out other alternative explanations of the relationship between the cause and the effect. Often, an outcome appears to be caused by another spurious effect due to a variable that is correlated with both the cause and the effect in a setting.

An interesting and informative source of fundamental information

[6] A further example concerning the analysis of factors in decision-making is from recent research that we conducted at the Living Analytics Research Centre (LARC) at Singapore Management University. It suggests that new levels of consumer informedness affect the sequences of actions that consumers make that result in purchase conversion and higher average revenue per user for digital entertainment (Hoang and Kauffman, 2018a,b). The insights obtained for household digital entertainment can be used for both retrospective explanation and theory-based prediction, since they are a product of causal inference. The research uncovered empirical regularities for consumer preferences that made it possible to understand the firm's customers more deeply.

on the development of causal research designs, and how to develop strategies for statistical analysis to make it possible to capture the relationship between the causes and the effects of motor vehicle crashes from driver fatigue was published by the National Academies of Sciences, Engineering, and Medicine (NASEM) (2016). The authors (p. 87) noted "*the relatively new subdiscipline* of *causal inference*," which they indicated includes design and analysis techniques useful in "*separating out the impact of fatigue and other causal factors on crash risk and … determin[e] the extent to which fatigue is causal.*"

They also discussed the challenges associated with data collection, including situations that arise where the data are of low quality or incomplete, where informants may not be truthful, and where true values of key variables may be systematically under-reported, leading to biases and inaccuracies. The NASEM (2016, p. 89) authors emphasized that developing appropriate inferences about causality requires understanding and controlling for causal factors. They pointed to confounding factors outside the interest of the analyst as well, such as are often present in e-commerce research.

Some other useful recommendations were offered that similarly apply to research in e-commerce. For example, there is the issue of *standardizing the dependent variable*. In many contexts, it is very hard to identify the impact on a dependent variable that is changing over time, such as conversions from consumer searches to purchase. Differential exposure to advertising, multi-channel marketing, and social media participation influences typically make it difficult to assess the number of purchases that occur at different points in time. The key question that arises is: How should *exposure variables* that lead to purchase conversions (e.g., with the number of search engine searches, consumer average exposures by channel, or ads in a period of time as possible denominators) be normalized? This is not an easy problem to circumvent, since the intensity of Internet sales changes over time due to sector growth in demand and a variety of competitive intensity-related factors.

Some of the methods that can enable a data scientist to achieve more effective causal inference include *randomized experiments*. The study groups (popularly referred to as "Facebook A/B experiments") (Cohen, 2014; Goel, 2014) only differ based on the treatment (e.g., due to the receipt of different news stories by email, or different ad exposures). So the possible confounding variables are not correlated with the behavior exhibited by the treatment and control group subjects (Kohavi et al., 2014). But a problem is that in many business and social situations the "perfect randomized treatment/control experiment" cannot be constructed because the researcher has no control over real-world industry and social activities, and the world does not have a controllable lab environment.

Other observational study approaches are possible though. They include the opportunity to take advantage of industry and social settings that have the natural structure of a quasi-experiment, which do not offer the researcher any opportunities to affect the treatment and control conditions before the activities are observed in public. We did this in Singapore, with a study of the entry of smart phones and their impacts on the triple-play telecom services (broadband, digital entertainment on TV, and phone) by leveraging customer account information and broadband service choices before and after smartphone entry occurred (Kim et al., 2018). For this research, we used the *case-control study* design approach for observational studies, in which the primary concern was to recover the *causal factors of interest*.

Other methods are possible as well. For example, one is a *longitudinal cohort study*, which tracks subjects in an historical manner to follow their outcomes relative to the treatments they were observed to have received (e.g., as with financial market investments by investors who have had more involvement with financial advisors yet are similar in other ways). A second is a *case-crossover study*, in which subjects are assigned to different longitudinal treatment groups to receive a sequence of different treatments (Stufken, 1996). Such research designs can be fully experimental or observational and may be strongly subject to the loss of participants over time – a problem that is referred to as *data censoring.* This has been addressed in recent IS research that focuses on strategies for data recovery to perfect temporal sequences to support causal inference (Hoang and Kauffman, 2018a,b; Hoang et al., 2018). Thus, we encourage authors in China and around the world to be cautious with respect to the issue of *correlation* versus *causality*, and to design their empirical research inquiries so as to distinguish between them.

3.4. Digital traces of human behavior

Chang et al. (2013, p. 6) have noted that "*the emergence of methods for the collection of historically large data sets in Social Science, and changes in the relative costs of conducting research that has rigorous scientific controls, while maintaining realism relative to real-world business activities and generality across research contexts, reflects a paradigm shift toward [CSS]*." This supports data and policy analytics researchers in doing more innovative and insightful empirical research. This is true for those whose home base for study is China, with its leading tech firms, social and business innovations, continuing economic growth, and progress in increasing social welfare. Such contexts are appropriate for the study of business, consumer and social insights based on the use of research designs that support causal inference and policy-related actions.

Big data now can speak for themselves without our imposing unnatural assumptions on the study context, which is reminiscent of the views of Sims (1980). In addition, very large data sets support support iterative resampling, stratification, simulation of statistical distributions, and "needle-in-a-haystack search" for deep behavioral patterns (Kauffman and Lee, 2010). They also support knowledge discovery through appropriate natural experiments, fully-experimental methods, and closed-loop experiments to assess consumer incentives and the nature of (Chang et al., 2014; Edin et al., 2004). Indeed, it is likely that the application of innovative methods which go beyond traditional surveys of behavioral intentions and consumer response will yield new knowledge at a rate that exceeds what the traditional methods have been able to do.

The basis of such new approaches is founded on researchers' interests and capabilities to acquire large-scale digital trace data from e-commerce platforms and sharing economy businesses. The data can support tests of new theories to explain what can be observed to understand the new nature of the changes in collaborative commerce. They also are relevant because the scale of the data makes it possible to do extraordinary natural experiments, since extensive data create many opportunities to develop event-based treatments and controls. We also advocate the use of data analytics methods to select subjects for analysis or to weight observations for balance between a treatment and a control group (NASEM, 2016, p. 96). An example of such a method for data analytics that matches this description is *propensity score matching* (Rosenbaum and Rubin, 1983; Shadish et al., 2002; Pearl, 2009).

A valuable consumer insight must be new, enduring in its potential impact, and suggestive of managerial actions that can be taken to improve organizational performance (Kauffman et al., 2017). These authors (Kauffman et al., 2016) note the role of digital traces of consumer behavior. They are among the most important sources of data for insight creation, and much more powerful in causal inference terms than the use of behavioral surveys of consumer attitudes and latent variables. Different research approaches have been suggeste, including information visualization and network analysis (Aigner et al., 2008; Lazer et al., 2009). Their efficacy has improved with the emergence of big data.

4. Can data science be relevant for practice in China?

An important way to describe the quality of research is to assess its relevance for practice. We provide suggestions as to the research approach that will lead to relevant results. We suggest to follow a number of principles: (1) look for big data applications in China; (2) pay attention to context-specific elements; (3) search for specific practical problems in China; (4) exploit the richness of data from China; and (5) investigate the practical consequences.

First, it's important to focus on an area where data applications have been deployed in China. Wide choices are available, since data analytics has been applied in various sectors there, including healthcare, social media, finance, e-commerce and e-retailing, transportation, and so on (Xu et al., 2014). Attention also should be paid to contextual elements that are specific to the Chinese nations. Examples are the huge market size, the generation of data in Chinese character text, the long-standing cultural traditions (Hofstede, 1984), and the centralized system (Biuk-Aghai et al., 2016; Zhang et al., 2018).

In addition to the industry and the national context, researchers should be able to draw inspiration from the problems emerging in Chinese data analytics practices. They include the recent event of the "runaway boss" from an online P2P lending platform (Liu, 2018), threats to passenger safety in vehicle-sharing, higher prices charged to repeat customers by taking advantage of their consumption information, and favorable comments for home-sharing vendors via payments to fake consumers, who act as Internet mercenaries. Many of the related news items have appeared in the general press, but they have posed challenges to the growth of big data practices in the relevant industries in China. Research that can address these problems has significant implications for the development of data applications there, according to Feng et al. (2013).

