Big Data as Strategic Enabler - Insights from Central European Enterprises

Rainer Schmidt[1], Michael Möhring[2], Stefan Maier[2], Julia Pietsch[2], and Ralf-Christian Härting[2]

[1] Hochschule München, Lothstraße 64, 80335 München, Germany
[2] Hochschule Aalen, Beethovenstraße 1, 73430 Aalen, Germany
Rainer.Schmidt@hm.edu,
{Michael.Moehring,Stefan.Maier,
Julia.Pietsch,Ralf.Haerting}@htw-aalen.de

Abstract. Big Data increases the amount of data available for analysis by significantly increasing the volume, velocity and variety. Big Data is the coincidence of technological developments with a radical transformation of information flows. Beyond technological considerations only few, general analyses of the strategic impact of Big Data exist. Therefore, we designed a study analyzing the strategic impact factors of Big Data. The study is based on a survey from 148 responses of enterprise specialists and managers in the field of Big Data, originating from central European enterprises. Key findings are that idle data have a negative moderate effect and the type of business process has a positive effect on the perceived advantages of big data. Furthermore there is an effect of the perceived advantages of big data on the planned and current use.

Keywords: big data, strategy, business intelligence, empirical research, business information system.

1 Introduction

The continuous growth of data is a new challenge for enterprises and organizations. The volume of data surpassed 2.8 zettabytes in 2012 – a growth by a factor of 9 over five years [1] and is expected to continue its rise to a 50 times higher volume in 2020 [2]. Big Data is used for processing and analyzing large amounts of heterogeneously structured data with a high velocity. Therefore, Big Data [3] [4] has become a central topic in management [5]. The importance of Big Data for management is supported by an empirical study (worldwide online survey of over 1300 IT managers) from ZDNet which indicates that "70% will use data analytics by 2013" [6]. Big Data is regarded as a highly disruptive information technological development [7]. It has been declared a national challenge and priority by President Obama's administration [8]. Big Data is also a new frontier for business and social science [9]. Big Data is considered as positive factor for the firm performance [10]. Companies using Big

W. Abramowicz and A. Kokkinaki (Eds.): BIS 2014, LNBIP 176, pp. 50–60, 2014.

Data and Analytics outperform peers 5 percent in productivity and 6 percent in profitability [11]. Big Data helps companies to learn more about the alternatives customers look for when considering buying a product. In this way, important factors in the customers' purchasing decision can be identified. Furthermore, companies gain knowledge how customers compose their shopping baskets. In spite of these expectations associated with Big Data, there are also critical thoughts [12] [13]. E.g. the problems created by too much trust into data are discussed in [6].

Despite its huge impact on enterprises, research on the strategic impact of Big Data in detail is just beginning. An initial analysis of the transformational effect of Big Data on the IT-department is undertaken in [14], [15]. The influence of Big Data on the design of business processes is demonstrated with several examples in [4]. In [5] a Big Data adoption model is introduced. It differentiates three levels of analytics adoption: aspirational, experienced and transformed. The highest level (transformed) is achieved by enterprises using analytics not only to guide but to prescribe actions. One precondition for achieving the transformed layer is a strong ability to capture, aggregate and analyze data. However, there is no deeper analysis of the impact of Big Data on business processes. An initial classification of the impact of Big Data in different sectors is given in [16]. The benefits from the use of Big Data in different sectors are analyzed in [7] .

In summary, both a thorough analysis of Big Data and its enabling potential as well as a detailed analysis of the impact of Big Data on business processes are still missing. To close this gap, we designed a study that analyses the impact factors of the perceived advantages of Big Data, the effects of Big Data, the current and planned implementation of Big Data in different business processes and the use of Big Data for strategic initiatives. The study is based on a survey comprising 148 responses of enterprise specialists and managers in relation to Big Data, originating from central European enterprises. Having eliminated incomplete and inconsistent cases, 110 responses were used for data analysis. The online survey started in January 2013 and ended in May 2013.

Our paper proceeds as follows. First, we discuss the basic properties of Big Data. In particular, Big Data is distinguished from traditional business intelligence concepts. Furthermore, enterprise architecture patterns for Big Data in enterprises are identified. In chapter 3 we introduce the research design and create four hypotheses. In the following we present our research methods and data collection. The results of our research are given in chapter 4, we give an outlook and conclusion.

