Marketing Analytics for Data-Rich Environments

Michel Wedel & P. K. Kannan

May 13, 2016

Michel Wedel is PepsiCo Chaired Professor of Consumer Science at the Robert H. Smith School of
Business, and a Distinguished University Professor at the University of Maryland, College Park, MD
20742; e-mail: mwedel@rhsmith.umd.edu.

P. K. Kannan is Ralph J. Tyser Professor of Marketing Science at the Robert H. Smith School of
Business, University of Maryland, College Park, MD 20742; e-mail: pkannan@rhsmith.umd.edu.

Marketing Analytics for Data-Rich Environments

Abstract

We provide a critical examination of marketing analytics methods, tracing their historical development, examining their applications to structured and unstructured data generated within or external to a firm, and reviewing their potential to support marketing decisions. We identify directions for new analytical research methods, addressing (a) analytics for optimizing marketing mix spending in a data-rich environment, (b) analytics for personalization, and (c) analytics in the context of customers' privacy and data security. We review the implications for organizations that seek to implement big data analytics. Finally, turning to the future, we identify trends that will shape marketing analytics as a discipline and marketing analytics education.

Keywords: Big Data, Marketing Analytics, Unstructured Data, Marketing Mix, Personalization, Privacy, Data Security.

Marketing Analytics for Data-Rich Environments

1. Introduction and Motivation

Data has been called "the Oil" of the digital economy. The routine capture of digital information via online and mobile applications produces vast data-streams on how consumers feel, behave and interact around products and services, and how they respond to marketing efforts. Data is assuming an ever more central role in organizations, as marketers seek to harness it to build and maintain customer relationships, personalize products, services and the marketing mix, and automate marketing processes in real time. The explosive growth of media, channels, digital devices and software applications has provided firms with unprecedented opportunities to leverage data to provide more value to customers, enhance their experiences, increase their satisfaction and loyalty, and extract value. Although initially the potential of big data may have been over hyped and companies may have invested too much in data capture and storage and not enough in analytics, it is becoming clear that the availability of big data is spawning data-driven decision cultures in companies, is providing them with competitive advantages, and is having a significant impact on their financial performance. The increasingly widespread recognition that big data can be leveraged effectively to support marketing decisions is highlighted by the success of industry leaders, and entirely new forms of marketing have emerged, including recommendations, geo-fencing, search marketing, and retargeting. Marketing analytics has come to play a central role in these developments, and there is urgent demand for new, more powerful metrics and analytical methods that make data-driven marketing operations more efficient and effective. However, it is yet not sufficiently clear which types of analytics work for which types of problems and data, what new methods are needed for analyzing new types of data, or how

companies and their management should evolve to develop and implement skills and procedures to compete in this new environment.

The *Marketing Science Institute* (2014-2016) has outlined the scope of research priorities around these issues. The present paper provides a review of research on one of these priorities: analytics for data rich environments. We have structured our thoughts using the framework in Figure 1. At the center is the use of analytics to support marketing decisions, which is founded on the one hand on the availability of data, and on the other hand on advances in analytical methods. Key domains for analytics applications are (1) in customer relationship management (CRM) with methods that help acquisition, retention and satisfaction of customers to improve their lifetime value to the firm[1], (2) the marketing mix, with methods, models and algorithms that support the allocation of resources to enhance the effectiveness of marketing effort, (3) personalization of the marketing mix to individual consumers, where because of the development of various approaches to capture customer heterogeneity significant advances have been made, and (4) privacy and security, an area that is of growing concern to firms and regulators. This leads to two pillars of the successful development and implementation of marketing analytics in firms: the adoption of organizational structures and cultures that foster data driven decision making, and the education and training of analytics professionals.

The agenda for this paper is as follows. Using the framework in Figure 1, we provide a brief review of the history of marketing data and analytics, followed by a critical examination of the extent to which specific analytical methods are applicable in data rich environments and support marketing decision making in core domains. This analysis leads to the identification of future directions. We choose to focus on (a) analytics for optimizing marketing mix spending, (b)

---

[1] We do not focus on CRM issues other than personalization in this paper as CRM is covered in depth by another paper in this special issue.

analytics for personalization of the marketing mix, and (c) analytics in the context of data security and customer privacy. We review the implications for implementing big data analytics in organizations and for analytics education and training. In doing so, we identify trends that will shape marketing as a discipline, and discuss actual and aspired interconnections between marketing practice and academia.
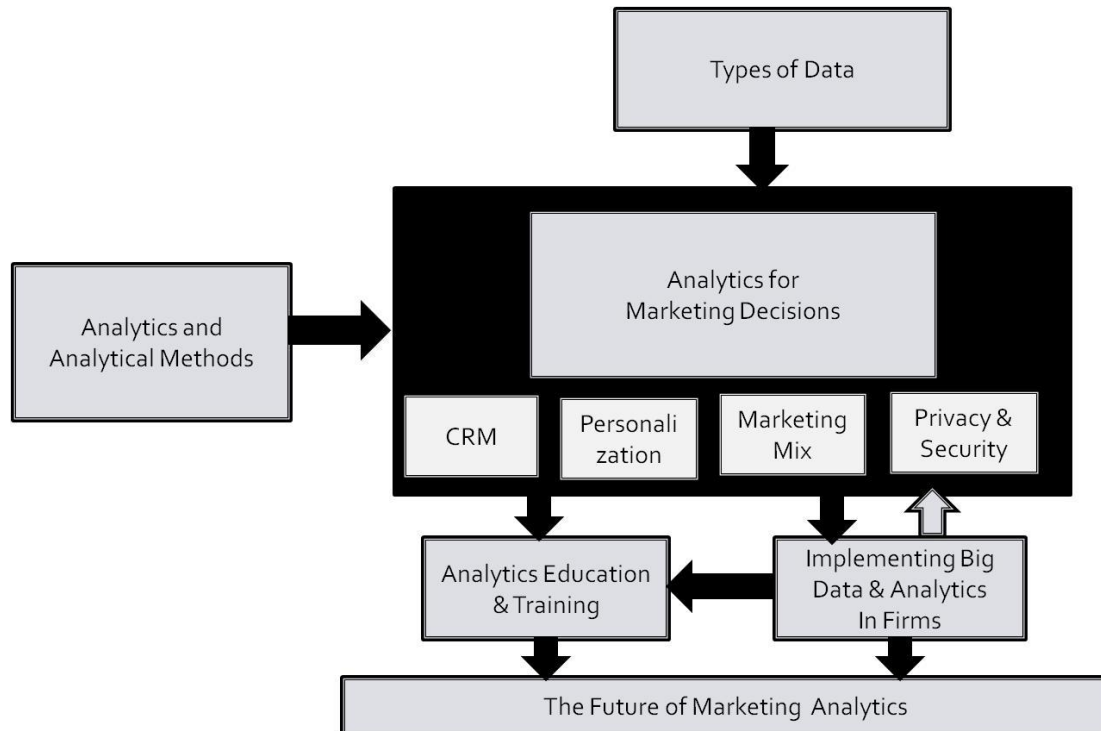

Figure 1: Framework of the Paper
Marketing data and analytical methods are used in four main areas of marketing decisions. Their implementation in firms depends on firm culture and organizational structure, and poses requirements for education and training, which will shape the future of Marketing Analytics.

