Quantitative Analyses in Digital Marketing and Business Intelligence

Inaugural-Dissertation

zur Erlangung

der wirtschaftswissenschaftlichen Doktorwürde

des Fachbereichs Wirtschaftswissenschaften

der Philipps-Universität Marburg

eingereicht von

Patrick Winter, B. Sc.

aus Neustadt an der Waldnaab

Erstgutachter:	Prof. Dr. Paul Alpar
Zweitgutachter:	Prof. Dr. Ulrich Hasenkamp
Einreichungstermin:	02. August 2016
Prüfungstermin:	07. September 2016
Erscheinungsort:	Marburg
Hochschulkennziffer:	1180

Quantitative Analyses in Digital Marketing and Business Intelligence

Patrick Winter

Meinen Eltern

Abstract

This work is divided into two parts. The first part consists of four essays on questions in digital marketing; this term refers to all marketing activities on the Internet, regardless of whether they primarily address users of stationary devices (e.g., a desktop PC) or users of mobile devices (e.g., a smartphone).

In Essay I, we model the time it takes until an item that is offered in the popular buy-it-now offer format is sold. Our model allows drawing inference from the observation of this time on how many consumers are interested in the item and on how much they value it. By this approach, several problems can be bypassed that often arise when these factors are estimated from data on items that are offered in an auction. We demonstrate the application of our model by an example.

Essay II investigates which effects ads that are displayed on search engine results pages have on the click behavior and the purchase behavior of users. For this purpose, a model and a corresponding decision rule are developed and applied to a dataset that we have obtained in a field experiment. The results show that search engine advertising can be beneficial even for search queries for which the website of the advertising firm already ranks high among the regular, so-called organic search results, and even for users who already search with one of the firm's brand names.

In Essay III, we argue theoretically and show empirically that online product ratings by customers do not represent the rated product's quality, as it has been assumed in previous studies, but rather the customers' satisfaction with the product. Customer satisfaction does not only depend on product quality as observed after the purchase but also on the expectations the customers had of the product before the purchase.

Essay IV investigates the relationship between the offline and the mobile content delivery channel. For this purpose, we study whether a publisher can retain existing subscribers to a print medium longer if he offers a mobile app through which a digital version of the print medium can be accessed. The application of our model to the case of a respected German daily newspaper confirms the existence of such an effect, which indicates a complementary relationship between the two content delivery channels. We analyze how this relationship affects the value of a customer to the publisher.

The second part of this work consists of three essays that explore various approaches for simplifying the use of business intelligence (BI) systems. The necessity of such a simplification is emphasized by the fact that BI systems are nowadays employed for the analysis of more and more heterogeneous data than in the past, especially transactional data. This has also extended their audience, which now also includes inexperienced knowledge workers.

Essay V analyzes by an experiment that we have conducted among knowledge workers from different firms how the presentation of data in a BI system affects how fast and how accurate the system users answer typical tasks. With regard to this, we compare the three currently most common data models: the multidimensional one, the relational one, and the flat one. The results show that it depends on the type of the task considered which of these data models supports users best.

In Essay VI, a framework for the integration of an archiving component into a BI system is developed. Such a component can identify and automatically archive reports that have become irrelevant. This is in order to reduce the system users' effort associated with searching for relevant reports. We show by a simulation study that the proposed approach of estimating the reports' future relevance from the log files of the BI system's search component (and other data) is suitable for this purpose.

In Essay VII, we develop a reference algorithm for searching documents in a firm context (such as reports in a BI system). Our algorithm combines aspects of several search paradigms and can easily be adapted by firms to their specificities. We evaluate an instance of our algorithm by an experiment; the results show that it outperforms traditional algorithms with regard to several measures.

The work begins with a synopsis that gives further details on the essays.

Inhalt (German Abstract)

Diese Arbeit ist in zwei Teile untergliedert. Ihr erster Teil besteht aus vier Essays zu verschiedenen Fragestellungen des digitalen Marketings; unter diesem Begriff werden alle Marketingaktivitäten im Internet verstanden, unabhängig davon, ob damit primär Nutzer stationärer Endgeräte (wie eines Desktop-PCs) oder Nutzer mobiler Endgeräte (wie eines Smartphones) adressiert werden.

In Essay I wird ein Modell entwickelt und beispielhaft angewandt, mit dem aus der Beobachtung, wie lange es dauert, bis ein im gängigen Sofort-Kaufen-Format angebotener Artikel verkauft wird, darauf zurückgeschlossen werden kann, wie viele Konsumenten sich für ihn interessieren und wie sehr sie ihn wertschätzen. Dadurch können einige Probleme umgangen werden, die häufig auftreten, wenn die genannten Faktoren anhand von Daten über versteigerte Artikel geschätzt werden.

Essay II untersucht, wie sich auf Suchmaschinenergebnisseiten platzierte Werbeanzeigen auf das Klick- und Kaufverhalten der Nutzer auswirken. Dazu werden ein Modell und eine zugehörige Entscheidungsregel entwickelt und auf einen Datensatz aus einem eigens durchgeführten Feldexperiment angewandt. Die Ergebnisse zeigen, dass sich Suchmaschinenwerbung selbst für Suchbegriffe lohnen kann, für die die Webseite des Werbetreibenden bereits unter den regulären, sogenannten organischen Suchergebnissen gut positioniert ist, und sogar für Nutzer, die bereits Markenbegriffe des Werbetreibenden verwenden.

In Essay III wird theoretisch argumentiert und empirisch gezeigt, dass von Kunden im Internet verfasste Produktbewertungen weniger ein Ausdruck der Qualität des bewerteten Produktes sind, wie in früheren Studien angenommen wurde, sondern eher ein Ausdruck der Zufriedenheit der Kunden mit dem Produkt. Diese hängt nicht nur von der Qualität des Produktes ab, wie sie nach dem Kauf beobachtet werden kann, sondern auch von den Erwartungen, die die Kunden bereits vor dem Kauf über das Produkt hatten.