2 Big Data

Big Data is the coincidence of technological developments with a radical transformation of information flows within enterprises and organizations. Big Data is not a replacement for concepts such as Business Intelligence [17] and Business Analytics [18]. Instead it increases the set of data available for analysis by significantly increasing volume, velocity and variety of data processing [19]. Existing approaches are mostly constricted to the use of transactional data, e.g. describing the purchasing of

goods. These data are well-structured, often stored in relational databases, and follow clearly defined semantics. Big Data extends this data in several ways. First, non-transactional data, such as customer interactions logged in customer relationship management systems is integrated into analysis. Nevertheless, such data are still clearly defined in terms of semantics and structures. The next extension is the integration of web logs etc. from customer-interfacing information systems. These data are no longer well-structured but subject to frequent structural changes. However these data are machine-created which results in homogeneous semantics.

Big Data supplements existing business intelligence and business analytics approaches in three ways as shown in Fig. 1**.** Integration of Business Intelligence and Big Data [15]. The first way is to use Big Data as an extension to an existing business intelligence application. The Business Intelligence application consists of a data processing layer, an information processing layer and a presentation layer. The Big Data application provides information created from data that are too huge and/or too unstructured to be processed by the legacy business intelligence system. The second way for using Big Data is to present the information created by the Big Data application (with own data processing and information generation) using an own presentation layer. The third possibility is to support other applications such as web-shops (with standardized interfaces using data processing and information generation). E.g. Big Data provides product suggestions for increasing cross-and upselling. This approach can be used by different applications in the enterprise to generate cost savings by cutting down maintenance, support and total unit costs of transactions.

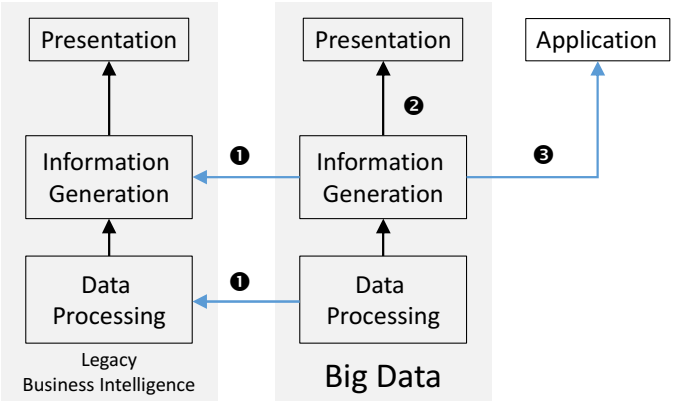

Fig. 1. Integration of Business Intelligence and Big Data [15]

3 Research Design and Method

In order to explore the potential of Big Data to enable a new performance of business processes and profitability we designed a study to identify key impact factors for implementing Big Data technologies and applications in business processes in enterprises and other organizations.

3.1 Design of the Study

The design of our study contains four hypotheses shown in Fig. 2. In companies and organizations enormous amounts of idle data exist [1]. Idle data are data available for analysis but not used today, due to restricted capacities and non-existent analytical capabilities of enterprise information technology. Therefore, the extent of existing idle data in a business process can be considered as an indicator for possible advantages by using Big Data.

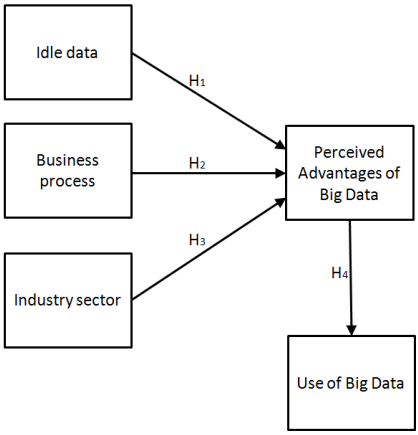

Fig. 2. Research Model

The first hypothesis of our study is, that the amount of idle data influences the perceived advantages of Big Data [19]. The perceived advantages of Big Data are defined as the capability of the use of Big Data in business processes (e.g. marketing) by data scientists and managers.

