Economics Bulletin

Volume 37, Issue 2

The bigger, the better? An investigation of optimal volume of big data

Bonyoung Koo
Logos Logistics Inc.

Seung Ho Yoo
Sunmoon University

Byung Cho Kim
Korea University

Abstract

Big data has changed the way modern businesses run. Recently, big data has drawn huge attention from academic researchers, industry practitioners, and policy makers. While the common belief about big data is that many benefits come from the data volume, existing studies identify challenges associated with big data volume such as privacy concerns and management cost. Given the trade-off between the opportunities and the challenges from big data, we presume that there exists an optimal volume of data which may contradicts to a common misperception that the bigger volume guarantees more benefit. Grounded on a game-theoretic model, we analyze a monopolistic e-tailer's decision on the optimal volume of data to collect and the price of the product to sell in the presence of two different types of customers in terms of their privacy-sensitivity. Our preliminary results show how the number of privacy-sensitive customers and their utility loss influence the optimal information level and the price. We then explore the welfare implications by comparing the profit-maximizing data volume with welfare-maximizing level. We aim to contribute to the literature by characterizing the optimal volume of big data while modelling customers' privacy-sensitivity. In the presence of privacy and cost challenges, our results may give them implications for strategic decision-making for practitioners and policy makers.

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A8017277)

Citation: Bonyoung Koo and Seung Ho Yoo and Byung Cho Kim, (2017) "The bigger, the better? An investigation of optimal volume of big data", *Economics Bulletin*, Volume 37, Issue 2, pages 871-879
Contact: Bonyoung Koo - byk8774@gmail.com, Seung Ho Yoo - shy1228@sunmoon.ac.kr, Byung Cho Kim - bkim@korea.ac.kr.
Submitted: August 02, 2016. **Published:** April 22, 2017.

1. Introduction

Big data has changed the way modern businesses run. Information extracted from big data enables managers to know more about the nature of their consumers and the business circumstances, and thus, make better decisions under uncertainty. In the current information age, ability to collect and analyze big data has provided companies with competitive advantage, and it is now becoming almost a commodity for modern businesses (Gartner 2015). In academia as well, big data analytics is one of the current and future research frontiers.

Companies use big data to draw useful information, make measurements, and detect trends. According to McKinsey, big data will become a key basis of competition, thus companies who miss big data opportunities will miss new waves of productivity growth, innovation and consumer surplus (McKinsey Global Institute, 2011). Big data is changing game plans and shifting paradigm for many industries. Consider retail for example. Walmart, the largest retailer in the world with 20,000 stores in 28 countries, has seen the value of big data for long. Walmart captures and analyzes an immense amount of data through its Data Café, a concept with a broad range of data warehousing capabilities, aiming to provide a large cross-section of employees with timely information so that they make better strategic decisions (Forbes, 2016). In the manufacturing industry, big data has been an enabler of the vast improvements in supply planning and product quality. Big data provides a manufacturer with process transparency, which resolves uncertainties in the manufacturing process, such as inconsistent performance or availability. Predictive manufacturing is an applicable approach toward near-zero downtime, which requires a large amount of data and advanced prediction tools for the systematic process of turning data into useful information (TATA Consultancy Services, 2013). The financial services sector is not an exception. Morgan Stanley who used to do a portfolio analysis on traditional databases, shifted to Hadoop to perform a sentiment analysis, predictive analytics, and financial trade on a large scale (Forbes, 2012). Bitcoin, the fast growing but still chaotic digital currency, is another example as it is gaining much from the power of big data analytics. Big data opens up numerous opportunities for the public sector as well, helping combat poverty, crime and pollution (New York Times, 2012). Big data has significant implications on national security as demonstrated by a joint effort made by the American Association for the Advancement of Science (AAAS), the Federal Bureau of Investigation (FBI) and the United Nations Interregional Crime and Justice Research Institute (UNICRI) to develop frameworks for risk and benefit assessments of big data and identify options for U.S. government action to mitigate potential risks by utilizing big data analytics (AAAS-FBI-UNICRI, 2014).