Since the market is very large and generates a very high volume of data, researchers should make full use of it. As mentioned above, the volume and the velocity of data in various forms generated in China have been increasing across many platforms and industries, including the e-commerce, m-commerce, s-commerce, and fintech sectors. Research grounded on these data has potential in revealing new patterns of consumer behavior and evaluating the effectiveness of firms (Müller et al., 2016).

Finally, it's important to make an effort to foresee the implications for practice rather than highlight the theoretical contributions only. Researchers ought to think more about the practical implications of their findings, such as the potential impact on business practices, the related product and service cost structures, demand and market structure, and how business may react (Grover et al., 2018).

4.1. Usefulness of data analytics research in China

The usefulness of research acivity and results related to data analytics in China can be demonstrated in several contexts, considering four main categories of stakeholders: companies, consumers, government, and society. Companies may take advantage of the results of big data research to make more informed decisions. Firms may gain insights about the targeted market segments and consumers to support more informed business decision-making (Chen et al., 2012). For example, Lien and Cao (2014) found that Chinese consumers were influenced by social media in making purchase decisions much more than consumers in other countries, and they were more inclined to purchase products and services exposed on social media. They were more motivated to post feedback for products and services, and to make recommendations to their relatives, friends and colleagues for favored products and services. This suggests it will be beneficial for companies to increase their presence on social media, and pay more attention to the feedback of customers when they are exposed to the firms' platforms.

Another example of the usefulness of data analytics research is represented by the study of online P2P lending published by Wang et al. (2018b), who built a novel user a behavioral and a credit risk model based on the analysis of historical data with fine granularity in the financial sector. The results showed that more precise user profiles can lower business decision risks. With regard to the value of the research, the authors noted that timely intervention for potential defaulters diminishes loan defaults and bad debt accounts, which reduces the losses banks have to absorb.

Consumers and users may also benefit from the research and findings. When firms draw better user profiles that comprehensively and precisely reflect users' consumption habits and preferences, they are able to provide users with more resonant and individualized services (Schlager et al., 2018). There has been plenty of research about the social media behavior of Chinese users in recent years. The results have shown that the intentions of Chinese users when using social media usage are manifold, such as receiving and sharing information (Lien et al., 2017), making new friends, maintaining relationships with old friends (Chang and Zhu, 2011), and consuming entertainment services (Gan and Wang, 2015).

Moreover, users may perceive value in dimensions such as social networking, and its informational, emotional, and hedonic effects in using social media, while emotional and hedonic value may result in consumers' continuous usage (Zhang et al., 2016). The feature of a seasonal "Red Envelope" initiated by WeChat gained extensive attention from Chinese users, mostly due to the elegant integration of social networking and entertainment. Extended functionality for transportation, healthcare, and a variety of public services were gradually integrated in the platform of instant messaging, providing convenience for users in life, work, and entertainment (Lien and Cao, 2014).

Finally, from the perspectives of government and society, the research and findings related to data analytics may enhance public services and regulatory capability for government agencies, and improve social governance effectiveness. In 2015, the State Council of China issued a document, on "Guidelines for Enhancing the Services and Regulatory for Market Entities by Deploying Big Data Analytics." It declared that: *"… making full use of the advanced ideas, technologies, and resources of big data analytics, …* [*it*] *is a strategic choice for enhancing national competitiveness and an inevitable requirement for enhancing the services and regulatory capability of government. It helps government to capture and utilize extensive information, to gain precisely understanding of the needs of market entities, to improve the pertinence and the effectiveness of services and regulation."* [7,8]

4.2. Value assessment for e-commerce research with big data

The value of scientific research manifests itself in both academic and practical usage, especially with social science and technology research (Fiorini et al., 2018). In research on big data practices in healthcare in China, for example, Zhang et al. (2018) have argued that for value from big data, new approaches and tools like ML ought be explored, and effort should be made to ensure that the information created is assessed for clinical effectiveness and use in practice.

In fact, the practical valuation of research is not only important in the field of healthcare, but elsewhere as well. Valuation should concern all the stakeholders categories, including business, consumers, and society (Wu et al., 2017). But how can we ensure an objective and valid valuation of a research appication, especially for Social Science research with big data analytics? Researchers should approach practical valuation by considering several aspects: the method, data sources, indicators, frequency of valuation, and scale at which the

[7] See www.gov.cn/zhengce/content/2015–07/01/content_9994.htm.

[8] In the same year, the State Council of China issued another document, "The Initiatives for Advancing Big Data Development," in which it is emphasized that *"… big data analytics can uncover associations which it is difficult for traditional technologies to reveal, promote government to open and share public data, promote data fusion of government and society and resource integration, and enhance overall capability data analysis of government to a great extent."* See www.gov.cn/zhengce/content/2015-07/01/content_9994.htm.

valuation is conducted.[9] Qualitative research and multiple case study methods can be used to assess the business, consumer, and social value of research findings qualitatively, while clustering, time-series analysis, regression models, decision trees, ML and other emerging techniques can be used to assess the value of research findings quantitatively with more detailed information (Wang et al., 2018a). Practical valuation also can be conducted using online or offline data, and the application context may be an online or real-world activity.[10]

Also standard cost-benefit analysis can be used to assess the value of a research project for business, consumers, and society (Quah and Haldane, 2007; Van Den Broeke et al., 2018). Generally, the value of adopting data analytics in business comes from cost reductions in business finance, human resources, and logistics, and the benefits increase in manufacturing and sales processes (Ghasemaghaei et al., 2018). Zhong et al. (2016) demonstrated that such adoption can promote production efficiency and product innovation by mining manufacturing data using data technologies.

From the perspective of consumers or users, valuation should cover aspects like time and effort spent, the quality of products or services, the consumption experience, and the level of satisfaction. For instance, by extracting and analyzing events from Chinese online news, Chen and Yin (2017) demonstrated that they could be used to predict stock returns and industry trends. Chen et al. (2018a,b) found that the fluctuation of prices of two stocks shared a similar pattern if the firms were frequently mentioned in online media. Such research findings can be valuable for investors' portfolios (Han et al., 2018a).

The social value of a research project can be shown by its implications for policy-making and macro-governance mechanisms. For example, by collecting and analyzing data for a variety of diseases, Zhou et al. (2016) depicted the trends in epidemic diseases across regions in China and assessed various healthcare problems. The results helped government agencies to issue disease warnings.[11,12]

[9] An example of usefulness evaluation for companies concerns the analysis of data from the devices and sensors provided by the Topaxi Energy Efficiency Management System of BONC at UniLever. The system allows the quantification of facility operations, including energy consumption, early warning, failure prediction, and overall efficiency of a variety of equipment, and measures improvement with respect to a baseline case. Deploying such a system is reported to have resulted in energy savings of 3–10%, machine downtime reduction by 5–7%, and maintenance cost reduction of 20% (AII, 2018).

[10] Two examples of research in an offline context are illustrative. They respectively aimed at improving the efficiency of industrial equipment by adopting big data technologies (AII, 2018), and enhancing the management of traffic congestion in metropolitan areas (Bauza and Gozálvez, 2013). Also, the valuation of a research effort based on online data fits more with research in online contexts such as e-commerce and m-commerce. For instance, by using online review data from streaming media in China, including Youku, Ku6, and Tudou, Zhang et al. (2013) were able to design a novel recommendation algorithm based on extracting emotions from comments posted by Chinese users, which improved the effectiveness of recommendation.

[11] Through the acceleration of interactions between companies and consumers over the Internet, before large-scale launches of new products or services, Chinese online companies such as JD.com and YHD.com have been able to make small-scale pilot tests. The outcome of research can be assessed more frequently online at a small scale, which can improve the application of research methods. For example, Chan and Chong (2017) demonstrated that by mining financial text data, market emotions and the price index of the stock market can be predicted better in real time.

[12] Two aspects that must also be considered when defining a valuation approach are the frequency and the scale of assessment. Traditionally, the value of research studies has been assessed infrequently. This has been due to technological restrictions and cost considerations. The presence of uncertainty has led to requests for increasing flexibility too (Van Den Broeke et al., 2018). The incorporation of the value of assessment flexibility, especially the frequent re-evaluation of the research results, into the value of a technology project has been made possible thanks to the theory of real options (Benaroch, 2001).

4.3. Evidence-based policy-making

The roots of evidence-based policy-making can be found in the U.K.'s 1997 Labor Party Manifesto (Wells, 2014, p. 1), which called for it. The goal was to move away from opinions that were unsupported or could be refuted. This reflected a pushback on the more traditional view of policy-making. It involves the forces of power, politics and people that shape government and agency policy outcomes. This view is a government-focused perspective first, and a business and enterprise perspective second.