2. A Brief History of Marketing Data and Analytics

Marketing Analytics involves collection, management, and analysis – descriptive, diagnostic, predictive and prescriptive – of data to obtain insights into marketing performance, to maximize the effectiveness of instruments of marketing control, and optimize their return on investment (ROI). It is interdisciplinary, being at the nexus of marketing and other areas of

business, mathematics, statistics, economics, econometrics, psychology, psychometrics, and more recently also computer science. While it has a long history, due to explosive growth in the availability of data in the digital economy in the last two decades firms' have increasingly recognized the key competitive advantages that analytics may afford, which has propelled its development and deployment (Davenport 2006).

2.1. Available Data

The history of the systematic use of data in marketing starts around 1910 with the work of Parlin for the Curtis Publishing Company in Boston (Bartels 1988, p. 125). Parlin gathered information on markets to guide advertising and other business practices, prompting several major US companies to establish commercial research departments. Duncan (1919) emphasized the use of external in addition to internal data by these departments. Questionnaire survey research, already done in the context of opinion polls by Gallup in the 1820s, became increasingly popular in the 1920s (Reilly 1929). Around that time concepts from psychology were being brought into marketing to foster greater understanding of the consumer. Starch's (1923) Attention, Interest, Desire, Action (AIDA) model is a prime example, and Starch is credited for the widespread adoption of copy research. This era also saw the first use of eye-tracking data (Nixon 1924).

In 1923, A. C. Nielsen founded one of the first market research companies. Nielsen started with measuring product sales in stores, and in the 1930s and 1950s began assessing radio and television audiences. Burke was founded in the US (1931) and initially did product testing research for P&G. Around the same time (1934), the market research firm GfK was established in Germany. The next decade saw the rise of field experiments and the increased use of telephone surveys (White 1931). Panel data became increasingly popular, at first mostly for

measuring media exposure, but in the 1940s they begun to be used for recording consumer purchases (Stonborough 1942). The use of companies' own customer data was stimulated around 1961 by Cullinan, who introduced the "Recency, Frequency, Monetary" (RFM) metrics that became central in CRM (Customer Relationship Management) (Neslin 2014). In 1966, the SAMI (Selling Areas Marketing Institute) was founded, which focused on warehouse withdrawal data. The importance of computers for marketing research was first recognized around that time as well (Casher 1969).

Starting in the late 1970s, geo-demographic data was amassed from government databases and credit agencies by Claritas, founded on the work by the sociologist Booth around 1890. The introduction of the Universal Product Code (UPC) code and IBM's computerized POS (Point of Sale) scanning devices in food retailing in 1972 marked the first automated capture of data by retailers. Companies such as Nielsen quickly recognized the promise of using POS scanner data for research purposes, and replaced bi-monthly store audits with more granular scanner data. Soon, individual customers could be traced through loyalty cards, which led to the emergence of scanner panel data (Guadagni and Little 1983). IRI (Information Resources, Inc.), which since its inception in 1979 measured TV advertising, rolled out its in-home barcode scanning service in 1995.

The use of internal customer data was greatly propelled by the introduction of the personal computer (PC) to the mass market by IBM in 1981. The PC allowed marketers to store data on current and prospective customers, which contributed to the emergence of data-base marketing, pioneered by the Kestnbaums and by Shaw (1987). CRM software emerged around 1990, for which earlier work on Sales Force Automation at Siebel Systems paved the way. The PC also facilitated survey research via personal (CAPI) and telephone (CATI) interviewing.

In1995, the world-wide-web came into existence after more than two decades of development at the Defense Advanced Research Projects Agency (DARPA) and other organizations and this led to the availability of large volumes of marketing data. Click-stream data extracted from server logs were used to track page-views and clicks via cookies. Click-through data yielded measures of the effectiveness of online advertising. The internet stimulated the development of CRM systems by firms such as Oracle, and in 1999 Salesforce was the first company to deliver CRM systems via cloud computing.

In 1998 Google was founded, which championed key-word search and the capture of search data. Search engines had been around since about a decade earlier: the first FTP search engine Archie was developed at McGill University. The advent of user generated content (UGC), including online product reviews, blogs and video, resulted in increasing volume and variety of data. The launch of Facebook in 2004 opened up an era of social network data. Vast amounts of data in the form of text and video uploaded by users, with the advent of Youtube in 2005, became the raw material for behavioral targeting. Twitter, with its much simpler 140-character messages, followed suit in 2006. Smart-phones existed since the early 1990s, but the introduction of the Apple iPhone in 2007 with global positioning (GPS) capabilities marked the onset of the capture of consumer location data at an unprecedented scale.

2.2. Analytics

The initiative of the Ford Foundation and the Harvard Institute of Basic Mathematics for Applications in Business (1959/1960) is widely credited for having provided the major impetus for the application of analytics to marketing (Winer and Neslin 2014). It led to the founding of the Marketing Science Institute (MSI) in 1961, which has had a continued role in bridging marketing academia and practice ever since. Statistical methods such as Analysis of Variance

had been applied in marketing research for over a decade (Ferber 1949), but the development of statistical and econometric models tailored to specific marketing problems took off when marketing was recognized as a field of decision making through the Ford/Harvard initiative (Bartels, 1988, p. 125). The development of Bayesian decision theory at the Harvard Institute (Raiffa and Schlaifer 1961) also played a role, exemplified by its successful application to, amongst others, pricing decisions by Green (1963). Academic research in marketing began to focus more on the development of statistical models and predictive analytics. While is not possible to review all subsequent developments here (see Winer and Neslin 2014 for an extensive review), a few landmarks will be noted.

New product diffusion models (Bass 1969) involved applications of differential equations from epidemiology. Stochastic models of buyer behavior (Massy, Montgomery and Morrison 1970) were rooted in statistics and involved distributional assumptions on measures of consumers' purchase behavior. The application of decision calculus (Little and Lodish 1969; Lodish 1971) to optimize spending on advertising and sales force became popular after its introduction to marketing by Little in 1970. Market share and demand models for store-level scanner data (Nakanishi and Cooper 1974) were derived from econometric models of demand. Multidimensional scaling and unfolding techniques, founded in psychometrics (Coombs 1950), became an active area of research with key contributions by Green (1969) and DeSarbo (DeSarbo and Rao 1986). These techniques enabled market structure and product positioning research by deriving spatial maps from proximity and preference judgments, and choice. Conjoint analysis (Green and Srinivasan 1978), and later conjoint choice analysis (Louviére and Woodworth 1983), are unique contributions that evolved from work in psychometrics by Luce on the quantification of psychological attributes (Luce and Tukey 1964). Scanner panel based

multinomial logit models (Guadagni and Little 1983) were built directly upon the work in econometrics by McFadden (1974). The nested logit model that captures hierarchical consumer decision making was introduced in marketing (Kannan and Wright 1991), and it was recognized that models of multiple aspects of consumer behavior (incidence, choice, timing, quantity) could be integrated (Gupta 1988). This proved to be a powerful insight for models of RFM (Schmittlein and Peterson 1994). Whereas earlier methods to identify competitive market structures were based on estimated cross-price elasticities, models that derive competitive maps from panel choice data were developed based on the notion that competitive market structures arise from consumer perceptions of substitutability, revealed through choices of products (Elrod 1988). Time-series methods (DeKimpe and Hansens 1995) enabled researchers to test whether marketing instruments resulted in permanent or transient changes in sales.