Certainly, many of the benefits come from the scale of big data. However, the core of big data analytics is the "analytics" part, in which useful information is drawn through data mining and analysis. But the term "big" data often forms misperception that the bigger scale is always better. Laney (2001) identifies volume, velocity, and variety, known as the 3Vs, to characterize the concept of big data, where volume is the size of the data set, velocity indicates the speed of the data transmitting in and out, and variety describes the range of data types and sources. Now, a fourth V can be considered value, variability, or virtual (Zikopoulos, 2011). Despite the benefits, the volume of big data presents challenges as well (Ahrens, 2011). When handling big data problems, difficulties lie in data capture, storage, searching, sharing, analysis, and visualization. While big data is meant to facilitate decision-making, big data must be complemented by "big judgment" (Shvetank, 2012). Handling and managing "too" big data creates unnecessary and significant burden on the firms. Another challenge presented by big data is consumers' privacy concern. Some types of data could lead to privacy issues, including communications records, location information, video surveillance, and financial status (Cumbley, 2013). Furthermore, considerable cultural challenges and privacy concerns are also going to become more significant (Shvetank, 2012).

Grounded on a game-theoretic model, we analyze a monopolistic online retailer's decision on the optimal volume of data to collect and the price of the product to sell. Reflecting reality, our model considers two different types of customers in terms of their privacy-sensitivity. Our preliminary results show how the number of privacy-sensitive customers and their utility loss influence the optimal data volume and the price. While the firm collects more data with less privacy-sensitive customers and less utility loss, interestingly, their impact on price is not trivial in that it is moderated by the dynamics in the aggregate privacy loss to privacy-sensitive customers and the aggregate big data benefit to privacy-insensitive customers. We then explore the welfare implications by comparing the profit-maximizing data volume with welfare-maximizing level. Our findings indicate that, even from welfare-maximizing perspective, there exists an optimal data volume, implying that collecting too much data is socially harmful. The social optimum turns to be higher than the profit-maximizing data volume, implying that the social planner has an incentive to encourage the firms to increase data volume.

We aim to contribute to the literature by characterizing the optimal volume of big data while modelling customers' privacy-sensitivity. To our knowledge, this is one of the few attempts to adopt the cost and privacy challenges in the big data context which becomes a critical success factor for modern businesses. Practically, firms and policy makers may use our findings as a guideline when they determine the scope of data collection. In the presence of privacy and cost challenges, our results may give them implications for strategic decision-making.

This paper is organized as follows. Section 2 reviews a relevant literature. In Section 3, we present a baseline model and preliminary results. Section 4 explores welfare implications, and finally, Section 5 concludes.

2. Literature Review

Big data analytics is becoming critical for modern businesses, and thus, attracts interest from academic researchers in applied economics and business. Although big data analytics produces numerous benefits for businesses, it is not free for them, as some concerns are identified in the existing literature, including data management cost and privacy concerns. Big data is costly and challenging to deal with. George et al. (2014) describe the management issues around big data, specifically the service level agreements (SLA) that define the nature and quality of information technology services, and depict big data-sharing agreements as poorly structured and informal. They consider the methodologies of analyzing big data and state that it is easy to get false correlations when using typical statistical tools. Hazen et al. (2014) argue that handling data with poor quality is extremely costly for organizations whose competitive advantage comes from big data analytics since sizable data often has noise. According to Hazen et al. (2014), "The costs of poor data quality have been estimated to be as high as 8% to 12% of revenues for a typical organization and may generate up to 40% to 60% of a service organization's expenses." They introduce and emphasize the need for monitoring and controlling the data volume and quality in supply chain management processes and provide a starting point for future research and applications.