According to the Evidence-Based Policy-Making Collaborative (2016), a joint effort involving the U.S.-based Pew Charitable Trusts-MacArthur Foundation Results First Initiative, the Urban Institute, the American Enterprise Institute, and the Brookings Institution, *evidence-based policy-making* has two goals: to apply what is known from program evaluation to make policy decisions and to build new knowledge for future decisions. It uses research findings, data, analytics, and new innovations in surfacing useful information.[13]

Various authors have discussed the role of *policy analytics* – for government and for business. Additional consideration has been given to the role that data-at-scale can play in the process of crafting the most effective policies for the organizations that seek them. For example, Tsoukiàs et al. (2013, p. 124) have suggested that *policy analytics* involves development and application of skills, methodologies, methods and technologies to support stakeholders by providing hindsight, insight and foresight.[14,15]

The huge amount of data generated in China is a vital resource in supporting smart decision-making in the context of business, consumer, and social management (Goes, 2014). We highlight several guidelines for researchers when the principles of evidence-based policy-making may be applied in business contexts. First, data generated inside and outside of business should be integrated together so an organization is able to gain insights in making business decisions, such as business operations (including data from R&D, production, operations, customers, and supply chain partners), environment monitoring and equipment operations (AII, 2018; China Electronics Standardization Institute (CESI), 2017). In addition, relevant data on law and regulations, the macroeconomy, industrial events and trends, user interactions with websites, and social media can be integrated (Zhang et al., 2018).

Second, effective data technologies should be deployed in data processing and analysis. Distributed file systems such as Hadoop, MapReduce can be used to store and process big data (Biuk-Aghai et al., 2016). Technologies for text analysis, natural language processing, networking and graphic analysis, and visualization can be used in data modeling and analysis (Lu et al., 2015), in order to discover knowledge related to business and consumers. Chen et al. (2018a,b) discussed several emerging data transformation technologies based on various kinds of neural networks.

Third, the insights gained from data analytics should be used in decision-making, such as streamlining production processes, reducing

[13] This approach has four notable principles (The Collaborative, 2016). The principles are to: (1) build and compile rigorous evidence about what works, including costs and benefits, to measure program impact; (2) monitor program delivery and use impact evaluation to measure program effectiveness; (3) use rigorous evidence to improve programs, scale what works, and redirect funds away from consistently ineffective programs; and (4) encourage innovation and test new approaches.

[14] For additional details on the authors' formal perspective on policy analytics, and its connections to Operations Research (OR), the interested reader should see De Marchi et al. (2016).

[15] Such information sources as facts, scientific knowledge, and expert knowledge are all appropriate to help the views of policy stakeholders to be understood. Although the authors primarily characterize policy analytics in public policy terms, the reader should recognize, as we suggest, that similar concepts are applicable for business, consumer and social policy analytics.

failure rates of equipment, and controlling resource consumption (Jeske et al., 2013; Zhong et al., 2016), improving human resource allocation and financial performance (Chen et al., 2012; Gunasekaran et al., 2017), improving product design and reducing time to market of new prodcts (Yu and Zhu, 2016; Wen and Zhou, 2016), reducing supply chain costs and alleviating related risks, increasing operational agility, and improving service quality (Ittmann, 2015).

4.4. Problems, contexts and outcomes

Several interesting studies in the context of e-commerce platforms that are practically relevant to China illustrate the above principles. For example, many online retailing platforms operating in China provide real-time chat functionality (Ali Wangwang, for instance) to facilitate the interaction between vendors and consumers. This is not available in the western platforms such as Amazon and eBay. By analyzing data from vendors on Taobao, Lv et al. (2018) found that the usage, responsiveness and frequency of real-time chats have significant effects on the purchase decision of Chinese consumers. They suggested that vendors on these platforms should adopt real-time chat functionality to maintain constant communications with consumers, provide rich and timely information, increase the credibility of vendors, and stimulate purchases from consumers.

E-commerce markets that operate in China, especially customer-to-customer (C2C) online markets, share several features with other digital marketplaces around the world: they are rapidly changing, support many new entrants and are highly competitive, have low entry barriers, and are physically dispersed. Gao et al. (2016) investigated the relationships between the age, scale and growth of online stores in such platforms. By analyzing panel data from Taobao, they discovered that small-scale, newer entrants suffered from higher market pressures and gained rapid growth, while established, large-scale online stores had early-mover advantage, but tended to grow more slowly. This provides evidence for C2C platforms to redesign their business policies, and leverage reputation systems to enhance the growth of established online stores.[16]

Data collected by e-commerce platforms also can be used to provide data analytics services to vendors. In an empirical study of Tmall, the biggest B2C online platform in China, Song et al. (2018) found that data analytics services based on data collected on a platform have significant effects on the sales of platform-based vendors. They further recommended that, when there is a wide variety of products and highly competitive market segments, vendors should use more data services on the demand-side and less on the supply-side. In addition, the provision of effective data analytics services to vendors can help improve the competitiveness of the platform.

5. Research directions

We next discuss some critical issues that are associated with the development of data analytics in Chinese digital commerce research. They suggest significant opportunities for research as well, and particularly research that is founded on the CSS paradigm, by combining empirical research methods for causal inference, with new machine and algorithm-based techniques. With the rapid advances of data analytics technologies and their widespread applications, new problems have emerged. They include open data campaigns, data security and privacy protection, higher prices charged to repeat customers by taking advantage of their purchase information, forged and fake commodities, and so on. These things pose great challenges to practitioners and policy-makers alike. We next focus on topics that call for investigation.

5.1. Open data and sharing

In 2009, the U.S. government launched an open data website, called data.gov. After that, more and more countries created websites to release public data regularly. In China also, the central government launched its own public data service platform. The acceleration of the provision of open public data provides academic institutions with a huge opportunity to carry out extensive research by exploring these data. They exhibit interesting properties, such as covering a wide area (typically the nation), and being certified by an institutional source.

5.2. Data security and privacy protection

The rapid development of big data applications in China has also brought challenges to data security and user privacy. Users are increasingly releasing their data on the Internet, and more and more companies are collecting an ever large volume of user information for business purposes. While the awareness of privacy protection is being promoted for both users and business in China, the ability to safeguard all of the stakeholders may not measure up to the challenge of the task (China Academy of Information and Communications Technology (CAICT), 2018).

Most Chinese are becoming increasingly concerned about the protection of their privacy. They have few alternatives in the national information environment that shapes their lifestyles. Data leakages have occurred one after another, bringing panic to consumers and losses to companies. Those with data concerning their customers are crying out for technologies to ensure data security and privacy protection. In addition, ethical protection is also a serious problem that such companies are required to deal with. On the other hand, digital commerce consumers and users of social networks should seek greater awareness and take proactive steps for protecting their privacy.

This will, no doubt, represent a source of concern for faculty members and doctoral students who wish to acquire a lot of potentially sensitive data for their research. In fact, more and more organizations have been trying to lock up their data due to considerations of protecting consumers' privacy and the concerns of potential leakage of these data. As a result, academic institutions face increasing difficulties in carrying out relevant research, leaving a lot of valuable data on the shelf. Researchers and academic institutions have been signing confidentiality agreements with organizations to secure anonymized data, in order to carry out their research in an environment with greater mutual trust. And yet we all know that this is not easy: too many instances have occurred in which such data sharing agreements have led to significant confidentiality and reputation problems, and profit impacts.

5.3. Using big data to exploit consumers

With the accumulation of data about the interactions between consumers and business in today's world, companies know more about the habits and preferences of consumers than ever before. They now routinely use data analytics to build individual profiles for each consumer. While companies are also able to provide customized service experiences for their customers as a result, the information access asymmetries between consumers and companies will increase, leading to a raft of new undesirable impacts for consumers that are still unknown. In China, a special kind of price discrimination has occurred in

[16] The majority of revenue of China-based e-commerce platforms has been coming from fees for search ads, listings, and subscriptions. Lee et al. (2018) found that consumer perceptions of the quality of e-commerce platforms have significant impacts on the effectiveness of marketing strategies of platform operators like Taobao. By providing transparent information to consumers, an online C2C platform is unlikely to turn into a "market for lemons." Its customers will be well-informed about the items they wish to buy, and sellers can simultaneously avoid the "death spiral" associated with proct information asymmetries. E-commerce platforms should formulate a strategy that fits with a clear-cut revenue model in order to increase their revenues and improve their competitiveness.

e-commerce, for example. Repeat customers of travel, commuting, and hotel websites have been asked to pay higher prices than newly-registered customers for the same products, as the menus are customized for each individual customer. This has increased revenues in the short term, but at the cost of concern and discontent of consumers, which may result in revenue losses for companies in the long run. The characteristics of such customer behavior, its consequences, and the related coping strategies are potentially relevant research topics as a result.