Heterogeneity in the behaviors of individual consumers became a core premise upon which marketing strategy was based, and the mixture choice model was the first to enable managers to identify response-based consumer segments from scanner data (Kamakura and Russell 1989). This model was generalized to accommodate a wide range of models of consumer behavior (Wedel and DeSarbo 1995). Heterogeneity of consumers was represented in a continuous fashion in Hierarchical Bayes models (Rossi, McCulloch and Allenby 1996). While initially researchers debated which of these two approaches represented heterogeneity best, it was shown that they each match specific types of marketing problems, with few differences between them (Andrews, Ainslie and Currim 2002). It can be safely said that the Bayesian approach is now one of the dominant modeling approaches in marketing, offering a powerful framework to develop integrated models of consumer behavior (Rossi and Allenby 2003). Such models have been successfully applied to eye-tracking of advertisements (Wedel and Pieters

2000), email marketing (Ansari and Mela 2003), web-browsing (Montgomery et al. 2004), social networks (Moe and Trusov 2011), and paid search advertising (Rutz, Trusov and Bucklin 2011).

The derivation of profit maximizing decisions, inspired by the work of Dorfman and Steiner (1954) in economics formed the basis of the operations research (OR) approach to optimal decision making in advertising (Parsons and Bass 1971), sales force allocation (Mantrala, Sinha and Zoltners 1994), target selection in direct marketing (Bult and Wansbeek 1995), and customization of online price discounts (Zhang and Krishnamurthi 2004). Structural models founded in economics include approaches that supplement aggregate demand equations with supply side equilibrium assumptions (Chintagunta 2002), based on the work of the economists Berry, Levinsohn and Pakes (1995). A second class of structural models accommodates forward looking behavior (Erdem and Keane 1996), based on the work in economics by Rust (1987). Structural models allow for predictions of shifts in behavior of agents when policy changes are implemented (Chintagunta, Rossi, Erdem and Wedel 2006).

2.3. From Theory to Practice

The impact of these academic developments on marketing practice was empirically demonstrated by Roberts et al. (2014). Through interviews among managers, they found a significant impact of several analytics tools on firm decision making. The relevance of the above developments for the practice of marketing is further evidenced by examples of companies that spun off from academic work. Early cases of successful companies include *Starch and Associates*, a company that specialized in ad copy testing based on Starch's academic work, and Little and Urban's *Management Decision Systems,* which was later sold to *IRI*. Zoltman and Sinha's work on Sales force allocation was implemented in practice through *ZS Associates*. The work by Fornell on the measurement of satisfaction led to the American Consumer Satisfaction

Index, produced by his company the *CFI group*. Hanssens' models on long run effectiveness of the marketing mix were successfully implemented by the company *MarketShare* that he cofounded. Steenkamp founded *Aimark*, a joint venture with GfK that applies academic methods and concepts especially in international marketing. Virtually all of these companies became successful through the application of analytics.

Examples of companies with very close ties to academia include Johnson's *Sawtooth Software*, which specializes in the design and analysis of and software for conjoint studies, and Cohen and Garratt's *In4mation Insights*, which applies comprehensive Bayesian statistical models to a wide range of applied problems including marketing mix modeling. In some cases, marketing academia lags behind developments in practice and so focuses instead on their impact and validity. In other cases, academics are co-investigators relying on data and problems provided by companies and working together with them to develop implementable analytics solutions. Yet in a growing number of application areas in the digital economy, reviewed below, academics are leading the development of new concepts and methods.

2.4. Synthesis

The development of data-driven analytics in marketing from around 1900 until the introduction of the World Wide Web in 1995 has progressed through roughly three stages: (1) description of observable market conditions through simple statistical approaches, (2) development of models to provide insights and diagnostics using theories from economics and psychology, and (3) evaluation of marketing policies, predicting their effects and supporting marketing decision making using statistical, econometric and OR approaches. In many cases, soon after new sources of data became available, methods to analyze them were introduced or developed (see Figure 2 for an outline of the history of data and analytical methods; Table 1 summarizes state-

of-the-art approaches). Many of the methods developed by marketing academics since the 1960s

have now found their way into practice, where they support decision making in areas such as

CRM, marketing mix and personalization, and have increased the financial performance of the
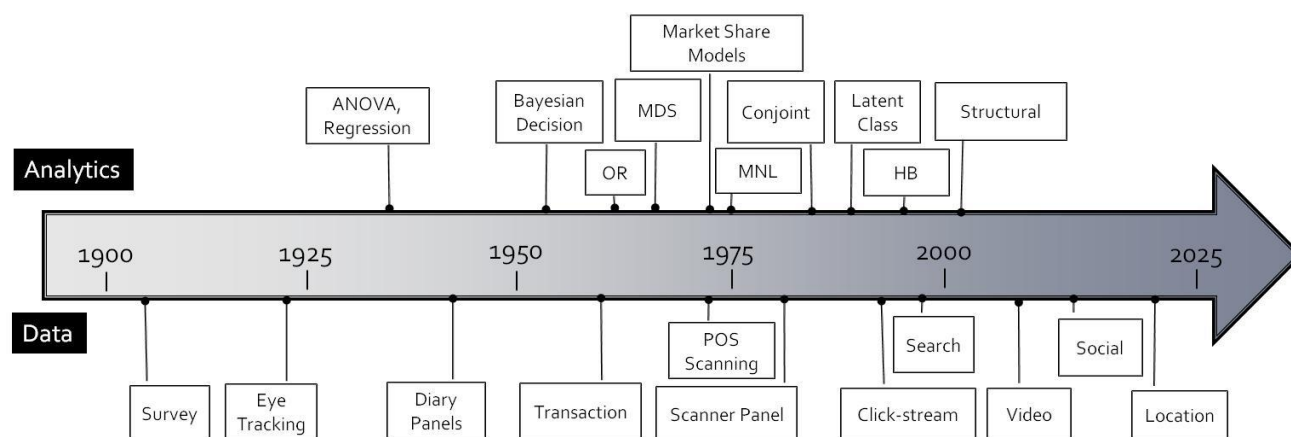
firms deploying them.

Figure 2: An Outline of the Timeline of Marketing Data and Analytics
This timeline summarizes the availability of new marketing data and the development of the major classes
of marketing models. As new types of data became available, new models to analyze them followed.