Essay IV betrachtet die Beziehung zwischen dem traditionellen und dem mobilen Distributionskanal. Dazu wird untersucht, ob ein Verleger bestehende Abonnenten eines Printmediums länger an sich binden kann, indem er eine App anbietet, die den mobilen Zugang zu einer digitalen Version des Printmediums ermöglicht. Die Anwendung des dafür entwickelten Modells für eine große deutsche Tageszeitung bestätigt einen solchen Zusammenhang, was auf eine komplementäre Beziehung der beiden Distributionskanäle hinweist. Es wird analysiert, wie sich diese auf den Wert eines Kunden für den Verleger auswirkt.

Der zweite Teil dieser Arbeit besteht aus drei Essays, die verschiedene Möglichkeiten thematisieren, wie die Arbeit mit Business Intelligence (BI)-Systemen vereinfacht werden kann. Die Notwendigkeit einer solchen Vereinfachung wird dadurch betont, dass in BI-Systemen heute mehr und verschiedenartigere Daten analysiert werden als früher, insbesondere Transaktionsdaten. Dadurch hat sich auch ihr Nutzerkreis vergrößert, zu dem nun auch unerfahrene Wissensarbeiter zählen.

Essay V untersucht mittels eines Experiments unter Wissensarbeitern verschiedener Firmen, wie sich die Darstellung von Daten in einem BI-System darauf auswirkt, wie schnell und wie akkurat dessen Nutzer typische Aufgaben beantworten. Dabei werden die drei derzeit gängigsten Datenmodelle verglichen: das multidimensionale, das relationale und das flache. Die Ergebnisse zeigen, dass es von der Art der Aufgabe abhängt, welches dieser Datenmodelle die Nutzer am besten unterstützt.

In Essay VI wird ein Framework zur Integration einer Archivierungskomponente in ein BI-System vorgestellt, die irrelevant gewordene Berichte automatisch erkennen und archivieren kann, um dadurch den mit der Suche nach relevanten Berichten verbundenen Aufwand der Systemnutzer zu reduzieren. In einer Simulationsstudie wird gezeigt, dass der entwickelte Ansatz, die zukünftige Relevanz von Berichten u.a. anhand der Log-Dateien der Suchkomponente des BI-Systems zu schätzen, dafür geeignet ist.

In Essay VII wird ein Referenz-Algorithmus zur Suche nach Dokumenten in einem Unternehmen (wie z.B. Berichten in einem BI-System) entworfen, der Aspekte verschiedener Suchparadigmen kombiniert und von Unternehmen leicht an ihre Spezifika angepasst werden kann. Mittels eines Experiments wird eine konkrete Instanz dieses Algorithmus untersucht; die Ergebnisse zeigen, dass er traditionelle Algorithmen hinsichtlich mehrerer Kennzahlen übertrifft.

Die Arbeit beginnt mit einer Synopsis, aus der weitere Details zu den einzelnen Essays hervorgehen.

Synopsis

1. Introduction and Contents

This work is divided into two parts. The first part contains four essays on topics in digital marketing, while the second part contains three essays on topics in business intelligence (see Table 1). In the following, we will introduce these disciplines, their interrelationship, and our essays.

Part	Essay		Title	Page
	#	**Label**		
Digital marketing	**I**	Buy-It-Now	Quantifying Consumer Interest and Consumer Valuation from Buy-It-Now Offers	15
	II	Search Ads	Effects of Search Engine Advertising and the Case of Brand Bidding	28
	III	Ratings	Understanding Online Product Ratings: A Customer Satisfaction Model	67
	IV	Mobile Apps	Can Mobile Apps Defend Print Media?	83
Business intelligence	**V**	Data Models	Should the World Look Flat for Business Intelligence Users?	102
	VI	Archiving	On the Relevance of Reports – Integrating an Automated Archiving Component into a Business Intelligence System	123
	VII	Search Algo.	Finding Information Faster by Tracing My Colleagues' Trails – A Reference Algorithm for Enterprise Search	141

Table 1. Contents of This Work.

1.1 Digital Marketing

When a firm intends to market a product, it faces a multitude of decisions. These decisions can be classified according to various criteria; the arguably most famous (if somewhat simplistic) approach is the "4P" marketing mix model of McCarthy (1960), which distinguishes between four categories: product, price, promotion, and place. Under "product", all decisions are subsumed that relate to the product to be marketed, such as which features it should provide or how it should be designed. "Price" refers to all decisions that affect how much customers effectively have to pay for the product, such as ones on its list price or on incentive sales promotions (e.g., temporary price cuts or coupons). Decisions on all other (i.e., non-incentive) sales promotions, especially on advertising, are subsumed under "promotion". Finally, "place" refers to decisions on the (physical or virtual) distribution of the product.

Obviously, all these decisions cannot be made effectively without consideration of consumers and their demand for certain products, product characteristics, possibilities of consumption, and so on. To emphasize this, McCarthy has put the consumer in the center of his marketing mix model; in his words: "*[M]arketing planners develop a product to satisfy […] consumer[s]. They seek to make [a] product available in the right channels at the right time. They advise consumers of its availability through advertising and sales effort, and seek to set a price in tune with consumer demand […]*" (p. 45). The consumer-centered view on marketing is widely accepted in research and practice (see Bell et al. 2002 for some examples), so that it is agreed that controlling consumer demand is a crucial task for firms.

In the last decades, the rise of the Internet and, more recently, the spread of mobile devices that provide access to it (such as smartphones) have fundamentally changed the relationship between firms and consumers. This, of course, has created several new opportunities for marketing, which make the difference between digital (i.e., online or mobile) marketing and traditional (i.e., offline) marketing. The first four essays of this work each address one of these opportunities, as is described in the following.