Hypothesis 1: The amount of idle data in business processes positively influences the perceived advantage of Big Data in business processes.

To study the impact of idle data, the online-based responses of the participants ranged on a scale of one to five (1: very low; 2: low; 3: moderate; 4: high; 5: very high) for business processes according to Value Chain introduced by Porter and Millar [3].

The advantages of Big Data may be different from business process to business process [3]. Literature often describe business cases of Big Data in marketing processes [4] [14]. Other core business processes (e.g. business processes of the human resources management) often ignored in publications. To investigate this topic, we created the second hypotheses, expressing that the type of business process has an influence on the perceived advantages of Big Data.

Hypothesis 2: The perceived advantages of Big Data are influenced by the respective business process.

The perceived advantages of Big Data ranged on a scale of one to five (1: very low; 2: low; 3: moderate; 4: high; 5: very high) for the respective business processes. The strategic effects of information technology vary strongly from industry sector to industry sector. Thus, we developed a hypothesis expressing that the perceived advantages of Big Data are dependent on the industry sector [20] [21].

The advantages of Big Data vary depending on business processes and industry sectors because of individual business cases. According to Gartner [22] investments in relation to Big Data vary from industry sectors. For instance retail and education enterprises more invested in Big Data as Manufacturing and Insurance enterprises [22]. To test the influence of industry sector, we created hypothesis 3:

Hypothesis 3: The industry sector exerts an effect on the perceived advantages of Big Data.

Our classification of industry sectors is based on the European Classification of Economic Activities (NACE) version 2 [23] .

Furthermore, it is interesting to understand how strong the perceived advantages of Big Data influence the current or planned use of Big Data in enterprises and organizations [24] because of the different current and planned investments related to Big Data [22]. So we defined the hypothesis 4:

Hypothesis 4: The perceived advantages of Big Data exert a positive effect on the current or planned use of Big Data.

The use of Big Data ranged from 1 to 4 as the following: 1: no plan of use; 2: use planned later than 2013; 3: use planned in 2013; 4: in use. The idle data, business process and the industry sector are important impact factors of the perceived advantages of Big Data and based on the current and planned implementation of Big Data in each business processes.

3.2 Research Methods and Data Collection

Regression analysis [25] [26] [27] is characterized as a family of methods for exploring and establishing a functional relationships between independent and dependent variables. The application of regression models is widespread use and applicable to many subject areas. Regression analysis can follow two basic approaches – simple regression and multiple regression. Generally, the regression analysis examines the relationship between a quantitative dependent variable and one or more quantitative independent variables (explanatory or predictor variables) [28].

To investigate the perceived and current or planned use of Big Data a local online-based survey in the German language (e.g. for enterprises of the countries Germany, Austria and Switzerland) were implemented. First a pre-test to reduce problems was executed, and findings were considered. The online survey started in January 2013

and ended in May 2013 (during five months). Hundreds of enterprise specialists in this area were invited to participate in this study. The final sample consists of n = 148 responses of data scientists and managers in relation to Big Data. After elimination of incomplete and inconsistent cases 110 responses were used for data analysis. To ensure that the participants were qualified to respond to this survey, we asked only specialists or managers from business or IT-departments with relations to Big Data and used validation questions (e.g. questions of their current job position within the enterprise).

The respondents cover a wide variety of industry sectors (based on European Classification of Economic Activities (NACE) version 2 [23] (Table 1). The majority of participating enterprises operate in the fields of transportation and storage, information and communications as well as manufacturing.

Table 1. Industry sectors of the participants enterprises

Industry sector	**Percent**
Transportation and storage, information and communication	37.27%
Manufacturing	20.91%
Activities of households as employers, undifferentiated goods- and services-producing activities of households for own use	10.00%
Real estate activities, professional, scientific and technical activities, administrative and support service activities	8.18%
Wholesale and retail trade, repair of motor vehicles and motorcycles	3.64%
Financial and insurance activities	3.64%
Education	3.64%
Arts, entertainment and recreation, other service activities	3.64%
Human health and social work activities	2.73%
Electricity, gas, steam and air conditioning supply, water supply, sewerage, waste management and remediation activities	1.82%
Construction	1.82%
Accommodation and food service activities	0.91%
Public administration and defence, compulsory social security	0.91%
Activities of extraterritorial organisations and bodies	0.91%
Agriculture, forestry and fishing	0.00%
Mining and quarrying	0.00%

Most respondents work for companies with more than 250 employees (30 percent work for companies with 500 or more employees).