Since 2000, the automated capture of online click-stream, messaging, word-of-mouth,

transaction and location data has greatly reduced the variable cost of data collection, and has

resulted in unprecedented volumes of data that provide insights on consumer behavior at

exceptional levels of depth and granularity. Although academics have taken up the challenge to

develop diagnostic and predictive models for these data in the last decade, these developments

are admittedly still in their infancy. On the one hand, descriptive metrics displayed on

dashboards are popular in practice. This could be the result of constraints on computing power

and the need for rapid real-time insights, a lack of trained analysts, and/or organizational barriers

to implementing advanced analytics. Especially unstructured data in the form of blogs, reviews

and tweets, offer opportunities for deep insights into the economics and psychology of consumer

behavior, which, once appropriate models are developed and applied, could usher in the second

stage in digital marketing analytics. On the other hand, machine learning methods from computer science (including deep neural networks and cognitive systems discussed below; see Table 1) have become popular in practice, but have seen little research in marketing academia. Their popularity may stem from their excellent predictive performance, and black-box nature which enables routine application with limited analyst intervention. The question is whether marketing academics should jump on that bandwagon, which they may have been reluctant to do because these techniques do not establish causal effects, or produce generalizable theoretical insights. However, combining these approaches with more classical models for marketing analytics may address these shortcomings and hold promise for future research (Table 2). It is reasonable to expect that the third step in the evolution of analytics in the digital economy, the development of models to generate diagnostic insights and support real time decisions from big data, is imminent. However, marketing academia will need to develop analytical methods with a keen eye for data volume and variety and speed of computation, which have thus far been largely ignored (see Table 2). The remainder of this paper reviews recent developments, and identifies potential barriers and opportunities towards successful implementation of analytics in data-rich environments to support marketing decisions.

3. Data and Analytics

3.1 Types of Data

Big data is often characterized by the "4Vs" – *Volume* (from Terabytes to Petabytes), *Velocity* (from one-time snap shots to high frequency and streaming data), *Variety* (numeric, network, text, images, and video), and *Veracity* (reliability and validity). The first two characteristics are important from a computing standpoint, the second two are important from an analytics standpoint. Sometimes a fifth "V" is added: *Value*. It transcends the first four and is important from a business standpoint. Big data is mostly observational, but surveys, field

experiments, and lab experiments may yield data of large variety and high velocity. Much of the excitement surrounding big data is exemplified by the scale and scope of *observational data* generated by the "big three of big data" – Google, Amazon and Facebook. Google receives over 4 million search queries per minute from the 2.4 billion Internet users around the world and processes 20 petabytes of information per day. Facebook's 1.3 billion users share 2.5 million pieces of content in a minute. Amazon has created a marketplace with 278 million active customers from which it records data on online browsing and purchasing behavior. These and other firms have changed the landscape of marketing in the last decade through the generation, provision and utilization of big data. Emerging solutions to link customer data across online and offline channels and across TV, tablet, mobile and other digital devices will contribute further to the availability of data. Further, in 2014 well over 15 billion devices were equipped with sensors, which enable them to connect and transfer data over networks without human interaction. This Internet of Things (IoT) may become a major source of new product and service development, and generate massive data in the process.

Surveys have become much easier to administer with the advances in technology enabling online and mobile data collection (Amazon's MTurk). Firms continuously assess customer satisfaction; new digital interfaces require this to be done with short surveys to reduce fatigue and attrition. For example, loyalty is often evaluated with single-item Net Promoter scores. As a consequence, longitudinal and repeated cross-section data is becoming more common. Mittal, Kumar and Tsiros (1999) use such data to track the drivers of customer loyalty over time. To address the issue of shorter questionnaires, analytic techniques have been developed to enable personalized surveys that are adaptive in nature based on the responses to earlier questions

(Kamakura and Wedel 1995), and the design of tailored split questionnaire designs for massive surveys (Adigüzel and Wedel 2008).

Digital technologies facilitate large-scale *field experiments* that produce big data and have become powerful tools for eliciting answers to questions on the causal effects of marketing actions. For example, large scale A/B testing enables firms to "test and learn" for optimizing website designs, (search, social and mobile) advertising, behavioral targeting, and other aspects of the marketing mix. Hui et al. (2013) use field experiments to evaluate mobile promotions in retail stores. Alternatively, natural (or quasi-) experiments capitalize on exogenous shocks occurring naturally in the data to establish causal relations, but often more extensive analytical methods are required to establish causality, including matching and instrumental variables methods. For example, Ailawadi et al. (2010) show how quasi-experimental designs can be used to evaluate the impact of the entry of Wal-Mart stores on retailers, using a before-after design with control group of stores matched on a variety of measures. Another way to leverage big data to assess causality is to examine thin slices of data around policy changes that occur naturally in the data, which can reveal the impact of those actions on dependent variables of interest via so called regression-discontinuity designs (Hartman, Nair and Narayanan 2011).

Finally, *lab experiments* typically generate smaller volumes of data, but technological advances have allowed administration online and collection of audio, video, eye-tracking, face-tracking (Teixeira, Wedel and Pieters 2010), and neuromarketing data obtained from EEG and brain imaging (Telpaz, Webb and Levy 2015). Such data are for example collected routinely by A.C. Nielsen, and often yields "$p > n$" data with more variables than respondents. Meta-analysis techniques can be used to generalize findings across large numbers of these experiments (Bijmolt, Van Heerde and Pieters, 2005).

3.2 Software for Big Data Processing[2]

Figure 3 provides an overview of the classes of marketing data discussed above and methods to store and manipulate it. For small to medium-sized structured data, the conventional methods such as Excel spreadsheets, ASCII files or datasets of statistical packages such as SAS, S-Plus, STATA and SPSS are adequate. SAS holds up particularly well as data size grows, and is popular in many industry sectors (e.g. retailing, financial services and government) for that reason. As the number of records goes into the millions, relational databases such as MySQL, used by, for example, Wikipedia, are increasingly effective for data manipulation and for querying. For big and real-time web applications where volume, variety and velocity is high, databases such as NoSQL (not only SQL) are the preferred choice because they provide a mechanism for storage and retrieval of data that does not require tabular relations like those in relational databases, and can be scaled out across commodity hardware. Apache Cassandra, an open-source software initially developed by Facebook, is a good example of such a distributed data base management system. Hadoop, originally developed at Yahoo!, is a system to store and manipulate data across a multitude of computers, written in the Java programming language. At its core are the Hadoop Distributed File Management System (HDFS) for data storage, and the MapReduce programming framework for data processing. Typically, applications are written in a language such as Pig, which maps queries across pieces of data that are stored across hundreds of computers in a parallel fashion, and then combines the information from all to answer the query. SQL engines such as Dremel (Google), Hive (Hortonworks), and Spark (Databricks) allow very short response times. For post-processing, however, such high frequency data are often still stored in relational databases with greater functionality.

---

[2] The web Appendix provides links to explanations of terms used in this and other sections.