A first new opportunity, addressed in *Essay I*, is that consumer demand can be quantified from consumer purchase behavior that is observable on the Internet, specifically on online auction platforms such as eBay. This is because the bids that consumers may place in auctions contain information on their demand that seems not to be accessible otherwise. In addition, firms can monitor auctions not only for their own but also for competing products, allowing competitive insights, and not only for new but also for used items, providing insights into the secondary market. Researchers have recognized this potential and have developed models to quantify consumer demand from the ending prices or bid histories of auctions (e.g., Bajari & Hortaçsu 2003). However, there is also a major drawback to this: Consumer behavior (i.e., bidding) and seller behavior in auctions are often strategic and may even be fraudulent (see Trevathan & Read 2006 for an overview), so that many models are valid only under restrictive assumptions. In Essay I, we argue that these problems can be bypassed by considering buy-it-now offers instead of auctions. In this format, which is available and popular on many platforms, consumers decide only on whether to buy an item, so that there is not much room for strategic or fraudulent behavior. We show that by analyzing how long it takes until an item is sold, it is still possible to reconstruct the same two aspects of consumer demand as from auctions: how many consumers are interested in the item and how much they value it.

When a firm is not satisfied with the demand for its products, it can try to create additional demand by advertising. In digital marketing, there are several new opportunities to do so (see Wordstream 2016 for an overview). In *Essay II*, we investigate the form of digital advertising that is currently most popular (in terms of revenues, according to PwC 2015): search engine advertising (SEA). SEA means placing ads that link to the firm's website on the results pages of search engines such as Google. Usually, a firm does not have to pay for the mere display of an ad but only if a search engine user clicks on it. This leads to the following problem: While the main goal pursued with SEA is to attract users who would not visit the firm's website without the ad, it may cause some users who would do so via the regular (organic) search result to instead click on the ad. In this case, the firm has to pay for clicks that it could have had for free. In addition, SEA may also affect the users' purchase behavior on the firm's website. In Essay II, we show how these effects can be quantified from the data provided by search engines to their customers and how they can be weighed against each other in order to support firms in selecting search queries for ad display. A special focus is given to search queries that contain one of the firm's brand names, as placing ads for them (termed brand bidding) is especially controversial (see the essay for the reasons).

If a firm, potentially after having run an advertising campaign, is satisfied with the demand for its products, it needs to ensure that this demand remains stable. For this purpose, it is beneficial to the firm if it can monitor the experiences that its customers make with its products. This is nowadays possible more easily and more continuously than in the past (when, e.g., surveys had to be used to collect customer feedback) since consumers often voluntarily report their experiences with products in form of reviews and ratings on dedicated websites or on the websites of online stores such as Amazon. Exploiting these data is the next new opportunity in digital marketing, which we address in *Essay III* from a theoretical perspective. While previous research has largely interpreted online product ratings as a representation of product quality (e.g., Hu et al. 2006, Koh et al. 2010), we argue and show that they rather represent customer satisfaction. Customer satisfaction depends on product performance as reflected by product quality, but also on the customers' expectations of the product in the pre-purchase phase (Westbrook & Reiley 1983, Fornell 1992). That is, our interpretation of online product ratings adds the perspective of pre-purchase expectations to the perspective of post-purchase quality assessments.

When a firm notices the threat of a decline in demand (e.g., because it learns from online product reviews and ratings that its customers are not satisfied, or at least less satisfied than the customers of its competitors) or when this decline has already occurred, the firm needs to reinforce demand. How this can be accomplished depends on the reasons for the (threatening) decline. If customers are not satisfied with how the firm sells or delivers its products, it can consider serving additional channels for these purposes, depending on the nature of its products. New opportunities for marketing here arise from the availability of the digital channels (online and mobile). In *Essay IV*, we study the relationship (complementary, zero, or substitutive) between the offline channel and the mobile channel. More concretely, we investigate whether a firm that publishes a print medium (in our case a newspaper) can retain existing subscribers to it longer (i.e., reinforce their demand) by additionally offering a mobile app for consuming a digital version of the content. We also analyze how this affects the value a customer has to the firm.

In summary, our essays on digital marketing can be classified according to the "4P" model as follows. Essay I belongs to "product" and "price" because it shows how consumer demand for certain items can be quantified in terms of both how many consumers are interested in them and how much these consumers value them. Essay II investigates the effects of SEA as a form of digital advertising and belongs, therefore, to "promotion". In Essay III, it is argued that online product ratings reflect customer satisfaction, which is determined by customer expectation and product performance. Therefore, it clearly relates to "product". It is also related to "price" because the product price is a determinant of customer expectation, as it is shown in the essay. Finally, Essay IV belongs to "place", as it investigates the effects of offering an additional delivery channel. One could also speak of Essays I–IV as presenting approaches to quantify, create, ensure, and reinforce consumer demand, respectively.

1.2 Business Intelligence

As we have elaborated on in the previous section, a firm cannot make effective marketing decisions without consideration of consumers and their demand. Of course, this is true not only with regard to marketing but also with regard to several other business functions (e.g., production). For decision support, many firms store aggregated data on their transactions with consumers, which here means their customers or potential customers, in dedicated information systems and integrate them with data from other sources (such as official statistics, industry studies, consumer panels, and so on). The stored data can then be used for monitoring key performance indicators, creating reports on or finding patterns in past consumer behavior, predicting future consumer behavior, and other analytical purposes. This approach is referred to as business intelligence (BI; see Chen et al. 2012 for various views on BI).

The changes in the firm–consumer relationship due to the rise of the Internet and the spread of mobile devices have not only greatly affected marketing as explained above; they also have had effects on BI, which are summarized by Chen et al. (2012). In fact, strong relationships between (digital) marketing and BI have become obvious in the last decade (e.g., Kar et al. 2010, Kurniawan et al. 2014). This can be seen from several examples that relate to the first essays of this work. Most importantly, new sources for data on the opinions, attitudes, needs, etc. of consumers are available due to the Internet. Internet users often generate such data by themselves, either explicitly (e.g., when they rate products online, Essay III) or implicitly (e.g., when they buy items on online auctions platforms, Essay I). This has led Chen et al. to the view that the integration of such data into BI systems has constituted a new era of BI, called BI 2.0[1]. They expect that this era will soon evolve into another one (BI 3.0), in which sensor-based and mobile content (Essay IV) will be in the focus of BI. Another relationship between digital marketing and BI is that many digital marketing platforms provide BI-like functionalities. For example, the SEA (Essay II) platform of Google, AdWords, offers a dashboard view, user-specified reports, and predictions on ad performance, in addition to several other typical functionalities such as an automatic aggregation and historization of data.