In addition to the cost side of big data management, consumers have privacy concerns. Lu et al. (2014) formalize the general architecture of big data analytics. While big data creates value for economic growth and technological innovation, the deluge of data also raises new privacy concerns. Thus, existing studies identify corresponding privacy requirements and introduce methods that efficiently preserve privacy in response to the rapid development of data mining techniques. Lee et al. (2013) investigate how perceived personal information in social networking services (SNS) affects user's risk perception and information privacy concerns. They also show that information on the SNS can significantly increase users' perceived level of privacy risks. In turn, these perceived privacy risks lead to an increased level of privacy concerns. Malhotra et al. (2004) argue that the lack of consumer confidence in information privacy is a major problem hampering the growth of e-commerce. Privacy and security risks from high volume of data, complex data-sharing and accessibility-related issues exist in a big data environment. The existing security solutions are not sufficient to handle the scale, speed, variety and complexity of big data. Most organizations lack systematic approaches for ensuring appropriate data access mechanisms. Boyd and Crawford (2012) discuss six provocations to spark conversations about the issues of big data by listing examples. They argue that collecting more data is not always good given the ethical and privacy issues that can arise with sizable data.

Given the trade-off between the benefits and privacy loss from big data, it would be reasonable to presume that the optimal data size is finite. Pursuing optimal scale in the presence of positive and negative externalities is not new in the existing literature. A common view in the domain of finance is that the optimal market size is determined based on the trade-off between positive externalities and costs of intermediation (Madhavan 2000, Bias et al. 2005). Theoretically, without costs, the optimal scale should be infinity which is rarely observed in reality. Researchers have identified other factors than cost which contributes to the finite optimal scale at equilibrium. One of those is negative externality (e.g., Spiegel and Subrahmanyam 1992). In a recent study, Kawakami (2013) shows that the optimal market size is finite even in the absence of intermediation costs when the negative network externalities are present.

In the stream of big data research, little has examined the optimal volume of data despite the importance. This paper aims to understand how the optimal volume of big data is determined at equilibrium. Grounded on a game-theoretic model, we analyze a monopolistic online retailer's decision on the optimal volume of data in the presence of customers who are different in sensitivity to privacy. Our work is based on the model from the literature on information economics where security and privacy concerns matter (Chen and Png 2003). Considering the scarcity of analytical work in the big data research stream, our work aims to make modelling contribution by considering two customer types, including privacy-sensitive customers and privacy-insensitive ones and analysing the impact on optimal data volume. Our paper aims to bridge the gap in the big data literature by characterizing the optimal data volume in the presence of the trade-off associated with data collection and handling.

3. Model

We presume that there exists an optimal volume of big data. In this section, we present a modelling framework that characterizes the optimal data volume to balance the advantages (e.g., forecasting accuracy) and disadvantages (e.g., management cost and privacy risk) of big data. As a benchmark, our model considers a monopolistic online retailer (or e-tailer) who gathers data from its customers at the time of purchase. For example, Amazon collects customer data at the time of purchase. The flow of decision making is as follows. The monopolistic online retailer determines the data volume, i.e., how much information to request from customers, denoted with v. Then the online retailer sets the price for the product, represented with p. Given v and p, customers make their purchase decision, i.e., they decide whether they pay price for the product while agreeing to offer their private information requested by the online retailer, or they do not buy.

Our model considers two types of customers regarding their sensitivity to privacy. Privacy-sensitive customers with fraction ρ, incur severe utility loss when they provide their personal information to the seller, i.e., utility loss due to privacy concern exceeds benefit from services enabled by big data, such as customization. The other type of customers, i.e., privacy-insensitive ones with fraction $1-\rho$, are not seriously concerned about privacy-related issues, thus, they enjoy benefits more than privacy loss from services enabled by big data analytics. For simplicity, we assume marginal cost to be negligible.

Analyzing customer behaviour leads to privacy-sensitive customers' utility as represented below:

$$U_S = q - \alpha v - p,$$

where q is the customers' perceived value of the product which is uniformly distributed on [0,1], α is the scale parameter for the net utility loss of privacy-sensitive customers from offering their private information, v is the information level, and p is the price of the product. Note that S in the subscript denotes privacy-"S"ensitive customers. The privacy-"I"nsensitive customers' utility becomes:

$$U_I = q + \beta v - p,$$

where β is the scale parameter for the net utility gain of the privacy-insensitive customers. Customers whose utility turns positive would buy the product, leading to the demand as $1-\alpha v - p$ and $1+\beta v - p$, from privacy-sensitive customers and privacy-insensitive ones, respectively. Thus, the total demand for the product becomes the following:

$$D = \rho(1-\alpha v - p) + (1-\rho)(1+\beta v - p).$$

Let $C(v)$ be the cost for data management, which is increasing in v, implying that the firm incurs more cost to manage higher volume of data. We assume a quadratic form, i.e., cv^2 for simplicity. Then the firm's profit function becomes

$$\pi = Dp - C(v) = \big(\rho(1-\alpha v - p) + (1-\rho)(1+\beta v - p)\big)p - cv^2.$$