C++, FORTRAN and Java are powerful and fast low-level programming tools for analytics that come with large libraries of routines. Java programs are often embedded as applets within the code of web pages. R, used by Google, is a considerably slower but often used open-source higher level programming language with functionality comparable to languages such as Matlab. Perl is software that is suited for processing unstructured clickstream (HTML) data, and was initially used by Amazon, but has been mostly supplanted by its rival Python (used by Dropbox), which is a more intuitive programming language that enables MapReduce implementation. Academic research in Marketing Analytics today already relies on many of these programming languages, where R appears to be the most popular. Much of this software for big data management and processing likely will become an integral part of the ecosystem of marketing academics and applied marketing analysts in the near future.
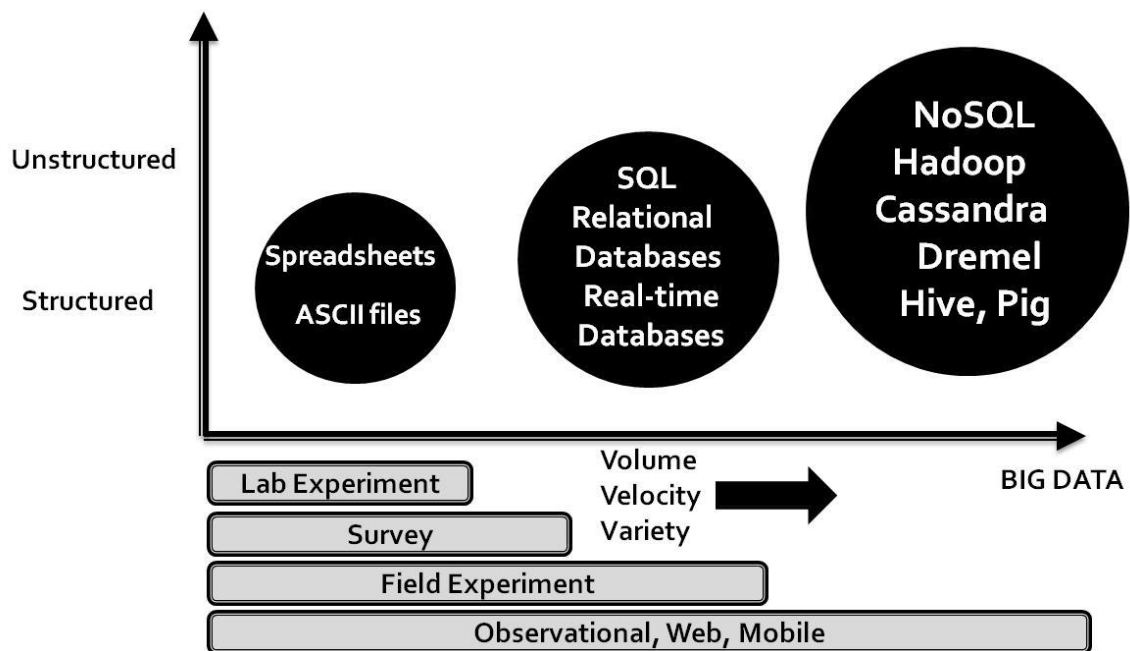
Figure 3: Managing Types of Data
The vertical axis shows the degree of structure in the data and the horizontal axis shows the dimensions resulting in big data. Software to manage these data is shown in the core of the figure.

3.3 Volume, Variety, Velocity: Implications for Big Data Analytics

The question is whether better business decisions require more data or better models. Some of the debate surrounding that question goes back to research at Microsoft, where Banko and Brill (2001) showed that in the context of text mining, algorithms of different complexity performed similarly, but adding data greatly improved performance. Indeed, throughout the academic marketing literature complex models barely outperform simpler ones on datasets of small to moderate size. The answer to the question is rooted in the bias-variance tradeoff. On the one hand, bias results from an incomplete representation of the true Data Generating Mechanism (DGM) by a model because of simplifying assumptions. A less complex model (one that contains fewer parameters) often has a higher bias, but a model needs to simplify reality to provide generalizable insights. To quote George Box: "*All models are wrong but some are useful.*" A simple model may produce tractable closed form solutions, but numerical and sampling methods allow for examining more complex models at higher computational cost. Model averaging and ensemble methods such as bagging or boosting address the bias in simpler models by averaging over many of them (Hastie, Tibshirani, and Friedman 2008). In marketing, researchers routinely use model free evidence to provide confidence that more complex models accurately capture the DGM (see for an example Bronnenberg, Dube and Gentzkow 2012). Field experiments are increasingly popular because data quality (veracity) can substitute for model complexity: when the DGM is under the researchers' control, simpler models can be used to make causal inferences (Hui et al. 2013). Variance, on the other hand, results from random variation in the data due to sampling and measurement error. A larger volume of data reduces the variance. Complex models calibrated on smaller datasets often over-fit. That is, they capture random error rather than the DGM. That more data reduces the error is well known to benefit

machine learning methods such as neural networks, which are highly parameterized (Geman, Bienenstock, and Doursat 1992). But, not all data is created equal. Bigger volume of data reduces variance and even simpler models will fit better. But, as data variety increases and data gets richer, the underlying DGM expands. Much of the appeal of big data in marketing is that it provides traces of consumer behaviors that were previously costly to observe even on small samples, including consumers' activities, interests, opinions and interactions. To fully capture the information value of these data, more complex models are needed. Those models will support deeper insights and better decisions, while at the same time large volumes of data will support such richer representations of the DGM. However, these models come at greater computational costs.

Many current statistical and econometric models and the estimation methods used in the marketing literature are not designed to efficiently handle large volumes of data. Solutions to this problem involve data reduction, faster algorithms, model simplification and/or computational solutions, which will be discussed below. In order to fully support data-driven marketing decision making, marketing analytics needs to encompass four levels of analysis: (1) descriptive data summarization and visualization for exploratory purposes, (2) diagnostic explanatory models that estimate relationships between variables and allow for hypothesis testing, (3) predictive models that enable forecasts of variables of interest and simulation of the effect of marketing control settings, and (4) prescriptive optimization models that are used to determine optimal levels of marketing control variables. Figure 4 shows that the feasibility of these higher levels of analysis decreases as a function of big data dimensions. It illustrates that the information value of the data grows as its volume, variety and velocity increases, but that the

decision value derived from analytical methods increases at the expense of increased complexity and computational cost of models.

In the realm of *structured data*, where much of the advances in marketing analytics have been so far, all four levels of analysis are encountered. Many of the developments in marketing engineering (Lilien and Rangaswamy 2006) have been in this space, spanning a very wide range of areas of marketing (including pricing, advertising, promotions, sales force, sales management, competition, distribution, marketing mix, branding, segmentation and positioning, new product development, product portfolio, loyalty, and acquisition and retention). Explanatory and predictive models, such as linear and logistic regression and time-series models, have traditionally used standard econometric estimation methods – generalized least squares, method of moments, and Maximum Likelihood. These optimization-based estimation methods become unwieldy for complex models with a large number of parameters. For complex models simulation based likelihood and Bayesian Markov Chain Monte Carlo (MCMC) methods are being used extensively. MCMC is a class of Bayesian estimation methods, the primary objective of which is to characterize the posterior distribution of model parameters. MCMC involves recursively drawing samples of subsets of parameters from their conditional posterior distributions (Gelman et al. 2003). This enables on to fit models that generate deep insight into the underlying phenomenon with the aim to generate predictions that generalize across categories, contexts, and markets. Optimization models have been deployed for, for instance, sales force allocation, optimal pricing, conjoint analysis, optimal product/service design, optimal targeting and marketing mix applications.

The realm of *unstructured data* has seen a growing number of marketing analytics applications. Technological developments in processing unstructured data and the development

of metrics from data summaries– such as provided by text mining, eye-tracking, and pattern recognition software – allow researchers to provide a data structure to facilitate the application of analytical methods. An example of the use of metrics as a gateway to predictive analytics include the application by Netzer et al. (2012), who use text mining on User Generated Content (UGC) to develop competitive market structures. Once a data structure is put in place using metrics, explanatory, prediction and optimization models can be built. Although especially in practice the application of predictive and prescriptive approaches for unstructured data still lags, analysing unstructured data in marketing seems to primarily boil down to transforming it into structured data, using appropriate metrics.