When Chen et al. speak of different eras of BI, they understand this in terms of the increased variety of data that are analyzed with BI systems. However, not only the variety of data has increased in the last decade, but also their volume. This has partly resulted from the integration of new data sources as well, but even more from a new BI practice: Firms, enabled by technological progress having greatly sped up the processing of data, nowadays conduct analyses at the level of transactional data (e.g., individual sales) instead of only at the level of aggregated data (e.g., daily sales figures) in order to support operational decisions (operational BI; White 2005) in a similar way as strategic decisions. For this purpose, however, huge amounts of transactional data have to be stored in BI systems.

[1] This is an allusion to the term “Web 2.0”, which emphasizes the “new” interactivity between users and websites that is in contrast to the passive role of users as mere consumers of content in the “traditional” Internet.

Besides, operational decisions are usually made by the lower management and operative staff, so that this large group has joined (and thereby greatly extended) the audience of BI. It has been recognized that the traditional approach of providing decision makers with analyses by the IT department is too slow to be effective under these circumstances (Imhoff & White 2011), especially as operational decisions are often particularly time-critical (e.g., Işik et al. 2013). This has motivated many firms to let their knowledge workers perform BI by themselves in order to accelerate the process and to increase its effectiveness (self-service BI; Imhoff & White 2011, Alpar & Schulz 2016). However, this has also greatly increased the number of BI system users and the share of them that have little to no experience with BI.

In summary, nowadays more and more heterogeneous users than in the past want to analyze more and more heterogeneous data with BI systems. This calls first for new approaches to support them in the process of data retrieval (which here means report generation). As users can be expected to perform the better in their tasks the better they have comprehended the available data, such an approach can be to present them the data in the structure that suits them best. In *Essay V*, we investigate which structure this is, considering three data models as alternatives: the multidimensional one, which has traditionally been used in BI systems, the relational one, which is common in operational systems, and the flat one, which is very simple and may, therefore, be particularly suited for presenting data to inexperienced users. We frame our analysis theoretically by research on the cognitive processing of structures.

The reports users create are usually not deleted, which under the abovementioned circumstances can quickly lead to an explosion of their total number. This is especially true when reports are created in such a way that they can hardly be reused, which is often the case for inexperienced users. An extreme example of this is reported by Eckerson (2008, p. 21): In a BI system with 450 users, 26,000 reports were available after some years of self-service BI, while only 300 (1.15%) would have sufficed to fulfill nearly the same information needs. Under such circumstances, users may not be able to find already existing reports that are relevant to them quickly, so that they have to spend time for an extensive search or for recreating such reports. This causes search costs and opportunity costs to the firm (e.g., Haas & Hansen 2007), which for a firm with 100 employees quickly can reach several million dollars annually (Schubmehl & Vesset 2014).

To counteract this, we propose in *Essay VI* the integration of an automated archiving component into the BI system. The key idea here is that the future relevance of each report can be estimated based on the log files of the system's search component and other indicators. Reports that are likely to be irrelevant can then be sorted out. In *Essay VII*, we develop a reference search algorithm for finding documents in a firm context (such as BI reports) faster by exploiting data on previous search sessions, taking into account the similarity of the respective users. This is an information retrieval-based approach to the abovementioned danger, which is complementary to the information storage-based approach presented in Essay VI (see there for details). Our algorithm can easily be adapted by firms to their specificities through concretization.

2. Data and Methods

Different (types of) data sources and different methods are used in the essays of this work. In the following, we will explain which data our models and approaches require in general and from which sources we obtained such data for their applications (left part of Table 2). We will also very briefly introduce the (data analysis) methods on which our models and approaches are based (right part of Table 2).

Essay		**Data source(s)**				**Method(s)**				**Esti-mation**
#	**Label**	**Internet (crawled)**	**Firm**	**Experi-ment**	**Simulation**	**GLM regression**	**Survival analysis**	**KPIs & tests**	**Other**	
I	Buy-It-Now	X			(X)	X	X		Auction theory	ML
II	Search Ads	(X)	X	X	(X)	X		(X)	HBM	MCMC
III	Ratings	X							SEM	PLS
IV	Mobile Apps		X		(X)	X	X		CLV analysis	ML
V	Data Models			X	(X)	X	X			ML
VI	Archiving				X			X		-
VII	Search Algo.			X	X			X		-

Notes: GLM = Generalized Linear Model, KPI = Key Performance Indicator, HBM = Hierarchical Bayesian Modeling, SEM = Structural Equation Modeling, CLV = Customer Lifetime Value, ML = Maximum Likelihood, MCMC = Markov Chain Monte Carlo, PLS = Partial Least Squares. (X) indicates non-essentiality.

Table 2. Data Sources and Methods Used in This Work.

2.1 Data and Data Sources

As mentioned above, the model presented in Essay I is based on the observation of how long it takes until an item offered in the buy-it-now offer format is sold. Items which are not sold (up to the time of data collection) can also be taken into account if this information is available. These data (and many potential covariates such as the item condition or information on the seller) can be *crawled* from several online auction platforms; we considered eBay for this purpose. We restricted ourselves to the case of a single product (the iPhone 5S) in order to reduce item heterogeneity. This was because our goal was rather to demonstrate the application of our model than to gain results for a large range of products.