5.4. *Fake products for sale in the online market*

In the huge markets of China, forged and fake products have existed for a long time, due to many reasons. They include imperfect market institutions, unfulfilled supervision, and the presence of many illegal and fraudulent companies cheating on labor and materials to pursue more profits. Three online market and six offline marketplaces in China were listed in a recent list of "notorious markets" issued by the Office of the United States Trade Representative (2017), including Taobao, the biggest player in e-commerce in China. Advances in data technologies and the continuous improvement of online markets provide opportunities for government agencies and companies to fight against forged and fake products both online and offline. Research on applications of data analytics are called for to combat fraudulent digital commerce activities in the Chinese context.

In this article, we have argued that the time is appropriate to shift toward CSS-based combinations of big data analytics that effectively tie in CSS approaches with others, such as CS and advanced econometrics methods, to achieve important business, consumer and social insights based on causal inference. We also have argued that this represents a set of "next steps" for digital commerce researchers in China, who must begin to implement research designs that include innovative real-world experiments, large-scale data collection, and the study of problems and research questions that were mostly inaccessible heretofore in research.

With these general goals in mind, we investigated a set of related issues that enabled us to draw conclusions about the research process for establishing causal inference. We did this based on the relevant methods from CS and econometrics, CSS thinking, the key techniques that researchers must use to construct research designs aimed at supporting causal inference, and their potential use of "digital trace" data for human behavior and other micro-level data phenomena. We encourage digital commerce researchers in China – and elsewhere in Asia – to implement the philosophy of science for big data-enabled empirical research that emphasizes causal inference. And to especially consider the variety of Computer Science methodologies that are discussed in Section 3.1 in their research designs. They offer possible bases for creating interdisciplinary technical and Social Science research approaches that can leverage the strengths of current machine learning and artificial intelligence methods to reveal new and interesting insights and findings.

Finally, based on our consideration of China as a "next major target" for the application of CSS methods and big data analytics, we laid out some of the reasons why data analytics research involving causal inference is a logical next step in a newly-evolving chain of research activities that will further enhance the power of the insights and research findings that can be obtained in Chinese digital commerce. We commented on pathways to discovering the business value of implementing big data analytics research designs in terms of the business and consumer social outcomes that they support, and the evidence-based policy-making approach they encourage.

Acknowledgments

This article benefited from the careful reviews of Dongsong Zhang and Alan Wang, as well as the Co-Editor of ECRA, Chris Yang. The authors also wish to acknowledge a number of people who encouraged us to undertake this effort several years ago: Han Zhang, Youwei Wang, Jing Zhao, and Zhen Zhu. Rob Kauffman thanks Alexis Tsoukiàs for his input on evidence-based policy analytics, during his visit to Singapore Management University's School of Information Systems to speak at the 2015 Workshop on Business, Consumer and Social Insights. We also acknowledge ECRA's Editorial Assistant for Special Issues, Ai-Phuong Hoang, and the peer reviewers who helped to develop all of the manuscripts for publication in the China Big Data and Causal Inference Special Issue. We had additional support from the organizers of the 2017 Wuhan International Conference on E-Business, for hosting a research workshop on this theme. All errors and omissions are due to the authors.

Appendix A. Empirical research on digital commerce published in ECRA

The reader should note that the appendix articles are not all big data articles with big data or CSS methods used. They simply represent an inventory of articles that use interesting theory, and mostly traditional methods to study the research questions on the issues they examine. They illustrate the opportunities to apply new methods. Not all of the research was conducted in China, though ECRA nevertheless published some of these works.

Table A1
Representative Electronic Commerce Empirical Research Published in ECRA (2012–2018).

Authors (Year)	Research Questions	Context, Data, Data Collection Level	Research Approach
Akter and Wamba (2016)	What can a systematic review of the big data analytics literature reveal? What kind of interpretive framework is useful to provide basic definitions, distinctive characteristics, impact types, business value and challenges for e-commerce landscape?	Context: 121 articles in e-commerce literature Data: Small data set from targeted literature	Large-scale lit survey yielding detailed analysis and classification 33 search strings for content, resulting in 5 categories: needs identification; mkt segments; decision-making; new product, mkt business innovations; infrastructure and transparency creation
Chang et al. (2012)	Are channel bundle preferences of TV viewers related to their viewing habits?	Context: Cable TV in Singapore Data: over 100,000 TV set-top boxes Meso-level data analysis	TV programming show genre classification through clustering with big data analytics data pre-processing
Liang and Yang (2018)	How do hedonic and utilitarian motivations influence participation in online travel communities? How is trust related to participation behavior in online travel communities?	Context: Tourism e-commerce Data: 285 questionnaires Meso-level data analysis	Structural equation model

(*continued on next page*)

Table A1 (*continued*)

Authors (Year)	Research Questions	Context, Data, Data Collection Level	Research Approach
Lv et al. (2018)	What is the effect of live chats between consumers and sellers on consumer purchase decisions?	Context: Online skin care product store Data: Surfing behavior, 29,801 consumers, 6 months, plus data on 6,517 live chats with 96,422 postings	Logit model Statistical hypothesis testing
Qin et al. (2017)	What is the optimal granularity for market segmentation?	Context: Online advertising Data: Theoretical model with simulated data	Integer programming optimization
Tian et al. (2016)	Can we train ML tools for sentiment analysis in online reviews when class distribution is imbalanced?	Context: E-commerce portals Data: Two meso-level datasets, Jingdong, Dangdang	Sentiment analysis Machine learning
Wang et al. (2018b)	Is it possible to do dynamic prediction of the probability of P2P loan default?	Context: P2P lending Data: Meso-level dataset from a P2P platform in China; 6079 defaults, 46,494 non-default loans	Random forest

Table A2
Representative Mobile Commerce Empirical Research Published in ECRA (2006–2017).

Authors (Year)	Research Questions	Context and Data	Research Approach
Chang et al. (2017)	Can push-pull mooring (PPM) theory be used to model consumer switching in multichannel shopping? Are factors for PPM theory good for m-commerce? Which affect consumer switching intention in multichannel shopping?	Context: PTT E-Shopping website in Taiwan Data: 403 m-shopping experienced subjects	10 theory-driven hypotheses tested using SEM methods to assess PPM applicability for context
Guo et al. (2016)	What is the privacy–personalization paradox in m-health service acceptance? Can privacy and personalization be balanced with consumer trust? Can privacy–personalization paradox factors provide different explanation of intention to adopt among young and elderly people?	Context: Acceptance of mobile health services among consumers in Harbin, China Data: 650 m-health services user-respondents participating in a survey questionnaire study	Measurement and structural models tested with SPSS and SMART PLS to assess the explanatory capacity of privacy concerns and perceived personalization on trust before adoption intention is affected
Hsu and Lin (2015)	Can mobile app users' purchase intentions be explained by expectation confirmation model? Does m-commerce app setting mimic try-first, purchase-later behavior captured by the model?	Context: Survey posted on Taiwan m-commerce sites and social communities for 45 days Data: 307 m-commerce app users participated	Application of SEM methods for assessing the structural model representing the authors' theory for m-commerce app try-outs and adoption for shopping
Hsiao and Chen (2016)	What motivates players' in-app purchase behavior? How do factors that influence intention differ for m-games and other settings?	Context: Tower of Saviors Internet-based game access via mobile phones in Taiwan Data: 3309 mobile game players, with 813 non-paying and 2469 paying participants in 2 Q 2014	SEM and AMOS software used to assess the structural model, and determine the appropriateness of selected study variables
Lin et al. (2011)	What is impact of inter-channel trust transfer on building initial trust in m-brokerage services? Direct/indirect impacts of trust in e-brokerage services on initial trust in m-brokerage?	Context: Branch customers of 4 leading brokerage firms in China with m-brokerage services Data: 357 sample respondents from Hubei, Beijing, Guangdong and Guangxi	Use of LISREL to assess a structural model involving traditional variables from TAM, trust and information quality theory for the m-brokerage context, with 8 of 9 hypotheses supported
Liu and Liou (2011)	What are the requirements for effective collaborative filtering in m-commerce? Can the hybrid multi-channel method help consumers with a lack of knowledge in this channel?	Context: Multi-channel e-retailer in Taiwan Data: 1692 mobile phone training (1353) and test (339) users with 184 phone models; 1416 products, 194 categories, 2006–2007 timeframe	Experimental simulation analysis of proposed hybrid multi-channel method compared to other contemporary and more common methods
Martens et al. (2006)	In a public good threshold contribution game, do mobile agents enable them to move to best location for investment and collaboration? What can be learned for this for development of mobile agent-based e-commerce?	Context: Multi-agent system design with agents that use 6 different strategies Data: 600 game-play results, contributions for 1–3 goods, 6 experimental settings, 60 rounds	Design science research methodology used to create new capability and test its performance capabilities; includes economic theory and analysis as a basis for the overall investigation
Raphaeli et al. (2017)	How does consumer behavior change for an e-commerce website when access is via PC versus mobile phones? Can differences be seen for shopping-to-purchase conversion rates?	Context: Israeli e-retailer with 40% mkt share Data: Weblog files for 1 month in 2012, 264,651 consumer sessions tracked, with 9.8% and 19.8% buying conversion rates for mobile devices and PCs.	Analysis of consumer behavior through web usage mining, clickstream data and server-side log files, with predictive modeling of consumer purchases; example of CS and explanatory methods that illustrate CSS
Sumita and Yoshii (2010)	How did the entrance of mobile devices affect m-commerce's development? Can behavioral differences be seen for e-commerce consumers when there are different access channels?	Context: Differences in diffusion and uptake of m-commerce in the presence of mobile phones compared to earlier e-commerce patterns Data: None, focus is on economic math model	Analysis of e-commerce via PC access only versus m-commerce with PC and mobile phone access; focus on distribution and achieving penetration for selling some number of products