Large-volume structured data comprises four main dimensions – Variables, Attributes, Subjects, and Time (VAST: Naik et al. 2008). The cost of modeling structured data for which one or more of these dimensions is large can be reduced in one of two ways. First, one may reduce one or more of the dimensions of the data through aggregation, sampling or selection, and/or using situation-appropriate simplifications in model specifications. Second, one can increase the speed and capacity of computational resources by using approximations, more efficient algorithms, and high performance computing. Techniques for reducing the dimensionality of data and speeding up computations are often deployed simultaneously, and will be touched upon later.
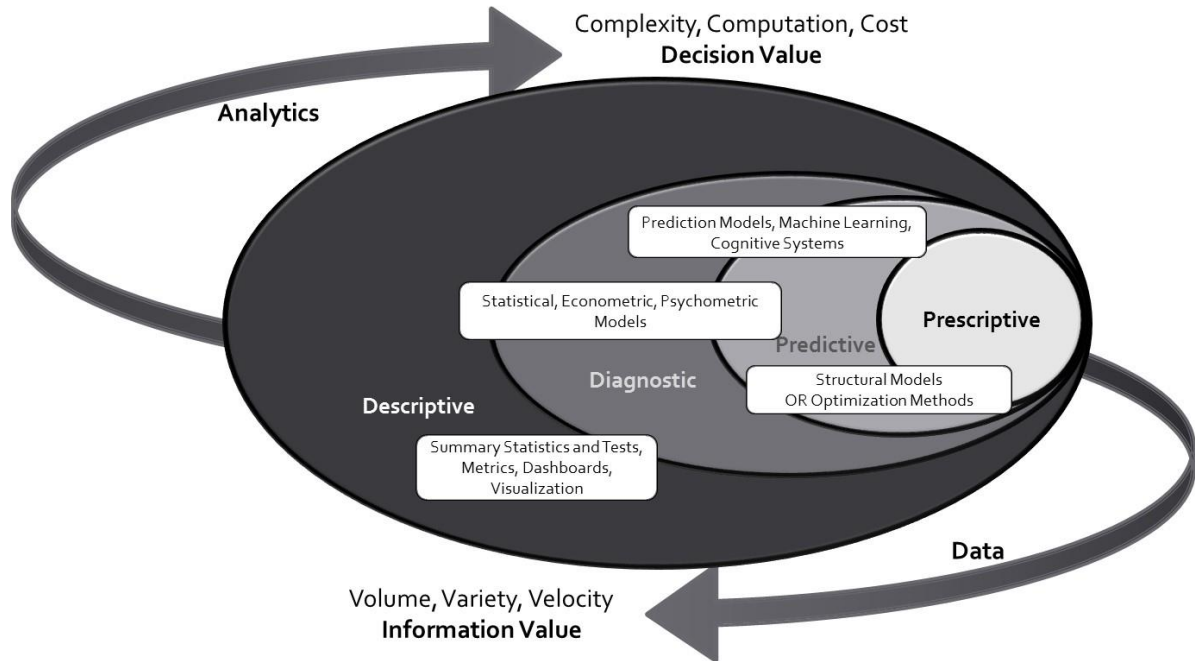
Figure 4: Data and Analytic Approaches
The figure shows the size and degree of structure in marketing data from right to left, and the extent to which analytical methods of increasing complexity are applied to that data from left to right.

3.4.1 Aggregation and Compression

Data volume can be reduced through *aggregation* of one or more of its dimensions, most frequently subjects, variables, or time. This can be done by simple averaging or summing -- which in several cases yields sufficient statistics of model parameters that make processing of the complete data unnecessary--, but also via variable-reduction methods such as Principal Component Analysis (PCA) and related methods, which are common in data mining, speech recognition and image processing. For example, Naik and Tsai (2004) propose a semi-parametric single-factor model that combines sliced inverse regression and isotonic regression. It reduces dimensionality in the analysis of high-dimensional customer transaction databases and is scalable because it avoids iterative solutions of an objective function. Naik, Wedel and Kamakura (2010) extend this to models with multiple factors, which they apply to the analysis of large data on customer churn.

Aggregation of data on different samples of customers (e.g., mobile, social, streaming, geo-demographic) can be accomplished by merging aggregated data along spatial (DMA, zip-code) or time (week, month) dimensions, or through data fusion methods (Kamakura and Wedel 1997; Gilula, McCulloch and Rossi 2006). Data fusion can help in *reducing* data requirements for specific applications through fusing data at different levels of aggregation. For example, if store-level sales data are available from a retailer, these could be fused with in-home scanner panel data. This creates new variables that can increase data veracity because the store data has better market coverage but no competitor information, and vice versa for the home scanning data. Fusion may also be useful when applying structural models of demand that recover individual level heterogeneity from aggregate data (store-level demand), in which case the fusion with individual level data (scanner panel data) can help identify the heterogeneity distribution. Feit et al. (2013) use Bayesian fusion techniques to merge such aggregate data (on customer usage of media over time) with disaggregate data (their individual-level usage at each touch-point) to make inferences about customer-level behavior patterns.

Bayesian approaches can be used in data compression. For example, in processing data that is collected over time, a Bayesian model can be estimated on an initial set of data for the first time period, to determine the posterior distributions for the parameters. Then one only needs to retain these posteriors for future usage, as priors for the parameters of the model calibrated on new data for subsequent time periods. Oravecz, Huentelman, and Vandekerckhove (2015) apply this method in the context of crowd-sourcing. There are several refinements of this general approach. Ridgeway and Madigan (2002) proposed to first perform traditional MCMC on a subset of the data to obtain an initial estimate of the posterior distribution, and then to apply importance sampling/re-sampling to the initial estimates based on the complete data. This

procedure can also be applied as new data comes in over the time. A related technique involves the use of information reweighed priors, which obviates the need to run MCMC chains each time new data comes in. Instead the new data is used to reweight the existing samples from the posterior distribution of the parameters (Wang, Bradlow and George 2014). This approach is related to the particle filter, applied for example by Chung, Rust and Wedel (2009) to reduce the computational burden in processing sequentially incoming data. All these sequential Bayesian updating techniques substantially reduce the computational burden of estimating complex models with MCMC on data of big volume and high velocity because they reweigh (or redraw) the original samples of the parameters from their posterior distributions, with often closed-form weights that are proportional to the likelihood computed from the new data. This class of algorithms thus holds promise for big data because it avoids running MCMC chains on the full data, or for new data that comes in. In addition, parallelizing these algorithms is much easier than standard MCMC because they do not involve iterative computations.