Crawled data were also used for the analysis of online product ratings in Essay III. However, here we combined three different data sources. The individual ratings of a product, which are the dependent variable in our model, were retrieved from Amazon. The same applies to the average score of the product's previous ratings, which, in addition to the product price, is an indicator of the rating customer's expectations of the product in our model. Another indicator, the reputation of the product's brand, was operationalized by a corresponding index (RepTrak®, see Ponzi et al. 2011). The indicator of product performance, called tested product quality, was operationalized by the grades that Stiftung Warentest, a German consumer organization, has assigned to the product based on standardized tests. We evaluated our model for two product groups (digital cameras and television sets) that represent typical search goods, that is, which can largely be assessed before the purchase based on technical functions, specifications, etc. (Nelson 1970). This was in order to ensure a common ground for customer expectation.

Our model for quantifying the impact that a mobile app has on the "lifetime" of subscribers to print media in Essay IV is based on their cancelation behavior. More concretely, it primarily requires data on whether a subscription to one of the two channels (offline or mobile) that was active at a certain time t_0 has been canceled up to another time t_1 and on whether the corresponding customer had a parallel subscription to the respective other channel at t_0. The duration of the subscription up to t_0 and customer demographics are included as covariates. All these data are easily accessible for the publisher but not for researchers. We received a large dataset from a *firm* that publishes a respected national German daily newspaper. This firm has served the offline channel for several decades and the mobile channel for a few years.

We also cooperated with a big and well-known German firm in order to analyze the effects of SEA in Essay II. Here, we conducted a field *experiment* with the firm, in which we instructed it with regard to when and for which search queries it should place ads. It was important to pause the ads for some periods of time because our model requires data on the behavior of search engine users with and without the ad. These data are provided by many search engines to their customers in aggregate form. They regularly include the numbers of searches, clicks, and resulting purchases, the average basket values, and several covariates, each figure per search query and day and separately for the ad and the firm's organic result. They usually do not include the number of competing ads displayed, however, so that we used a crawler to capture this number. We considered only search queries that contain one of the firm's brand names (see the essay for the reasons) and classified each search query manually according to various criteria.

The effect of the data presentation model on the data retrieval performance of BI users in Essay V was also evaluated by an experiment. The experiment was carried out online and, which is one of its distinctive features, among knowledge workers from many different firms as subjects. Each subject was randomly assigned a data model at one of two levels of model complexity and had to solve five typical tasks using a front-end. We measured for each subject and each task how much time the subject spent and whether the task was answered correctly. These data form the dependent variable in our model. The covariates include the subjects' demographics and their reported expertise in areas relevant to BI.

The (type of) data analysis in Essays VI and VII differs from that in the other essays in so far as its purpose is not to investigate user behavior but to demonstrate the suitability of the proposed approaches to reduce search effort for given user behavior. More concretely, we evaluated our archiving component and our search algorithm by measures that are based on or related to the number of reports (or, more generally, documents) that have to be looked through on average until a relevant one is found. For this purpose, we conducted a *simulation study* in Essay VI. In Essay VII, we used a combined approach of a simulation and a laboratory experiment among students.

Finally, we remark that we also conducted simulation studies to validate the models presented in Essays I, II, IV, and V (and their implementation). However, this is only described in detail for Essay II.

2.2 Methods

In this work, we take a quantitative view on the issues we address. Our aim is to analyze not only whether engaging in SEA, offering a mobile app, presenting data according to another model, integrating an archiving component into a BI system, or using a novel search algorithm has an effect on the performance (in the respective meanings) of people but also how large this effect is. In addition, we want to investigate not only which are the determinants of these effects or the figures that they relate to, such as the determinants of consumer demand or online product ratings, but also how strong their influence is.

The arguably most common approach to quantify the influence of a vector of variables $\mathbf{x}$ on a variable y is to regress y on $\mathbf{x}$, i.e., to specify a functional relationship $y = f(\mathbf{x})$. In the simplest case, f is assumed to be linear, but this is not always appropriate. If y must be positive by definition, for example, this should be accounted for by using, e.g., an exponential model. However, it has been recognized that a broad class of different models can be linearized by using a link function g to transform y before analysis, which is why they are called *generalized linear models* (GLMs; McCullagh & Nelder 1983). We make use of several members of this class in Essays I, II, IV, and V for various purposes.

In marketing research, one is often interested in what influences consumer decisions. However, one can usually only observe which alternative is chosen in the end, but not the utility consideration that has led to this choice. The latent utility behind a decision is still accessible for analysis because it can be related to the observable outcome variable. If the decision is between two alternatives, a particular GLM can be used for this purpose, the logit link function $g(y) = \log(y/(1-y))$. Here, y is the probability that the outcome variable (which in this case is a dummy variable) takes a certain value (indicating one of the two alternatives). Note that because y here is a probability, the logit model cannot be estimated using ordinary least squares, as it would be the case for a linear model; this is also true for several other GLMs. Instead, the estimation method is usually maximum likelihood; we refer to McCullagh & Nelder (1983) for further details. We use the logit link function (with two alternatives) in this work to model the decision of a customer to cancel a subscription in Essay IV. It can also be extended to the case when there are more than two alternatives. In Essay II, for example, we use a multinomial version of it (with three alternatives) to model why search engine users click on a firm's ad, its organic result, or not at all.

Decisions are often hierarchical. For example, (only) after a search engine user has decided whether (and where) to click to visit a firm's website, he faces the next decision whether to purchase something there. As the latent utilities behind these decisions may be correlated due to unobserved factors, modeling them by separate GLMs may lead to biased results. A common solution is to combine several GLM-type models using a *hierarchical Bayesian model* (see Rossi et al. 2005 for a marketing-related introduction). We use this approach in Essay II. However, it is difficult to estimate such models analytically, which is why they are often estimated using a simulation-like method such as Markov chain Monte Carlo.