A further analysis leads to Proposition 1 which characterizes the equilibrium:

Proposition 1: *There exists an equilibrium at which the optimal data volume and price are as follows:*

$$v^* = \frac{\beta(1-\rho)-\alpha\rho}{4c-(\beta(1-\rho)-\alpha\rho)^2} \quad \textit{and} \quad p^* = \frac{2c}{4c-(\beta(1-\rho)-\alpha\rho)^2}.$$

Proof: Solving for the profit-maximizing price, i.e., $\frac{\partial\pi}{\partial p}=0$, leads to $p^*(v) = \frac{1+v(\beta-(\alpha+\beta)\rho)}{\partial p}$. At $p^*(v)$, we solve for the profit-maximizing volume, i.e., $\frac{\partial\pi}{\partial v}\Big|_{p^*(v)} = 0$, which yields the optimal volume as $v^* = \frac{\beta(1-\rho)-\alpha\rho}{4c-(\beta(1-\rho)-\alpha\rho)^2}$. Note that the second-order condition is $\frac{\partial^2\pi}{\partial v^2}\Big|_{p^*(v)} = -\frac{4c-(\beta(1-\rho)-\alpha\rho)^2}{2}$. For the second-order condition to be negative, $4c-(\beta(1-\rho)-\alpha\rho)^2 > 0$. Replacing v^* in $p^*(v)$ leads to the optimal price level as $p^* = \frac{2c}{4c-(\beta(1-\rho)-\alpha\rho)^2}$. ■

Proposition 1 indicates that there exists an optimal data volume which is finite, given the trade-off between the benefit from big data analytics and the cost of managing data, combined with privacy concern. Our findings show that it is not always better to collect more data. Rather, firms are better off to collect data at the finite optimal level, considering their capability in data management and the potential customers' privacy sensitivity. A key factor that plays a role in determining the optimal data volume is the interplay between the aggregate benefit from big data and the aggregate privacy loss. Technically, when the aggregate loss exceeds benefit ($\beta(1-\rho) < \alpha\rho$), the maximization problem has a boundary solution, indicating that the firm has no incentive to collect data. Since this case is uninteresting, hereafter, we only consider the case when the aggregate benefit from big data is greater than the aggregate privacy loss ($\beta(1-\rho) > \alpha\rho$). Next, we analyze the impact of privacy-sensitivity parameters on the optimal data volume. Proposition 2 summarizes our findings:

Proposition 2: *As more customers become privacy-sensitive, it is optimal for the firm to reduce data volume* $\left(\frac{\partial v^*}{\partial \rho} < 0\right)$. *The optimal data volume decreases with the privacy-sensitive customers' utility loss* $\left(\frac{\partial v^*}{\partial \alpha} < 0\right)$ *while it increases with the privacy-insensitive customers' utility gain* $\left(\frac{\partial v^*}{\partial \beta} > 0\right)$.

Proof: Note that $\frac{\partial v^*}{\partial \rho} = -\frac{(\alpha+\beta)(4c-(\beta(1-\rho)-\alpha\rho)^2)}{(4c-(\beta(1-\rho)-\alpha\rho)^2)^2} < 0$, $\frac{\partial v^*}{\partial \alpha} = -\frac{\rho(4c-(\beta(1-\rho)-\alpha\rho)^2)}{(4c-(\beta(1-\rho)-\alpha\rho)^2)^2} < 0$ and $\frac{\partial v^*}{\partial \beta} = \frac{(1-\rho)(4c-(\beta(1-\rho)-\alpha\rho)^2)}{(4c-(\beta(1-\rho)-\alpha\rho)^2)^2} > 0$, which completes the proof. ■

The impact of privacy sensitivity parameters on the optimal data volume is straightforward. Proposition 2 shows that the optimal data volume decreases with the number of privacy-sensitive customers and with their utility loss due to privacy concerns. On the other hand, the firm collects more data as the market has more privacy-insensitive customers and they enjoy more benefits. This result is consistent with the practitioners' view that emerging privacy concerns become significant challenge of data modelling and this is one of the reasons for failure in customer relationship management (Forbes 2013). Thus, our findings imply that firms should be well aware of the characteristics of potential customers before determining the scale of big data that they want to collect. We then investigate how these parameters influence the optimal price.