4 Results

We used IBM Statistics SPSS 20 to test our hypotheses (Figure 2) employing linear regression models (H1, H3, H4) or descriptive analytics (H2). Hypotheses 1, 3 and 4 were verified by using bivariate linear regression modeling. The usage of this analysis consists of a dependent variable (e.g. perceived advantage of Big Data in each

business unit) and an independent variable (e.g. idle data in each business unit) in the same business process (e.g. marketing). The variable-based impact was analyzed by non-standardized and standardized (beta) coefficient of regression as well as the significance (p-value).

To discover how the idle data impacts the perceived advantages of Big Data in each business process a linear regression with the "perceived advantages of Big Data" as the dependent variable and the "idle data" as the independent variable in each business process is applied (Table 2).

Table 2. Linear regression for hypothesis 1

Variable	Regr. coeff.	Beta	Sign.
Inbound logistics	0.268	-0.462	0.000
Outbound logistics	-1.388	-0.361	0.000
Operations	-1.486	-0.438	0.000
Marketing	-1.299	-0.397	0.000
Management	-1.125	-0.329	0.002
Financial Service	-1.026	-0.265	0.015
Technology development / IT	-1.093	-0.346	0.001
Human Resource Management	-0.939	-0.333	0.002

The comparison of the beta values (effect of the independent variable on the dependent variable) of each business process shows a moderate negative and significant effect of the availability of idle data to the perceived advantage of Big Data in each business process (from -0.265 to – 0.462). Therefore, hypothesis 1 (The amount of idle data in business processes positively influences the perceived advantage of Big Data in business processes.) must be rejected.

Table 3. Descriptive Analytics for Hypothesis 2

Variable	Average
Inbound logistics	3.018
Outbound logistics	2.736
Operations	3.109
Marketing	3.409
Management	3.200
Financial Service	2.727
Technology development / IT	3.572
Human Resource Management	2.618

Furthermore, the study shows differences in the perceived advantage of Big Data for every business process. Table 3 shows the averages for the respective business processes. The perceived use of Big Data related to each business process is rated on a five-grade scale (1: very low; 2: low; 3: moderate; 4: high; 5: very high):

Technology development / IT (average: 3.572) and Marketing (average: 3.409) are the core processes with the highest perceived use of Big Data (see Table 3).

Especially in the marketing environment applications of Big Data enable new knowledge transfer to the management. That helps to get a better understanding of important marketing issues (e.g. forecast the success of a marketing campaign).

To answer these questions plenty of applications are available. Based on a Big Data fundament tools like decision trees, association, time series, clustering or neural nets are often used.

In contrast to the Marketing the processes Human Resource Management (average 2.618) and Outbound Logistics have the lowest use perceived. All core business processes have a noticeable significant influence of the perceived use of Big Data (highly two-side significant (0,000) of 1 and 2 (very low and low perceived advantages of Big Data)). Therefore, hypothesis 2 (The perceived advantages of Big Data are influenced by the respective business process) is supported by the survey data. Hypothesis 3 tests the influence of industry sector. The data collection for hypothesis 3 was focused on the usage of Big Data in different business sectors. Therefore, the identification of 8 business sectors was made.

Table 4. Linear regression for hypothesis 3

Variable	**Regr. coeff.**	**Beta**	**Sign.**
Inbound Logistics	0.209	0.099	0.397
Outbound Logistics	0.423	0.182	0.118
Operations	0.894	0.353	0.002
Marketing	-0.207	-0.084	0.472
Management	0.135	0.058	0.622
Financial Service	0.097	0.059	0.613
Technology development / IT	-0.256	-0.138	0.239
Human Resource Management	-0.010	-0.009	0.939

After analyzing the received data, it was obvious that - based on the measure of determination of the significance - the results could not be interpreted well. Consequently a reduction and selection of the three dominating business sectors within the sample was made to improve the chance for significant results:

- Transportation and storage, information and communication (n=41)
- Manufacturing sector (n=23)
- Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use (n=11)