Essay III also addresses a decision: the decision of a customer how to rate a product. However, our focus here is primarily on testing whether the hypothesized influence of two (rather than one) latent constructs, customer expectation and product performance, on this decision is confirmed by empirical data. For such research designs, a *structural equation model* is sometimes preferable to a GLM because it can explain the antecedents (or indicators) and the effects of both latent constructs in one step (while a GLM would need two steps) and because it needs less functional assumptions (see Gefen et al. 2000). Therefore, we use a (partial least squares-type) structural equation model in Essay III.

Research on BI is often more concerned with the analysis of performance or effort rather than decisions. If the goal is just to show that and to measure by how much a certain approach increases performance (or, equivalently, reduces effort), it may suffice to observe differences in key performance indicators with and without the approach and to *test statistically* whether they are significant. This is the case for Essays VI and VII. We also use statistical tests in Essay II, but only to gain first insights into the data.

If the goal is more elaborate (e.g., when various conditions are to be analyzed jointly), however, more complex methods are required. As performance and effort are usually related to the time it takes until a certain event occurs (e.g., until a task is answered), one such method is *survival analysis*, which has been developed specifically for the analysis of duration data. In its parametric form, it is a regression-type approach with the probability that the event of interest does not occur before a certain point in time (or a related figure) as the dependent variable. It also allows for so-called censored observations for which this event cannot be observed anymore (for any reasons). In Essay V, we use a semi-parametric form of survival analysis, the Cox regression (Cox 1972), to evaluate our experiment on user performance in different data models. More concretely, the event of interest here occurs when a task is answered correctly, while incorrect answers are treated as censored observations (see the essay for justification).

In some cases, duration data are also of interest in marketing research. A prominent example of this is *customer lifetime value analysis*. This approach, which may also be regarded a concept rather than a method, aims to estimate the value of a customer to a firm over the entire or future time of his relationship to the firm. For this purpose, often his expected total or remaining lifetime (as a customer) is estimated in a first step, using, e.g., survival analysis. In a second step, the expected profits he generates are projected over his estimated lifetime. This is, in essence, the approach we use in Essay IV to investigate whether offering a mobile app can increase the customer lifetime (value) of subscribers to print media.

In Essay I, we observe the time it takes until an item is sold. However, we are not interested in this time per se (in contrast to, e.g., the customer lifetime in Essay IV) but in what we can learn from it about consumer demand. For this purpose, we model it using survival analysis and derive the parametric form of our model (only) from established standard assumptions of *auction theory*. The variation in the item prices is used to separate how many consumers are interested in an item and how much they value it.

3. Main Findings and Implications

We now summarize our main findings and their implications.

The main contributions of *Essays I, II, and IV* are the models developed in them. These models can be used by practitioners to quantify consumer demand for their products, to assess the effects of SEA for their individual context, and to estimate the impact of serving a second content delivery channel (which does not have to be the mobile channel) on customer lifetime (value), respectively. Obviously, this knowledge supports their marketing decisions, such as which products they should offer (and at which price), for which search queries they should place ads, and which content delivery channels they should serve. Our models can also be used (and be further improved) by future research to address questions that are related to the ones we have investigated. In particular, the model developed in Essay I – although it currently has many important limitations (as described in the essay) – is a novel approach for assessing consumer demand, by which several problems of previous research on this topic can be bypassed. Its central insight is that the time until an item is sold follows (under standard assumptions) an exponential distribution, where the rate parameter of this distribution describes (the abovementioned two aspects of) consumer demand.

Essays I, II, and IV also contribute empirical findings to behavioral marketing research. An interesting result in Essay I is, for example, that consumers seem to overcompensate for shipping costs: We have found that charging 1€ more for shipping decreases their estimated valuation for an item by significantly more than 1€ (or, more concretely, by 4.53€ in our case). The results of Essay II show that engaging in SEA can be beneficial to a firm even for search queries for which its website ranks high among the organic search results and even for users who already search with one of its brand names. This is largely because, somewhat surprisingly, many users who are motivated by the ad to visit the firm's website do so by clicking on the firm's organic result (so that no costs occur to the firm). In addition, we have found that the ad has a positive effect on the probability of a purchase on the firm's website conditional on a click, which may be due to a self-selection process of users with regard to where they click. The magnitudes of the effects of SEA and, thus, its profitability differ between (the defined classes of) search queries. In Essay IV, our results imply a complementary relationship between the offline channel and the mobile channel; that is, offering a mobile app can indeed prolong the lifetime of offline customers and, thus, increase their value to the firm. We have also found evidence for the reverse effect. Of course, as it always is the case for empirical research, it is not clear how generalizable these results are. This is because we have only considered a single product in Essay I and only a single firm (each) in Essays II and IV (and big and well-known firms at that). Future research can, therefore, explore whether our results hold true also for other products or firms, respectively.

The contribution of *Essay III* is mainly a theoretical one. Our results here confirm that both a customer's expectations of a product before the purchase and the product's performance after the purchase have a significant effect on how the customer rates the product, as hypothesized. This finding supports the proposed interpretation of online product ratings as a representation of customer satisfaction rather than product quality, which can refine their current understanding in research. It also has implications for practice because it motivates the design of new rating mechanisms. An example of such a mechanism is described in the essay. Finally, all assumed indicators of the customer's expectations of the product – (the average score of) its previous ratings, its price, and the reputation of its brand – have indeed been found to have a significant influence on this construct, at least for some product groups. However, it has to be noted that our model seems to explain only a small proportion of the variance in online product ratings. This may partly be due to the presence of fake ratings (see the essay for details).

The results of *Essay V* indicate that BI users perform equally well (or poorly) in all data models considered for tasks of low or high difficulty. Which model suits them best for tasks of medium difficulty apparently depends on whether attributes from different dimensions (or relations) are needed. If this is the case, they seem to perform better with the flat model and the relational model than with the multidimensional model, while it is just the other way around otherwise. These findings are largely consistent with the theoretical predictions. They contribute to research on human–computer interaction and advise vendors of BI software to offer several models for data presentation in one product, so that users can choose the model that they prefer. In addition, we have found that a user's expertise in BI has a significant influence on his performance, which confirms that it is useful when firms offer their knowledge workers (more) BI training. This may be especially true for old users because our results suggest that performance decreases with age.