Proposition 3: *The optimal price decreases with the number of privacy-sensitive customers* $\left(\frac{\partial p^*}{\partial \rho} < 0\right)$ *and their utility loss* $\left(\frac{\partial p^*}{\partial \alpha} < 0\right)$. *On the other hand, when there exist more privacy-insensitive customers in the market and they enjoy more utility gain* $\left(\frac{\partial p^*}{\partial \beta} > 0\right)$, *the firm charges higher price.*

Proof: Note that $\frac{\partial p^*}{\partial \rho} = -\frac{4c(\alpha+\beta)(\beta(1-\rho)-\alpha\rho)}{(4c-(\beta(1-\rho)-\alpha\rho)^2)^2} < 0$, $\frac{\partial p^*}{\partial \alpha} = -\frac{4c\rho(\beta(1-\rho)-\alpha\rho)}{(4c-(\beta(1-\rho)-\alpha\rho)^2)^2} < 0$ and $\frac{\partial p^*}{\partial \beta} = \frac{4c(1-\rho)(\beta(1-\rho)-\alpha\rho)}{(4c-(\beta(1-\rho)-\alpha\rho)^2)^2} > 0$, which completes the proof. ■

Proposition 3 indicates that the optimal price responds to the privacy sensitivity parameters in the same way as the optimal data volume. Data collection and handling are often costly. Recall from Proposition 2 that when the number of privacy-sensitive customers and their utility loss increase, the firm reduces data volume. Thus, the natural reaction to the reduced data volume is cutting price. On the other hand, as the number of privacy-insensitive customers and their utility gain increase, the firm has an incentive to collect more data. Then the firm's optimal decision is setting higher price to make up for the increased cost.

4. Welfare Analysis

Our analysis indicates that privacy sensitivity of the customers matters when the firm determines the data volume and the price. Then an interesting question becomes what a welfare-maximizing data volume would be and how it would be different from the profit-maximizing level. Given the emerging concern around privacy and the social expense on big data collection and management, characterizing the welfare-maximizing data volume would provide implications to business managers and policy makers. In this section, we aim to answer the aforementioned questions through a welfare analysis.

To start, the net benefit for the two customer types, i.e., consumer surplus is as follows:

$$CS = \rho \int_{\alpha v+p}^{1} (q - \alpha v - p)dq + (1-\rho) \int_{-\beta v+p}^{1} (q + \beta v - p)dq.$$

Then, the social surplus becomes:

$$S = \pi + CS = \rho\left(\frac{1}{2} - p + \frac{p^2}{2} - v\alpha + pv\alpha + \frac{v^2\alpha^2}{2}\right) + (1-\rho)\left(\frac{1}{2} - p + \frac{p^2}{2} + v\beta - pv\beta + \frac{v^2\beta^2}{2}\right).$$

Maximizing the social surplus yields the following data volume and price:

$$v_S^* = \frac{\beta(1-\rho) - \alpha\rho}{2c - \beta^2(1-\rho) - \alpha^2\rho} \quad \text{and} \quad p_S^* = 0.$$

Recall that the profit-maximizing data volume is $v^* = \frac{\beta(1-\rho)-\alpha\rho}{4c-(\beta(1-\rho)-\alpha\rho)^2}$. A comparison between v^* and v_S^* leads to Proposition 4:

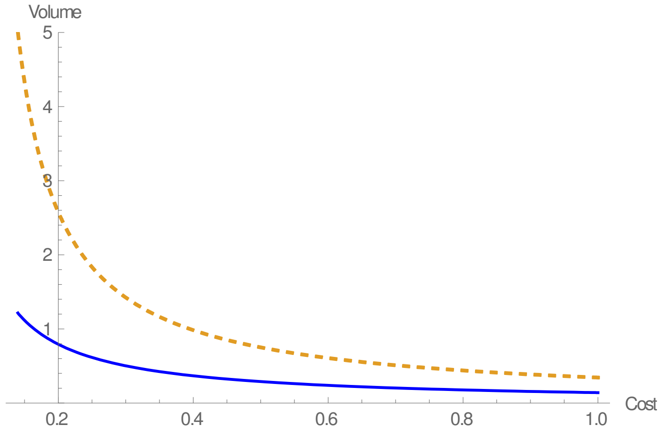
Proposition 4: *The welfare-maximizing data volume exceeds the profit-maximizing level* $(v_S^* > v^*)$.

Proof: Recall that $v^* = \frac{\beta(1-\rho)-\alpha\rho}{4c-(\beta(1-\rho)-\alpha\rho)^2}$ and $v_S^* = \frac{\beta(1-\rho)-\alpha\rho}{2c-\beta^2(1-\rho)-\alpha^2\rho}$. Since the numerators are identical, thus comparing denominators is sufficient for proof. Note that $(4c - (\beta(1-\rho) - \alpha\rho)^2) - (2c - \beta^2(1-\rho) - \alpha^2\rho) = 2c + \rho(1-\rho)(\alpha+\beta)^2 > 0$. Thus, $v_S^* - v^* > 0$. ■

Our findings from the welfare analysis provide the following implications. Even from the welfare perspective, there exists an optimal data volume, implying that collecting too much data is socially harmful. The socially optimal data volume exceeds the profit-maximizing level. Thus, the social planner has an incentive to push the firms to collect more data, which is consistent with what has been happening in reality. Figure 1 illustrates the results, i.e., the socially optimal data volume is always greater than the profit-maximizing data volume. Interestingly, the discrepancy between these two levels shrinks as the cost increases.

Figure 1. Profit maximization (solid) versus welfare-maximization (dotted)

$(\alpha = 0.3,\ \beta = 0.5,\ \rho = 0.1)$



5. Conclusion

Data volume is one of the key factors that define big data. But it does not necessarily mean that higher data volume guarantees higher utility or more profit. Firms incur more cost to collect, handle, and analyze the data and the customers bear higher privacy risks with bigger data. In this paper, we investigate the optimal information level, aiming to break the myth in the data volume of big data. Grounded on a game-theoretic model, we consider the two types of customers in terms of privacy sensitivity, and examine the impact of privacy sensitivity factor and the data management cost on a monopolistic online retailer's choice on the optimal data volume and the price.

Our result shows that there exists an equilibrium with finite optimal data volume and price. This contradicts to the common misperception of big data in that the size guarantees benefit. The findings indicate that both the number of privacy-sensitive customers and their utility loss are negatively related to the data volume and price. The welfare analysis reveals that there exists a socially optimal data volume, implying that collecting too much data is socially harmful as well. Also, we find that the welfare-maximizing data volume exceeds the profit-maximizing level. This explains the governments' incentive to encourage firms to join the big data movement, which has been observed in many countries for the last decade or so.

Despite the importance of characterizing the optimal volume of big data, the existing studies do not pay much attention to the optimal data volume. We explore cost and privacy issues that determine the optimal data volume and compare the profit- and welfare-maximizing levels. Our findings may give online retailers and policy makers with a guideline to determine the scope of data collection. Future research may consider the following venue. Firstly, it would be interesting to empirically validate our findings, i.e., the existence of optimal data volume given the cost and privacy concerns. Secondly, to add more reality, modelling the economies of scale, possibly with a network externality term and examining its impact on the data volume would be a nice direction to expand. Thirdly, it would be meaningful to propose and investigate policy alternatives that reduce the gap between profit-maximizing data volume and the social optimum. Finally, the impact of privacy sensitivity on consumer utility may be moderated by firm characteristics as we often observe that a consumer may share the data with one firm while she may not with another. It is a worthwhile venue to explore the moderating effect of firm characteristics under competition.