After that, the significance levels improved but remained mostly unsatisfactory. The only significant business process within this model is operations. That means that there exists a relation between the investigated business sector and the perceived advantage of the business process. Nevertheless if the attention is paid to the coefficient of regression and the Beta coefficient the data can be interpreted as follows: The coefficient of regression is between -0.256 and 0.894 what results in an arithmetic average of 0.160625. The Beta coefficient contains an averaged value of 0.065 (the single data range between -0.138 and 0.353). That means that in comparison of both standardized

and non-standardized values a significant impact is not visible. Ergo hypothesis 3 (The industry sector exerts an effect on the perceived advantages of Big Data.) cannot be confirmed - that is to say that the business sector has no impact on the perceived advantages of Big Data. But such a conclusion would be based on non-significant values.

The effect of the perceived advantage of Big Data on the planned and current use shows the following results based on a linear regression model (dependent variable: the planned and current use of Big Data; independent variable: perceived advantages of Big Data):

Table 5. Linear regression for hypothesis 4

Variable	Coeff. of regr.	Beta	Sign.
Inbound logistics	0.243	0.332	0.004
Outbound logistics	0.181	0.330	0.004
Operations	0.179	0.299	0.009
Marketing	0.402	0.537	0.000
Management	0.251	0.385	0.001
Financial Service	0.327	0.383	0.001
Technology development / IT	0.387	0.494	0.000
Human Resource Management	0.493	0.310	0.007

On average, there is a low moderate effect (average of beta value: 0.383) on the current and planned use of Big Data based on the perceived advantages of Big Data. The maximum value of beta is 0.537 (Marketing), the minimum value is 0.299 (Operations). Therefore, hypothesis 4 (The perceived advantages of Big Data exert a positive effect on the current or planned use of Big Data.) can be confirmed.

5 Conclusion

Research on the strategic impact of Big Data in detail is just taking off. Our research shows that Big Data has a high perceived advantage in many business processes and especially in Technology development and IT as well as in Marketing. Thus, enterprises can define their investments in Big Data technology related to these primary processes and their business needs. Especially small and medium sized enterprises are able to focus their Big Data initiatives on investments providing the highest business impact and can so reduce financial risks.

Furthermore, we found that the business sector shows no significant influence on the perceived advantages of Big Data. Future research based on larger samples may yield different findings. The idle data in each business process has a negative effect on perceived use of Big Data and the perceived advantages of Big Data influence its current and planned usage.

This research has both practical and theoretical implications. Managers can benefit from information about the use of Big Data in different business processes or units, especially in Marketing and Technology development / IT. Academic research can

benefit of new knowledge about differences in the use of data analysis in different core processes and sectors of central European enterprises. Consequently methods and theories of data analysis and Big Data / Business Intelligence can adopt these results in order to create or improve current approaches (e.g. for implementing management information systems).

Generally the study results are significant. But there are exceptional cases. So there is a need for future research and a comparison in few years later to document the effect of the use of Big Data after several implementation projects. In our study we focused on the most abstract business processes such as operations, marketing, etc.. In the future there is a need to explore business processes in detail (e.g. process of customer relationship management or sales) and to evaluate the best implementation points for Big Data. That helps to improve business processes (e.g. reduce process time and process costs, improve process quality) in detail. In addition a framework for implementation of Big Data can be built. Not in all business cases better decisions based on a larger data base can be implemented. Therefore a proof of concept for each business case of Big Data is needed. Future research will enlarge the number of countries involved. For example enterprises from North America and BRIC states are very interesting, because of their economic growth and standing. A comparison of the Big Data adoption in these countries with the one in central Europe, will be an interesting theme for research.