Our analyses in *Essays VI and VII* have confirmed that both the integration of an automated archiving component into a BI system following our framework and the use of our search algorithm can reduce the search effort of users. These results advise vendors of BI and similar software on how to design their products, similarly to the results obtained in Essay V. Note that this here is related to the design science research paradigm as described by Hevner et al. (2004) for information systems research. This paradigm "addresses research through the building and evaluation [emphases removed] of artifacts to meet […] identified business need[s]" (pp. 79–80) and "seeks to create what is effective" (p. 98). In our case, such artifacts are the archiving component and the search algorithm; however, it has to be clarified that we have not rigorously followed the guidelines of Hevner et al. when developing them. Still, they both are described in such a way that firms can easily tailor them to their individual contexts, which relates to one of these guidelines: "Technology-oriented audiences need sufficient detail to enable the described artifact[s] to be […] used within an appropriate organizational context" (p. 90).

4. Bibliographic Information and Editorial Notes

The majority of the essays contained in this work – concretely, Essays I, III, IV, VI, and VII – have already been published. Essays II and VII are extended versions of research that has been published earlier. Table 3 gives details on the respective publications.

The table also shows for each essay (or its earlier stage version) who were the respective co-authors. As can be seen, all essays except for Essay I have profited from contributions of other authors.

Essay		**Stage of work***	**Co-author(s)**	**Year**	**Outlet**		**Reference**
#	**Label**				**Name**	**HB**	
I	Buy-It-Now	current	-	2015	International Conference on Information Systems (ICIS)	0.2	Winter (2015)
II	Search Ads	earlier	Paul Alpar Christian Geißler	2014	International Conference on Information Systems (ICIS)	0.2	Winter et al. (2014)
		current	Paul Alpar	(not yet published)			
III	Ratings	current	Tobias H. Engler Michael Schulz	2015	Journal of Retailing and Consumer Services	0.3	Engler et al. (2015)
IV	Mobile Apps	current	Paul Alpar Christina Kunert	2016	European Conference on Information Systems (ECIS)	0.2	Winter et al. (2016a)
V	Data Models	current	Michael Schulz Paul Alpar	(not yet published)			
VI	Archiving	current	Michael Schulz Sang-Kyu Choi	2015	International Journal of Information Management	0.5	Schulz et al. (2015)
VII	Search Algo.	earlier	Tobias H. Engler Michael Schulz	2014	European Conference on Social Media	(not rated)	Engler et al. (2014)
		current	Michael Schulz Tobias H. Engler	2016	Analyzing the Strategic Role of Social Networking in Firm Growth and Productivity (Book)	(not rated)	Winter et al. (2016b)

Note: The column "HB" gives the rating assigned to the respective publication outlet by the German business newspaper Handelsblatt in 2014 in the category business administration (see Gygli et al. 2014 for details). This rating is the current German de facto standard for assessing the quality of a scientific outlet in business administration and related disciplines (such as information systems). Just being covered by Handelsblatt at all already indicates a respectable quality. The rating is ordinal with levels of 0.1, 0.2, 0.3, 0.5, 0.7, and 1.0, where higher values indicate better quality.

* "current" indicates that there is no difference in contribution between the respective published version of the essay and the version contained in this work. However, these versions still are not exactly identical (see the text for details).

Table 3. Details on the Publication of the Essays Contained in This Work.

The essays as contained in this work are not exactly identical to their (to-be-)published version. This is because they have been homogenized with regard to layout, formalities (e.g., citation format), and especially writing style. Typographical and other errors have been corrected if found. Many passages have been reworded (without changing their content). References to the earlier stage versions of Essays II and VII have been removed; similarly, cross-references between the essays have been resolved. Essay III has been partly restructured. In addition, the presentation of content has been improved (e.g., in terms of clarity or cogency) in a few cases by adding, extending, shortening, or deleting some passages or by revising some figures and tables; this relates primarily to Essays III and VII. However, the substantial content of each essay has remained untouched, so that there is no difference in contribution between its (to-be-)published version and the version contained in this work.

References

Alpar, P., & Schulz, M. (2016): Self-service business intelligence. *Business & Information Systems Engineering*, Vol. 58, No. 2, pp. 151–155.

Bajari, P., & Hortaçsu, A. (2003): The winner's curse, reserve prices, and endogenous entry: Empirical insights from eBay auctions. *The RAND Journal of Economics*, Vol. 34, No. 2, pp. 329–355.

Bell, D., Deighton, J., Reinartz, W. J., Rust, R. T., & Swartz, G. (2002): Seven barriers to customer equity management. *Journal of Service Research*, Vol. 5, No. 1, pp. 77–85.

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012): Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, Vol. 36, No. 4, pp. 1165–1188.

Cox, D. R. (1972): Regression models and life-tables. *Journal of the Royal Statistical Society – Series B (Methodological)*, Vol. 34, No. 2, pp. 187–220.

Eckerson, W. W. (2008): Pervasive business intelligence – Techniques and technologies to deploy BI on an enterprise scale. Best practice report, TDWI.

Engler, T. H., Schulz, M., & Winter, P. (2014): Towards a social data enriched search algorithm for business intelligence portals. In *Proceedings of the 1st European Conference on Social Media*, pp. 740–743.

Engler, T. H., Winter, P., & Schulz, M. (2015): Understanding online product ratings: A customer satisfaction model. *Journal of Retailing and Consumer Services*, Vol. 27, pp. 113–120.

Fornell, C. (1992): A national customer satisfaction barometer: The Swedish experience. *Journal of Marketing*, Vol. 56, No. 1, pp. 6–21.

Gefen, D., Straub, D., & Boudreau, M.-C. (2000): Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems*, Vol. 4, Art. 7.