Table A3
Representative Social Commerce Empirical Research Published in ECRA (2013–2018).

Authors (Year)	Research Questions	Context and Data	Research Approach
Curty and Zhang (2013)	Can website feature changes over time be used to characterize the shift from e-commerce to s-commerce?	Context: e-commerce to s-commerce Data: 5 top e-commerce companies screen captures since their websites were established	Framework for historical analysis for website screen captures to depict transition Identified 174 emerging technical features
Gibreel et al. (2018)	What factors influenced shift from e-commerce to s-commerce in emerging markets?	Context: Twitter and WhatsApp in Kuwait Data: 137 survey respondents Analysis conducted with SPSS and AMOS	Online questionnaire to probe 11 constructs based on TAM, utility, WOM, behavior intention issues
Han et al. (2018b)	What is s-commerce research? What research methods have been used? What are areas for future research?	Context: Research literature on s-commerce Data: 407 academic articles, 2006–2017 Used literature taxonomy, theory and methods	Systematic review of the literature for classification
Huang and Benyoucef (2013)	Can individual, conversation, community, commerce, and sub-constructs identify s-commerce design approaches?	Context: Literature for past 20 years Data: Coverage on design constructs for e-commerce and extensions to s-commerce	Literature review to identify s-commerce design principles and related constructs to support improved design
Huang and Benyoucef (2017)	How do ineffective s-commerce design choices negatively affect consumer online purchase behavior?	Context: Experience with e-commerce and s-commerce websites in China Data: 262 students from Shaanxi Normal Univ.	Empirical survey study of purchase model with usability, functionality, sociability as s-commerce design factors in 5-stage decision process
Ko (2018)	Is a goal-directed behavior theory useful to assess how social and commercial desire drive social sharing and shopping intentions?	Context: Purchasing Facebook users in Taiwan Data: 248 participants via online survey	Survey method used to assess contents of theoretical model of Facebook-based shopping intentions for advertised products
Lu et al. (2016)	Is user trust affected by perceived effectiveness of institutional mechanisms and social presence in a market environment? Does IT influence consumer purchase intentions?	Context: Students at 2 Chinese universities, with 7 most-visited group-buying sites in China Data: 260 valid responses from 317 participants	Simulation and experimental approach to mimic real-world setting, with tasks to perform, and post-experimental survey; data analysis of structural model with PLS-SEM
Sun et al. (2016)	Does collaborative shopping tie in with members intention of purchasing? Does social climate add value to understanding behavioral intentions in friendship groups in s-commerce?	Context: WeChat in China Data: 215 members win a friendship group Methods: Survey and PLS empirical analysis	Tested hypotheses involving social climate, information, self-discovery, group intentions, purchase frequency
Xiao et al. (2015)	What influences intentions to initiate ties in social networks? How do network closure mechanisms work for buyers and sellers?	Context: Taobao community in China Data: Longitudinal network data involving 336 sellers and 146 buyers over 20 weeks in 2012	Tested hypotheses involving observational learning, social contagion, structural equivalence, reciprocity and homophily in social relationships
Zhang et al. (2016)	Do self-, social-, brand page-characteristics influence consumer brand loyalty?	Context: Brand pages within Weibo Data: Online survey of 442 participants	PLS-SEM assessment of structural model based on survey questionnaire for the context
Zhang et al. (2018)	How does a firm align its social media presence and related UGC to support business strategy? How do social media opinions influence firm-level innovation investments and support firm performance?	Context: List firms in Chinese stock market and social media user-generated content (UGC) Data: 5 years of panel data on 886 listed firms, and 6.2 mn relevant microblogs via Weibo	Used latent Dirichlet allocation (LDA) for text and topic analytics for UGC sentiment postings; carries out CS data analytics, econometric explanations in CSS

References

Aggarwal, R., Gopal, R., Gupta, A., Singh, H., 2012. Putting money where the mouths are: the relation between venture financing and electronic word-of-mouth. Inf. Syst. Res. 23 (3-part-2), 976–992.

Aigner, W., Miksch, S., Müller, W., Schumann, H., Tominski, C., 2008. Visual methods for analyzing time-oriented data. IEEE Trans. Vis. Comput. Graphics 14 (1), 47–60.

Akter, S., Wamba, S.F., 2016. Big data analytics in e-commerce: a systematic review and agenda for future research. Electr. Markets 26 (2), 173–194.

Alliance of Industrial Internet (AII), 2018. White paper on China industrial big data technology and application, 2017. Available from: www.aii-alliance.org/index.php?m=content&c=index&a=show&catid=23&id=142.

Andrews, M., Luo, X., Fang, Z., Ghose, A., 2016. Mobile ad effectiveness: hyper-contextual targeting with crowdedness. Marketing Sci. 35 (2), 218–233.

Balasubramanian, S., Peterson, R.A., Jarvenpaa, S., 2002. Exploring the implications of m-commerce for markets and marketing. J. Acad. Mark. Sci. 30 (4), 348–361.

Banker, R.D., Kauffman, R.J., Morey, R.C., 1990. Measuring gains in operational efficiency from information technology: a study of the Positran deployment at Hardee's Inc. J. Manage. Inf. Syst. 7 (2), 29–54.

Bauza, R., Gozálvez, J., 2013. Traffic congestion detection in large-scale scenarios using vehicle-to-vehicle communications. J. Network Comput. Appl. 36 (5), 1295–1307.

Benaroch, M., 2001. Option-based management of technology investment risk. IEEE Trans. Eng. Manage. 48 (4) 428 428.

Biuk-Aghai, R.P., Kou, W.T., Fong, S., 2016. Big data analytics for transportation: problems and prospects for its application in China. In: Proceedings of the IEEE Region 10 Symposium. IEEE Computer Society Press, Washington, DC, pp. 173–178.

Chan, S.W., Chong, M.W., 2017. Sentiment analysis in financial texts. Decis. Support Syst. 94, 53–64.

Chang, M.R., Ghosh, P., Jung, G., Kauffman, R.J., Zhang, P. 2013. Do household cable TV viewing patterns demonstrate efficiency and concentration? In: Proceedings of the 2013 China Summer Workshop on Information Management, Tianjin, China.