3.4.2 Sampling and Selection

Sampling is mostly applied to subjects, products, or attributes. In many cases big data internal to the company is comprised of the entire population of customers. Using samples of that data enables classical sampling-based inference. Here one has full control over the size, nature and completeness of the sample and multiple samples can be analyzed. Some of the dominant estimation approaches in marketing academia, in particular Maximum Likelihood, are developed within a statistical framework that purports to make inferences from a sample to the population. But, because in many cases big data captures an entire population, statistical inference becomes mute as asymptotic confidence regions degenerate to point-masses under the weight of these massive data (Naik et al. 2008). Traditional statistical inference and hypothesis testing lose their

appeal, because the p-value, the probability of obtaining an effect in repeated samples that is at least as extreme as the effect in the data at hand, becomes meaningless in that case. Unless samples of the data are being analyzed, alternative methods are called for. A problem of using samples rather than the complete data, however, is that this approach may limit the ability to handle long-tail distributions and extreme observations, and is problematic when the modeling focus is on explaining or predicting rare events in the tail of high-dimensional data (see Naik and Tsai 2004). Further, problems with sampling arise when inferences are to be made on social networks. In this case a sampling frame may not be available, and simple random and other standard sampling methods may be inefficient or even detrimental to network properties (snowball, or random forest samples perform better, see Ebbes, Huang and Rangaswamy 2015). More importantly, sampling impedes personalization, for which data on each individual customer is needed, and thus eliminates a major point of leverage of big data (see below).

Bayesian statistical inference offers philosophical advantages in big data applications because inference is conditioned on the data and considers parameters to be random. Inference reflects subjective uncertainty of the researcher about the model and its parameters rather than random variation due to sampling (Berger 1985). This allows one to formulate a probabilistic statement about the underlying truth rather than about the data (e.g., "what is the probability that the null-hypothesis is true?"). However, a limitation of many MCMC algorithms is that they are iterative in nature and, therefore, are computationally intense. Solutions to this computational problem (see discussion above and below) will render comprehensive statistical modeling of big data feasible, which may then be used to drive metrics on dashboards and displays. It is a promising avenue for further development to combine deep insight with user dashboards, as is illustrated by Dew and Ansari (2015), who use semi-parametric prediction of customer base

dynamics on dashboards for computer games. These developments are important given the ubiquitous use of dashboards as the primary basis for decision making in industry, as is the case for example at P&G.

Selection can be used to reduce the dimensionality of big data in terms of variables, attributes, or subjects. Selection of subjects/customers can be used when interest focuses on specific well-defined subpopulations or segments. Even though big data may have a large number of variables ($p> n$ data), they may not all contribute to prediction. Bayesian Additive Regression Tree approaches produce tree structures that may be used to select relevant variables. In the computationally intense Bayesian variable section approach, the key idea is to use a mixture prior to obtain a posterior distribution over all possible subset models. Alternatively, Lasso-type methods can be used, which place a Laplace prior on coefficients (Genkin, Lewis and Madigan 2007). Routines have been developed for the estimation of these approaches using parallel computing (Allenby et al. 2014).

3.4.3 Approximations and Simplifications

A development gainfully employed for big data predictive analytics is the "divide-and-conquer" strategy. Several simpler models are fit to the data, and the results combined. Examples of this strategy include estimation of for example logistic regression, or classification and regression trees on sub-samples of the data, which then are tied-together through bootstrapping, bagging and boosting techniques (Varian 2014). To allow statistical inference in the context of structured big data, variations of this strategy have been used to overcome the disadvantages of using a single random sample. Within a Bayesian framework, analyses of subsamples of big data with a single or multiple models have been combined using meta-analysis

techniques (Bijmolt, Van Heerde and Pieters, 2005; Wang, Bradlow and George 2014), or model averaging methods (Chung, Rust and Wedel 2009).

Another approach to reduce the computational burden of MCMC for big data analytics is to derive analytical approximations to complex posterior distributions in Bayesian models: Bradlow, Hardie and Fader (2002) and Everson and Bradlow (2002) derive closed-form Bayesian inference for models with non-conjugate priors and likelihood, such as the Negative Binomial and Beta-Binomial models, using series expansions. A related technique that uses tractable deterministic approximations to the posterior distribution is Variational Inference (Braun and McAuliffe 2010). Here, the idea is to develop a (quadratic) approximation to the posterior distribution, the mode of which can be derived in closed form. Other work that promises to speed up the computations of MCMC is Scalable Rejection Sampling (Braun and Damien 2015), which relies on tractable stochastic approximations to the posterior distribution (rather than deterministic as in Variational Inference). Taken together, these developments make MCMC estimation of hierarchical models on big data increasingly feasible.

An alternate way to achieve tractability is to simplify the models themselves: one can use simple probability models without predictor variables that allow for closed-form solutions and fast computation. Work by Fader and Hardie (2009) is an example in the realm of CRM to assess lifetime value. More work is needed to support the application of model-free methods (Hastie et al. 2008, Wilson et al. 2010, Goldgar 2001). Model-free methods can reduce computational effort so that big data can be analyzed in real-time, but predictive validation is critical, for example though cross-validation or bagging (Hastie et al. 2008). In the case of unstructured data, the issue is more complex. Deep neural networks (Hinton 2007) provide good prediction results for voice recognition, natural language processing, visual recognition and classification

(especially objects and scenes in images and video), and playing computer games. These approaches, which are neural networks with many hidden layers that can be trained through stochastic gradient descent methods, provide viable approaches to the analysis of unstructured data with much predictive power (Nguyen et al. 2015). Both Facebook and Google have recently invested in their development and application. Marketing models for large-scale unstructured data are still in their infancy, but work is starting to emerge (Netzer et al. 2012; Lee and Bradlow 2011). In this work, the computation of metrics from text, image and video data using image processing methods facilitates the application of standard models for structured data. Examples are Pieters, Wedel and Batra (2010) who use file size of JPEG images as a measure of feature complexity of advertisement images, Landwehr, Labroo and Herman (2011) who apply image morphing to selected design points to compute visual similarity of car images, and Xiao and Ding (2014) who deploy eigenface methods to classify facial features of models in ads.

Relatively little work in the academic marketing literature has addressed deep neural networks and other machine learning methods. This may be because marketing academics favor methods that represent the underlying data generating mechanism and support the determination of marketing control variables, and may shy away from "one solution fits all" methodology and estimation methods, the convergence properties of which cannot be unequivocally established. Nevertheless, future gains can be made if some of these methods can be integrated with the more theory-driven approaches in marketing. This is a fruitful area for future research.

3.4.4. Computation

Many of the statistical end econometric models used in marketing are currently not scalable to big data. MapReduce algorithms (which are at the core of Hadoop) provide a solution, and allow the processing of very large data in a massively parallel way by bringing computation locally to

pieces of the data distributed across multiple cores, rather than copying the data in its entirety for input into analysis software. For example, MapReduce based clustering, naive Bayes classification, singular value decomposition, collaborative filtering, logistic regression, and neural networks have been developed. This framework was initially used by Google, and has been implemented for multi-core, desktop grids and mobile computing environments.

Likelihood maximization is well suited for MapReduce because the log-likelihood consists of a sum across individual log-likelihood terms that can easily be distributed and allow for Map() and Reduce() operations. In this context, Stochastic Gradient Descent (SGD) methods are often used to optimize the log-likelihood. Rather than evaluating the gradient of all terms in the sum, SGD samples a subset of these terms at every step and evaluates their gradient, which greatly economizes computations.