Gygli, S., Müller-Dröge, H. C., & Schläpfer, J. (2014): Handelsblatt BWL-Ranking 2014 – Methodik und Zeitschriftengewichte. http://htmldb-hosting.net/pls/htmldb/FMONITORING.download_my_file?p_file=921.

Haas, M. R., & Hansen, M. T. (2007): Different knowledge, different benefits: Toward a productivity perspective on knowledge sharing in organizations. *Strategic Management Journal*, Vol. 28, No. 11, pp. 1133–1153.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004): Design science in information systems research. *MIS Quarterly*, Vol. 28, No. 1, pp. 75–105.

Hu, N., Pavlou, P. A., & Zhang, J. (2006): Can online reviews reveal a product's true quality? In *Proceedings of the 7th Conference on Electronic Commerce*, pp. 324–330.

Imhoff, C., & White, C. (2011): Self-service business intelligence – Empowering users to generate insights. Best practice report, TDWI.

Işik, Ö., Jones, M. C., & Sidorova, A. (2013): Business intelligence success: The role of BI capabilities and decision environments. *Information & Management*, Vol. 50, No. 1, pp. 13–23.

Kar, A. K., Pani, A. K., & De, S. K. (2010): A study on using business intelligence for improving marketing efforts. *Business Intelligence Journal*, Vol. 3, No. 2, pp. 141–150.

Koh, N. S., Hu, N., & Clemons, E. K. (2010): Do online reviews reflect a product's true perceived quality? An investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications*, Vol. 9, No. 5, pp. 374–385.

Kurniawan, Y., Gunawan, A., & Kurnia, S. G. (2014): Application of business intelligence to support marketing strategies: A case study approach. *Journal of Theoretical and Applied Information Technology*, Vol. 64, No. 1, pp. 240–248.

McCarthy, J. (1960): Basic marketing: A managerial approach. Homewood: Richard D. Irwin.

McCullagh, P., & Nelder, J. A. (1983): Generalized linear models. London: Chapman and Hall.

Nelson, P. (1970): Information and consumer behavior. *Journal of Political Economy*, Vol. 78, No. 2, pp. 311–329.

Ponzi, L. J., Fombrun, C. J., & Gardberg, N. A. (2011): RepTrak^{TM} Pulse: Conceptualizing and validating a short-form measure of corporate reputation. *Corporate Reputation Review*, Vol. 14, No. 1, pp. 15–35.

PwC (2015): Global entertainment and media outlook 2015–2019. http://www.pwc.com/gx/en/global-entertainment-media-outlook/assets/2015/internet-advertising-key-insights-3-search.pdf.

Rossi, P. E., Allenby, G., & McCulloch, R. (2005): Bayesian statistics and marketing. Hoboken: John Wiley & Sons.

Schubmehl, D., & Vesset, D. (2014): The knowledge quotient: Unlocking the hidden value of information using search and content analytics. White paper, IDC.

Schulz, M., Winter, P., & Choi, S.-K. (2015): On the relevance of reports – Integrating an automated archiving component into a business intelligence system. *International Journal of Information Management*, Vol. 35, No. 6, pp. 662–671.

Trevathan, J., & Read, W. (2006): Undesirable and fraudulent behaviour in online auctions. In *Proceedings of the 1st International Conference on Security and Cryptography*, pp. 450–458.

Westbrook, R. A., & Reilly, M. D. (1983): Value-percept disparity: An alternative to the disconfirmation of expectations theory of consumer satisfaction. In *Proceedings of the 10th North American Advances in Consumer Research Conference*, pp. 256–261.

White, C. (2005): The next generation of business intelligence: Operational BI. *DM Review*, Vol. 15, No. 5, pp. 34–37.

Winter, P. (2015): Quantifying consumer interest and consumer valuation from buy-it-now offers. In *Proceedings of the 36th International Conference on Information Systems*.

Winter, P., Alpar, P., & Geißler, C. (2014): When does brand bidding pay off (even) if website competition is low? In *Proceedings of the 35th International Conference on Information Systems*.

Winter, P., Alpar, P., & Kunert, C. (2016a): Can mobile apps defend print media? In *Proceedings of the 24th European Conference on Information Systems*.

Winter, P., Schulz, M., & Engler, T. H. (2016b): Finding information faster by tracing my colleagues' trails – A reference algorithm for enterprise search. In Benson, V., Tuninga, R., & Saridakis, G. (Eds.): Analyzing the strategic role of social networking in firm growth and productivity, Hershey: IGI Global, pp. 406–426.

Wordstream (2016): Online ads: A guide to online ad types and formats. http://www.wordstream.com/online-ads.

Note on Electronic Version

The essays contained in (the printed version of) this work have not been included in its electronically published version in order to prevent copyright issues. The published papers on which they are based can be found at the following links (as of September 13, 2016):

Essay		**Stage of work***	**Link to published version**
#	**Label**		
I	Buy-It-Now	current	http://aisel.aisnet.org/icis2015/proceedings/eBizeGov/6/
II	Search Ads	earlier	http://aisel.aisnet.org/icis2014/proceedings/EBusiness/51/
		current	(not yet published)
III	Ratings	current	http://www.sciencedirect.com/science/article/pii/S0969698915300138
IV	Mobile Apps	current	http://aisel.aisnet.org/ecis2016_rp/133/
V	Data Models	current	(not yet published)
VI	Archiving	current	http://www.sciencedirect.com/science/article/pii/S0268401215000675
VII	Search Algo.	earlier	http://www.academic-bookshop.com/shop/search.php?q=ecsm&searchbutton=Go%21 or https://books.google.de/books?id=qVQJBAAAQBAJ, pp. 740–743.
		current	http://www.igi-global.com/chapter/finding-information-faster-by-tracing-my-colleagues-trails/161859

* "current" indicates that there is no difference in contribution between the respective published version of the essay and the version contained in this work. However, these versions still are not exactly identical (see the synopsis for details).