Chang, M.R., Kauffman, R.J., Kwon, Y.O., 2014. Understanding the paradigm shift to computational social science in the presence of big data. Decis. Support Syst. 63, 67–80.
Chang, M.R., Kauffman, R.J., Son, I., 2012. Consumer micro-behavior and TV viewership patterns: Data analytics for the two-way set-top box. In: Bichler, M., Kauffman, R.J., Lau, H.C., Yang, C., Yang, Y. (Eds.), Proceedings of the 14th Annual International Conference on Electronic Commerce. ACM Press, New York, pp. 272–273.
Chang, H.H., Wong, K.H., Li, S.Y., 2017. Applying push-pull-mooring to investigate channel switching behaviors: M-shopping self-efficacy and switching costs as moderators. Electron. Commer. Res. Appl. 24, 50–67.
Chang, Y.P., Zhu, D.H., 2011. Understanding social networking sites adoption in china: a comparison of pre-adoption and post-adoption. Comput. Hum. Behav. 27 (5), 1840–1848.
Chen, G.Q., Gang, W.U., Yuan-Dong, G.U., Ben-Jiang, L.U., Wei, Q., 2018a. The challenges for big data driven research and applications in the context of managerial decision-making: paradigm shift and research directions. Chin. J. Manage. Sci.
Chen, H., Chiang, R.H., Storey, V.C., 2012. Business intelligence and analytics: from big data to big impact. MIS Q. 1165–1188.
Chen, G., Kotz, D., 2000. A survey of context-aware mobile computing research. Working paper. Tuck School of Business, Dartmouth College, Hanover, NH.
Chen, K., Luo, P., Liu, L., 2018b. News, search and stock co-movement: Investigating information diffusion in the financial market. Electron. Commer. Res. Appl. 28, 159–171.
Chen, K., Yin, J., 2017. Information competition in product launch: evidence from the movie industry. Electron. Commer. Res. Appl. 26, 81–88.
China Academy of Information and Communications Technology (CAICT), 2018. White paper on big data. Beijing, China, April. Available at: www.caict.ac.cn/english/yjcg/bps/201805/t20180514_165149.htm.
China Electronics Standardization Institute (CESI), 2017. White paper on industrial big data. Beijing, China. Available from: www.cesi.cn/201703/2250.html.
Clemons, E.K., Dewan, R.D., Kauffman, R.J., Weber, T.A., 2017. Understanding the information-based transformation of strategy and society. J. Manage. Inf. Syst. 34 (2), 425–456.
Clemons, E.K., Gao, G.G., Hitt, L.M., 2006. When online reviews meet hyperdifferentiation: a study of the craft beer industry. J. Manage. Inf. Syst. 23 (2), 149–171.
Cohen, D., 2014. Advertisers on Facebook have some new ways to conduct A/B tests. Adweek, April 6. Available at: www.adweek.com/digital/advertisers-onfacebook-have-some-new-ways-to-conduct-a-b-tests/.
Consultancy.uk, 2017. Venture capital investment in fintech reaches record $27.4 billion high. April 17. Available from: www.consultancy.uk/news/16707/venture-capital-investment-in-fintech-reaches-record-274-billion-high.
Curty, R.G., Zhang, P., 2013. Website features that gave rise to social commerce: a historical analysis. Electron. Commer. Res. Appl. 12 (4), 260–279.
De Marchi, G., Lucertini, G., Tsoukias, A., 2016. From evidence-based policy-making to policy analytics. Ann. Oper. Res. 236 (1), 15–38.
Ding, Y., Xu, H., Tan, B., 2015. A conflicting relationship between online product page viewing and product returns. Proceedings of the 2015 International Conference on Information Systems. AIS Press, Atlanta, GA.
Edin, F., Machens, C.K., Schütze, H., Herz, A.V., 2004. Searching for optimal sensory signals: iterative stimulus reconstruction in closed-loop experiments. J. Comput. Neurosci. 17 (1), 47–56.
Efron, B., Hastie, T., 2016. Computer Age Statistical Inference Vol. 5 Cambridge University Press, Cambridge.
Erevelles, S., Fukawa, N., Swayne, L., 2016. Big data consumer analytics and the transformation of marketing. J. Bus. Res. 69 (2), 897–904.
Evidence-Based Policy-Making Collaborative, 2016. Principles of evidence-based policy-making. White paper, American Enterprise Institute, the Pew-MacArthur First Results Initiative, the Brookings Institution, and the Urban Institute, Washington, DC, September.
Fan, S., Kanbur, R., Zhang, X., 2011. China's regional disparities: experience and policy. Rev. Dev. Finance 1 (1), 47–56.
Fang, Z., Gu, B., Luo, X., Xu, Y., 2015. Contemporaneous and delayed sales impact of location-based mobile promotions. Inf. Syst. Res. 26 (3), 552–564.
Feng, Z.Y., Guo, X.H., Zeng, D.J., Chen, Y.B., 2013. On the research frontiers of business management in the context of big data. Chin. J. Manage. Sci.
Fiorini, P.D.C., Seles, B.M.R.P., Jabbour, C.J.C., Mariano, E.B., 2018. Management theory and big data literature: From a review to a research agenda. Int. J. Inf. Manage. 43, 112–129.
Fong, N.M., Fang, Z., Luo, X., 2015. Geo-conquesting: competitive locational targeting of mobile promotions. J. Mark. Res. 52 (5), 726–735.
Gan, C., Wang, W., 2015. Uses and gratifications of social media: a comparison of microblog and WeChat. J. Syst. Inf. Technol. 17 (4), 351–363.
Gao, B., Chan, W., Chi, L., Deng, X., 2016. Size and growth dynamics of online stores: a case of China's Taobao.com. Electr. Commer. Res. Appl. 17, 161–172.
Ghasemaghaei, M., Ebrahimi, S., Hassanein, K., 2018. Data analytics competency for improving firm decision making performance. J. Strat. Inf. Syst. 27 (1), 101–113.
Gibreel, O., Al Otaibi, D.A., Altmann, J., 2018. Social commerce development in emerging markets. Electron. Commer. Res. Appl. 27, 152–162.
Goel, V., 2014. Facebook tinkers with users' emotions in news feed experiment, stirring outcry. New York Times, June 30.
Goel, A.L., Okumoto, K., 1979. Time-dependent error-detection rate model for software reliability and other performance measures. IEEE Trans. Reliab. 28 (3), 206–211.
Goes, P.B., 2014. Big data and IS research. MIS Q. 38, 3, iii-viii.