References

AAAS-FBI-UNICRI (2014) "National and Transnational Security Implications of Big Data in the Life Sciences" available at http://www.aaas.org/sites/default/files/AAAS-FBI-UNICRI_Big_Data_Report_111014.pdf

Ahrens, J.P., B. Hendrickson, G. Long, S. Miller, R. Ross and D. Williams (2011) "Data-Intensive Science in the US DOE: Case Studies and Future Challenges" *Computing in Science Engineering* 13(6), 14-24.

Biais, B., L.R. Glosten and C. Spatt (2005) "Market Microstructure: A Survey of Microfoundations, Empirical Results, and Policy Implications" *Journal of Financial Markets* 8, 217-264.

Boyd, D. and K. Crawford (2012) "Critical Questions for Big Data" *Information, Communication & Society* 15(5), 662-679.

Chen, Y. and I. Png (2003) "Information Goods Pricing and Copyright Enforcement" *Information Systems Research* 14(1), 107-123.

Cumbley, R. and P. Church (2013) "Is Big Data Creepy?" *Computer Law & Security Review* 29(5), 601-609.

Forbes (2012) "Morgan Stanley Takes on Big Data with Hadoop" available at http://www.forbes.com/sites/tomgroenfeldt/2012/05/30/morgan-stanley-takes-on-big-data-with-hadoop/#7e5d0a172112.

Forbes (2013) "Why Your CRM Implementation is Quietly Failing" available at http://www.forbes.com/sites/danwoods/2013/09/30/why-your-crm-implementation-is-quietly-failing/#796ff9d97e3c.

Forbes (2016) "The Most Practical Big Data Use Cases of 2016" available at http://www.forbes.com/sites/bernardmarr/2016/08/25/the-most-practical-big-data-use-cases-of-2016/#2b5dcf8d7533.

Gartner (2015) "Top 10 Strategic Technology Trends for 2015" available at http://www.gartner.com/smarterwithgartner/gartners-top-10-strategic-technology-trends-for-2015/.

George, G., M.R. Haas and A. Pentland (2014) "Big Data and Management" *Academy of Management* 57(2), 321-326.

Hazen, B.T., C.A. Boone, J.D. Ezell and L.A. Jones-Farmer (2014) "Data Quality for Data Science, Predictive Analytics, and Big Data in Supply Chain Management: An Introduction to the Problem and Suggestions for Research and Applications" *International Journal of Production Economics* 154, 72-80.

Kawakami, K. (2013) "Optimal Market Size" The University of Melbourne working paper number 1168.

Laney, D. (2001) "3d Data Management: Controlling Data Volume, Velocity and Variety" META Group Research Note 6, 70.

Lee, H., D. Lim and H. Zo. (2013) "Personal Information Overload and User Resistance in the Big Data Age" *Korea Intelligent Information System Society* 19(1), 125-139.

Lu, R., H. Zhu, X. Liu, J.K. Liu and J. Shao (2014) "Toward Efficient and Privacy-Preserving Computing in Big Data Era" *IEEE Network* 28(4), 46-50.

Madhavan, A. (2000) “Market Microstructure: A Survey” *Journal of Financial Markets* 3, 205-258.

Malhotra, N.K., S.S. Kim and J. Agarwal (2004) “Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model” *Information Systems Research* 15(4), 336-355.

McKinsey Global Institute (2011) “Big Data: The Next Frontier for Innovation, Competition, and Productivity” available at http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation.

New York Times (2012) “The Age of Big Data, February 11” available at http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html.

Shvetank, S., A. Horne and J. Capella (2012) “Good Data Won’t Guarantee Good Decisions” *Harvard Business Review* 90(4).

Spiegel, M. and A. Subrahmanyam (1992) “Informed Speculation and Hedging in a Noncompetitive Securities Market” *The Review of Financial Studies* 5(2), 307-329.

TATA Consultancy Services (2013) “The Emerging Big Returns on Big Data” *A TCS 2013 Global Trend Study*.

Zikopoulos, P. and C. Eaton (2011) *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw Hill Professional.