References

1. Gantz, J., Reinsel, D.: Extracting value from chaos. IDC IView, 1–12 (2011)
2. Mearian, L.: World's data will grow by 50X in next decade, IDC study predicts, http://www.computerworld.com/s/article/9217988/World_s_data_will_grow_by_50X_in_next_decade_IDC_study_predicts
3. Porter, M.E., Millar, V.E.: How information gives you competitive advantage. Harv. Bus. Rev. 63, 149–160 (1985)
4. McAfee, A., Brynjolfsson, E.: Big data: the management revolution. Harv. Bus. Rev. 90, 60 (2012)
5. LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., Kruschwitz, N.: Big data, analytics and the path from insights to value. MIT Sloan Manag. Rev. 52, 21–32 (2011)
6. Bursting the Big Data Bubble | ZDNet, http://www.zdnet.com/bursting-the-big-data-bubble-7000002352/
7. Bughin, J., Chui, M., Manyika, J.: Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. McKinsey Q. 56 (2010)
8. Weiss, R.: OBAMA ADMINISTRATION UNVEILS "BIG DATA" INITIATIVE
9. Kaisler, S., Armour, F., Espinosa, J.A., Money, W.: Big Data: Issues and Challenges Moving Forward. In: 2013 46th Hawaii International Conference on System Sciences, pp. 995–1004. IEEE Computer Society, Los Alamitos (2013)
10. Brynjolfsson, E., Hitt, L., Kim, H.: Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? (2011)
11. Breuer, P., Forina, L., Moulton, J.: Beyond the hype: Capturing value from Big Data and advanced analytics, http://cmsoforum.mckinsey.com/article/beyond-the-hype-capturing-value-from-big-data-and-advanced-analytics

12. Gray, J.: What data can and cannot do, http://www.guardian.co.uk/news/datablog/2012/may/31/data-journalism-focused-critical
13. The Hidden Biases in Big Data - Kate Crawford - Harvard Business Review, http://blogs.hbr.org/cs/2013/04/the_hidden_biases_in_big_data.html
14. Möhring, M., Schmidt, R., Wolfrum, N., Kammerer, M., Maier, S., Höritz, S.: How Big Data Transforms the IT Department to a Strategic Weapon. In: Proceedings of the IADIS International Conference Information Systems 2013, Lisbon, pp. 323–326 (2013)
15. Schmidt, R., Möhring, M.: Strategic alignment of Cloud-based Architectures for Big Data. In: Proceedings of the 17th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOCW), Vancouver, Canada, pp. 136–143 (2013)
16. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data: The next frontier for innovation, competition, and productivity. McKinsey Glob. Inst., 1–137 (2011)
17. Kemper, H.-G., Baars, H., Lasi, H.: An Integrated Business Intelligence Framework. In: Rausch, P., Sheta, A.F., Ayesh, A. (eds.) Business Intelligence and Performance Management, pp. 13–26. Springer, London (2013)
18. Barton, D.: Making Advanced Analytics Work For You. Harv. Bus. Rev. 90, 78–83 (2012)
19. Chen, H., Chiang, R.H.L., Storey, V.C.: Business intelligence and analytics: From big data to big impact. MIS Q. 36, 1165–1188 (2012)
20. Brown, B., Chui, M., Manyika, J.: Are you ready for the era of "big data"? McKinsey Q. 4, 24–35 (2011)
21. Hannula, M., Pirttimäki, V.: Business intelligence empirical study on the top 50 Finnish companies. J. Am. Acad. Bus. 2, 593–599 (2003)
22. Kart, L., Heudecker, N., Buytendijk, F.: Survey analysis: big data adoption in 2013 Shows Substance behind the hype (2013)
23. Commission, E.: NACE Rev. 2–Statistical classification of economic activities in the European Community. Luxemb. Off. Off. Publ. Eur. Communities (2008)
24. Mayer-Schönberger, V., Cukier, K.: Big data: a revolution that will transform how we live, work and think. John Murray, London (2013)
25. Stigler, S.M.: The Story of Statistics: The Measurement of Uncertainty Before 1900. Harvard University Press (1986)
26. Boscovich, R.J.: De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa. Bononiensi Sci. Artum Inst. Atque Acad. Comment. 4, 353–396 (1757)
27. Boscovich, R.J.: De recentissimis graduum dimensionibus, et figura, ac magnitudine terrae inde derivanda. Philos. Recentioris Benedicto Stay Romano Arch. Publico Eloquentare Profr. Versibus Traditae Libri X Adnot. Suppl. P Rogerii Joseph Boscovich SJ Tomus II, 406–426 (1757)
28. Mason, C.H., Perreault Jr., W.D.: Collinearity, power, and interpretation of multiple regression analysis. J. Mark. Res., 268–280 (1991)