Gomber, P., Kauffman, R.J., Parker, C., Weber, B.W., 2017. On the fintech revolution: interpreting the forces of innovation, disruption and financial services transformation. J. Manage. Inf. Syst. 35 (1), 220–265.
Grover, V., Chiang, R.H., Liang, T.P., Zhang, D., 2018. Creating strategic business value from big data analytics: a research framework. J. Manage. Inf. Syst. 35 (2), 388–423.
Gu, B., Konana, P., Rajagopalan, B., Chen, H.W.M., 2007. Competition among virtual communities and user valuation: the case of investing-related communities. Inf. Syst. Res. 18 (1), 68–85.
Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S.F., Childe, S.J., Hazen, B., Akter, S., 2017. Big data and predictive analytics for supply chain and organizational performance. J. Bus. Res. 70, 308–317.
Guo, X., Zhang, X., Sun, Y., 2016. The privacy–personalization paradox in m-health services acceptance of different age groups. Electron. Commer. Res. Appl. 16, 55–65.
Han, H.Q., 2018. Fintech in China: An Introduction. Wharton Fintech, Philadelphia, PA Available from: medium.com/wharton-fintech/fintech-in-china-an-introduction-6b11abd9cb64.
Han, S., Hao, X., Huang, H., 2018a. An event-extraction approach for business analysis from online Chinese news. Electron. Commer. Res. Appl. 28, 244–260.
Han, H., Xu, H., Chen, H., 2018b. Social commerce: a systematic review and data synthesis. Electron. Commer. Res. Appl. 30, 38–50.
Hoang, A.P., Kauffman, R.J., 2018a. Content sampling, household informedness, and the consumption of digital information goods. J. Manage. Inf. Syst. 35 (2), 575–609.
Hoang, A.P., Kauffman, R.J., 2018b. Censored observation recovery for causal inference using temporal sequences and iterative data simulation. Presented at the 2018 Workshop on Statistical Challenges in Electronic Commerce. Erasmus University, Rotterdam, Netherlands June.
Hoang, A.P., Kauffman, R.J., Li, T., 2018. Uncensored: Iterative data simulation and probabilistic inference method to recover censored data. Working paper. School of Information Systems, Singapore Management University, Singapore.
Hofstede, G., 1984. National cultures revisited. Asia Pac. J. Manage. 2 (1), 22–28.
Hofstede, G., 2006. What did GLOBE really measure? Researchers' minds versus respondents' minds. J. Int. Bus. Stud. 37, 882–896.
Hsiao, K.L., Chen, C.C., 2016. What drives in-app purchase intention for mobile games? An examination of perceived values and loyalty. Electron. Commer. Res. Appl. 16, 18–29.
Hsu, C.L., Lin, J.C.C., 2015. What drives purchase intention for paid mobile apps? An expectation confirmation model with perceived value. Electron. Commer. Res. Appl. 14, 46–57.
Huang, J., Jiang, X., Tang, Q., 2009. An e-commerce performance assessment model: its development and an initial test on e-commerce applications in the retail sector of China. Inf. Manage. 46 (2), 100–108.
Huang, Z., Benyoucef, M., 2013. From e-commerce to social commerce: a close look at design features. Electron. Commer. Res. Appl. 12 (4), 246–259.
Huang, Z., Benyoucef, M., 2017. The effects of social commerce design on consumer purchase decision-making: an empirical study. Electron. Commer. Res. Appl. 25, 40–58.
Ittmann, H.W., 2015. The impact of big data and business analytics on supply chain management. J. Transp. Supply Chain Manage. 9 (1), 1–9.
Jeske, M., Grüner, M., Weiß, F., 2013. Big data in logistics: a DHL perspective on how to move beyond the hype. DHL Customer Solut. Innov. 12.
Kauffman, R.J., Kim, K., Lee, T.S.Y., 2016. Computational social science fusion analytics: combining machine-based methods with explanatory empiricism. In: Sprague, R., Bui, T. (Eds.), Proceedings of the 2016 Hawaii International Conference on System Sciences. ACM Press, New York, pp. 5048–5057.
Kauffman, R.J., Kim, K., Lee, T.S.Y., Hoang, A.P., Ren, J., 2017. Combining machine-based and econometric methods for policy analytics insights. Electron. Commer. Res. Appl. 25, 115–140.
Kauffman, R.J., Lee, D., 2010. A multi-level theory approach to understanding price rigidity in Internet retailing. J. Assoc. Inf. Syst. 11 (6), 303–338.
Kauffman, R.J., Spauding, T., Wood, C.A., 2009. Are online auction markets efficient? An empirical study of market liquidity and abnormal returns. Decis. Support Syst. 48 (1), 3–13.
Kharas, H., 2017. The unprecedented expansion of the global middle class: an update. Brookings, India, New Delhi.
Kim, Y., Kauffman, R.J., Hoang, A., Jung, G., 2018. Fixed and mobile broadband bandwidth choices: How did smartphone adoption transform Singapore's service market? Working paper. School of Information Systems, Singapore Management University, Singapore.
Kitchin, R., 2014. Big data, new epistemologies and paradigm shifts. Big Data Soc. 1, 1.
Ko, H.C., 2018. Social desire or commercial desire? The factors driving social sharing and shopping intentions on social commerce platforms. Electron. Commer. Res. Appl. 28, 1–15.
Koppel, M., Schler, J., 2006. The importance of neutral examples for learning sentiment. Comput. Intell. 22 (2), 100–109.
Kohavi, R., Deng, A., Longbotham, R., Xu, Y., 2014. Seven rules of thumb for web site experimenters. In: Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. ACM Press, New York, pp. 1857–1866.
Larmer, B., 2017. China's revealing spin on the sharing economy. New York Times November 20.
Lazer, D., Pentland, A.S., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M., 2009. Life in the network: the coming age of computational social science. Science 323 (5915), 721–723.
Lecun, Y., Benjio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Lee, J.Y., Fang, E., Kim, J.J., Li, X., Palmatier, R.W., 2018. The effect of online shopping platform strategies on search, display, and membership revenues. J. Retail. 94 (3), 247–264.
Li, T., Kauffman, R.J., van Heck, E., Vervest, P., Dellaert, J., 2014. Consumer informedness and firm information strategy. Inf. Syst. Res. 25 (2), 345–363.

Liang, L., Yang, Y., 2018. An experimental study of Chinese tourists using a company-hosted WeChat official account. Electron. Commer. Res. Appl. 27, 83–89.
Lien, C.H., Cao, Y., 2014. Examining WeChat users' motivations, trust, attitudes, and positive word-of-mouth: evidence from China. Comput. Hum. Behav. 41, 104–111.
Lien, C.H., Cao, Y., Zhou, X., 2017. Service quality, satisfaction, stickiness, and usage intentions: an exploratory evaluation in the context of WeChat services. Comput. Hum. Behav. 68, 403–410.
Lin, J., Lu, Y., Wang, B., Wei, K.K., 2011. The role of inter-channel trust transfer in establishing mobile commerce trust. Electron. Commer. Res. Appl. 10, 615–625.
Long, D., 2017. China's e-commerce market to pass $1.1 trillion in 2017. eMarketer July 5.
Liu, D.R., Liou, C.H., 2011. Mobile commerce product recommendations based on hybrid multiple channels. Electron. Commer. Res. Appl. 10, 94–104.
Liu, J., 2018. The dramatic rise and fall of online P2P lending in China. Techcrunch August 1. Available from: techcrunch.com/2018/08/01/the-dramatic-rise-and-fall-of-online-p2p-lending-in-china/.
Lu, B., Zeng, Q., Fan, W., 2016. Examining macro-sources of institution-based trust in social commerce marketplaces: an empirical study. Electron. Commer. Res. Appl. 20, 116–131.
Luo, X., Andrews, M., Fang, Z., Phang, C.W., 2013. Mobile targeting. Manage. Sci. 60 (7), 1738–1756.
Lu, M., Lai, C., Ye, T., Liang, J., Yuan, X., 2015. Visual analysis of route choice behavior based on GPS trajectories. In: Proceedings of the 2015 IEEE Conference on Visual Analytics Science and Technology. IEEE Computer Society Press, Washington, DC, pp. 203–204.
Lv, Z., Jin, Y., Huang, J., 2018. How do sellers use live chat to influence consumer purchase decision in China? Electron. Commer. Res. Appl. 28, 102–113.
Malhotra, R., Bahl, L., Sehgal, S., Priya, P., 2017. Empirical comparison of machine learning algorithms for bug prediction in open source software. Proceedings of the 2017 International Conference on Big Data Analytics and Computational Intelligence. IEEE Computer Society Press, Washington, DC.
Martens, R., Paranjape, R., Benedicenti, L., Sankaran, S., Sadanand, A., 2006. Mobile agent strategies for the provision of public goods: an experimental study. Electron. Commer. Res. Appl. 5, 140–146.
Morgan Stanley Research, 2015. China's E-commerce Revolution. Blue paper, New York.
Müller, O., Junglas, I., Brocke, J.V., Debortoli, S., 2016. Utilizing big data analytics for information systems research: challenges, promises and guidelines. Eur. J. Inf. Syst. 25 (4), 289–302.
Natali, F., Carley, K.M., Zhu, F., Huang, B., 2017. The role of different tie strength in disseminating different topics on microblogs. Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE Computer Society Press, Washington, DC.
National Academies of Sciences, Engineering, and Medicine (NASEM), 2016. Research methodology and principles: Assessing causality. Chapter 6 in Commercial Motor Vehicle Driver Fatigue, Long-Term Health, and Highway Safety: Research Needs. The National Academies Press, Washington, DC.
Office of the United States Trade Representative, 2017. 2017 Out-of-cycle Review of Notorious Markets. Executive Office of the President, U.S. Government, Washington, DC.
Pearl, J., 2009. Causality: Models, Reasoning, and Inference, second ed. Cambridge University Press, New York.
Prabu, K., 2014. Baidu leverages big data to visualize Chinese New Year travel patterns. Tnooz January 28.
Qin, R., Yuan, Y., Wang, F., 2017. Exploring the optimal granularity for market segmentation in RTB advertising via computational experiment approach. Electron. Commer. Res. Appl. 24, 68–83.
Quah, E., Haldane, J.B.S., 2007. Cost-Benefit Analysis. Routledge, London.
Raphaeli, O., Goldstein, A., Fink, L., 2017. Analyzing online consumer behavior in mobile and PC devices: a novel web usage mining approach. Electron. Commer. Res. Appl. 26, 1–12.
Rana, R., Stardon, M., Berger, C., Hansson, J., Nilsson, M., Meding, W., 2014. The adoption of machine learning techniques for software defect prediction: an initial industrial validation. In: Proceedings of the Joint Conference on Knowledge-Based Software Engineering, Communications in Computer and Information Science, Vol. 466. Springer, Berlin/Heidelberg, Germany, pp. 270–285.
Razeghian, M., Weber, T.A., 2019. The advent of the sharing culture and its effect on product pricing. Electron. Commer. Res. Appl. 33, 100801.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.
Savicky, P., Fürnkranz, J., 2003. Combining pairwise classifiers with stacking. Proceedings of the 2003 Cryptographic Hardware and Embedded Systems Conference. Springer, Berlin/Heidelberg, Germany.
Schlager, T., Hildebrand, C., Häubl, G., Franke, N., Herrmann, A., 2018. Social product-customization systems: peer input, conformity, and consumers' evaluation of customized products. J. Manage. Inf. Syst. 35 (1), 319–349.
Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin, Boston.
Sims, C., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.
Song, P., Zheng, C., Zhang, C., Yu, X., 2018. Data analytics and firm performance: an empirical study in an online B2C platform. Inf. Manage. 55 (5), 633–642.
Stufken, J., 1996. Optimal crossover designs. In: S. Ghosh, C.R. Rao (Eds.), Design and Analysis of Experiments. Handbook of Statistics, 13, North-Holland, Amsterdam, pp. 63–90.
Sumita, U., Yoshii, J., 2010. Enhancement of e-commerce via mobile accesses to the Internet. Electron. Commer. Res. Appl. 9, 217–227.
Sun, Y., Wei, K.K., Fan, C., Lu, Y., Gupta, S., 2016. Does social climate matter? On friendship groups in social commerce. Electron. Commer. Res. Appl. 18, 37–47.
Tan, C.H., Sutanto, J., Phang, C.W., Gasimov, A., 2014. Using personal communication technologies for commercial communications: a cross-country investigation of email and SMS. Inf. Syst. Res. 25 (2), 307–327.
Tian, F., Wu, F., Chao, K.M., Zheng, Q., Shah, N., Lan, T., Yue, J., 2016. A topic sentence-based instance transfer method for imbalanced sentiment classification of Chinese product reviews. Electron. Commer. Res. Appl. 16, 66–76.
Tian, J., Zhang, Y., Zhang, C., 2018. Predicting consumer variety-seeking through weather data analytics. Electron. Commer. Res. Appl. 28, 194–207.
Tong, F., 2017. Online retail sales in China soar past $1 trillion in 2017. DigitalCommerce360.com, February 8. Available at: www.digitalcommerce360.com/2018/02/08/online-retail-sales-china-soar-past-1-trillion-2017/.
Tosun, A., Bener, A., Kale, R., 2010. AI-based software defect predictors: Applications and benefits in a case study. Proceedings of the Twenty-Second Innovative Applications of Artificial Intelligence Conference. American Artificial Intelligence Association, Menlo Park, CA.
Tse, E., Mellor, I., 2016. Chinese companies lead the way in fintech innovation. Nikkei Asian Rev September 8.
Tsoukiàs, A., Montibeller, G., Lucertini, G., Belton, V., 2013. EURO J. Decis. Process. 1 (1–2), 115–134.
Van Den Broeke, M., Boute, R.N., Mieghem, J.A.V., 2018. Platform flexibility strategies: R&D investment versus production customization tradeoff. Eur. J. Oper. Res in press.
VentureScanner. Start-up market reports and data. San Francisco, CA, December 2017.
Vuylsteke, A., Wen, Z., Baesens, B., Poelmans, J., 2010. Consumers' search for information on the Internet: How and why China differs from Western Europe. J. Interact. Mark. 24, 309–331.
Walters, J., Kuo, Y., Jap, W., Hsu, H., 2011. The World's Next E-commerce Superpower. White paper, Boston Consulting Group, New York.
Wang, J., Du, J., Chiu, Y., Li, J., 2018a. Dynamic effects of customer experience levels on durable product satisfaction: Price and popularity moderation. Electron. Commer. Res. Appl. 28, 16–29.
Wang, W., Dollar, D., 2018. What's Happening with China's Fintech Industry? Brookings Institution, Washington, DC February 8.
Wang, N., Li, Q., Liang, H., Ye, T., Ge, S., 2018b. Understanding the importance of interaction between creators and backers in crowdfunding success. Electron. Commer. Res. Appl. 27, 106–117.
Weber, T.A., 2014. Intermediation in a sharing economy: Insurance, moral hazard, and rent extraction. J. Manage. Inf. Syst. 31 (3), 35–71.
Weber, T.A., 2017. Smart products for sharing. J. Manage. Inf. Syst. 34 (2), 341–368.
Wells, P., 2014. New labour and evidence-based policy-making. Seminar paper, Political Economy Research Centre/Centre for Regional Economic and Social Research, Sheffield Hallam University, Sheffield, UK May 16.
Wen, X., Zhou, X., 2016. Servitization of manufacturing industries based on cloud-based business model and the down-to-earth implementary path. Int. J. Adv. Manuf. Technol. 87 (5–8), 1491–1508.
Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., 2016. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Burlington, MA.
Wu, J., Li, H., Liu, L., Zheng, H., 2017. Adoption of big data and analytics in the mobile healthcare market: an economic perspective. Electron. Commer. Res. Appl. 22, 24–41.
Xiao, B., Huang, M., Barnes, M.J., 2015. Network closure among sellers and buyers in social commerce community. Electron. Commer. Res. Appl. 14, 641–653.
Xinhua, 2018. Winter has come for physical outlets in the age of e-commerce. ChinaDaily.com.cn. January 12. Available at: www.chinadaily.com.cn/a/201801/12/WS5a580f0ea3102c394518eb60.html.
Xu, Z.B., Feng, Z.Y., Guo, X.H., Zeng, D.J., Chen, G.Q., 2014. The research frontiers of big data driven management and decision-making. Manage. World 11, 158–163.
Yan, J., 2017. China's booming sharing economy: why a bust may be near. For. Affairs November 9.
Yang, M., Adomavicius, G., Burtch, G., Ren, Y., 2018. Mind the gap: accounting for measurement error and misclassification in variables generated via data mining. Inf. Syst. Res. 29 (1), 4–24.
Yao, X., Lu, X., Phang, C.W., Ba, S., 2017. Dynamic sales impacts of online physical product sampling. Inform. Manage. 5 (4), 599–612.
Yu, C., Zhu, L., 2016. Product design pattern based on big data-driven scenario. Adv. Mech. Eng. 8, 7.
Zhang, W., Ding, G., Chen, L., Li, C., Zhang, C., 2013. Generating virtual ratings from chinese reviews to augment online recommendations. ACM Trans. Intell. Syst. Technol. 4 (1), 9.
Zhang, C.B., Li, Y.N., Wu, B., Li, D.J., 2016. How WeChat can retain users: roles of network externalities, social interaction ties, and perceived values in building continuance intention. Comput. Hum. Behav. 69.
Zhang, L., Wang, H., Li, Q., Zhao, M.H., Zhan, Q.M., 2018. Big data and medical research in China. BMJ 360, j5910.
Zhong, R.Y., Newman, S.T., Huang, G.Q., Lan, S., 2016. Big data for supply chain management in the service and manufacturing sectors: challenges, opportunities, and future perspectives. Comput. Ind. Eng. 101, 572–591.
Zhou, M., Wang, H., Zhu, J., Chen, W., Wang, L., Liu, S., Li, Y., Wang, L., Liu, Y., Lin, P., Liu, J., Yu, S., Tan, F., Barber, R.S., Coates, M.M., Dicker, D., Fraser, M., Gonzalez-Medina, D., Hamavid, H., Hao, Y., Hu, G., Jiang, G., Kan, H., Lopez, A.D., Phillips, M.R., She, J., Vos, T., Wan, X., Xu, G., Yan, L.L., Yu, C., Zhao, Y., Zheng, Y., Zou, X., Naghavi, M., Wang, Y., Murray, C.J., Yang, G., Liang, X., 2016. Cause-specific mortality for 240 causes in China during 1990–2013: a systematic subnational analysis for the Global Burden of Disease Study 2013. Lancet 387 (10015), 251–272.