Parallelization of MCMC is also an active area of research, where several promising breakthroughs have been made recently (Tibbitts, Haran and Liechty, 2011, Brockwell and Kadane 2005; Scott et al. 2013, Neiswanger, Wang and Xing 2014). Work also appears underway to combine features of SGD and MCMC. With the continued growth of multi-core computing, formerly computationally prohibitive MCMC algorithms have now become feasible, as is illustrated by their large scale implementation by the analytics company In4mation Insights. Recent advances in parallelization using graphical processing units that promise to speed up likelihood maximization and MCMC sampling (Suchard et al. 2010) are equally promising but outside of the scope of the present exposition.

3.5 Synthesis

Currently only a few academic marketing applications take advantage of really large-scale data, especially rich unstructured data, and tackle the computational challenges that come

with it. Marketing applications favor comprehensive statistical and econometric models that capture the data generating mechanism in detail, but that are often computationally (too) burdensome for big data (Table 1). Solutions to Big Data analytics in the future will use:

1. Developments in high performance computing, including MapReduce frameworks for parallel processing, grid and cloud computing, and computing on graphic cards;

2. Simpler descriptive modeling approaches such as probability models, or computer science and machine learning approaches that facilitate closed form computations, possibly in combination with model averaging and other divide and conquer strategies to reduce bias;

3. Speed improvements in algorithms provided by Variational Inference, Scalable Rejection Sampling, re-sampling and re-weighting, Sequential MCMC, and parallelization of likelihood and MCMC algorithms;

4. Application of aggregation, data fusion, selection, and sampling methods that reduce the dimensionality of data.

Work in practice often deploys a combination of 1 and 2, focusing on exploration and description and generating actionable insights from unstructured data in real time, and can be called *Small Stats on Big Data*. The majority of academic research currently focuses on 3 and 4: rigorous and comprehensive process models that allow for statistical inference on underlying causal behavioral mechanisms and optimal decision making, mostly calibrated on small to moderately sized structured data, and can be called *Big Stats on Small Data*.

Future solutions will likely have an "all of the above" nature (Table 2). One-size-fits all approaches may not be as effective and techniques will need to be mixed and matched to fit the specific properties of the problem in question. Therefore, software for big data management and

processing and high performance computing likely will become an integral part of the ecosystem of marketing analysts in the near future.

4. Analytics and Models

Rich internal and/or external data enables marketing analytics to create value for companies and achieve their short-term and long-term objectives. We define Marketing Analytics as the methods for measuring, analyzing, predicting and managing marketing performance, with the purpose of maximizing effectiveness and ROI. Figure 5 shows how Big Data Marketing Analytics creates increasing diagnostic breath, which is often particularly beneficial for supporting firms' long-term objectives.
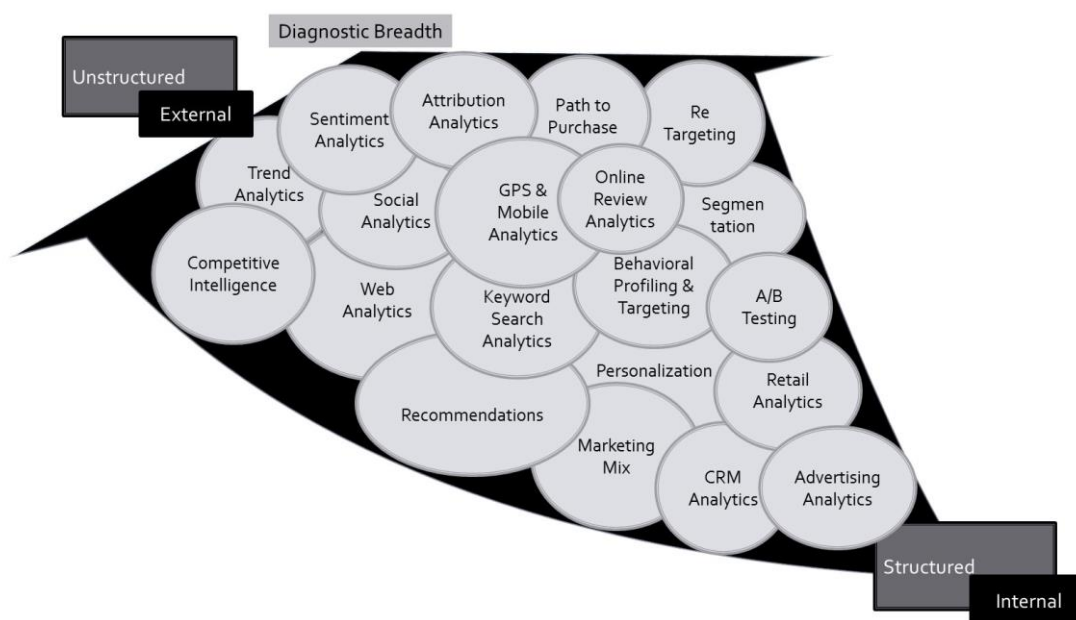
Figure 5: The Diagnostic Breadth of Big Data Marketing Analytics
The arrow shows the increasing breadth of diagnostic insights as a function of utilization of (mostly structured) internal data and (mostly unstructured) external data.

The following examples of recent research (illustrated in Figure 5) take advantage of new digital data sources to develop tailored analytical approaches that yield novel insights.

The *analysis of online reviews* may help a firm to fine tune its offerings and provide

better value to its customers. This was demonstrated by Chevalier and Mayzlin (2006) for online

(book) reviews, which were shown to positively affect book sales. *Keyword search analytics*

may help firms to assess profitability of the design of their websites and placement of their ads.

For example, Yao and Mela (2012) develop a dynamic structural model to explore the interaction

of consumers and advertisers in keyword search. They find that when consumers click more

frequently, the position of the sponsored advertising link has a bigger effect. Further, the study

shows that search tools, such as sorting/filtering based on price and ratings, may lead to

increased platform revenue and consumer welfare. *Analytics for mobile retail data* may help a

firm to provide better recommendations, better target promotions and personalize offerings, and

increase spending by existing customers. Through field experiments with retail stores, Hui et al.

(2013) found that mobile promotions motivate shoppers to travel further inside the store, which

induced greater unplanned spending. *Social analytics* can help firms evaluate and monitor their

brand equity and their competitive positions by identifying trending keywords. For example,

Nam and Kannan (2014) propose measures based on social tagging data and show how they can

be used to track customer-based brand equity and proactively improve brand performance.

Competitive intelligence and trend forecasting can help identify changes in the environment and

set up defenses to retain market share. Along these lines, Du and Kamakura (2012) show how to

spot market trends with Google trends data using factor analytic models. *Click-stream data*

analytics allows for pattern-matching between customer and non-customer behavior, to help

firms identify segments for behavioral targeting. Trusov, Ma, and Jamal (2016) show how to

combine a firm's data with third-party data to improve the recovery of customer profiles. *Mobile*

GPS data analytics provides opportunities to geo-target customers with promotional offers based

on situational contexts. Mobile data allows firms to test the efficacy of their targeting of customers as well as non-customers to increase revenues. Using field experiments, Andrews et al. (2015) show that commuters in crowded subway trains are twice as likely to respond to a mobile offer as commuters in non-crowded subway trains.

These illustrative examples make it easy to understand the importance of big data analytics for supporting marketing decision making in a wide range of areas. The marketing engineering approach, championed by Lilien and Rangaswamy (2006), has contributed to widespread recognition that if the problem drives the choice of models, the superior effectiveness of these models, the quality of the insights they yield, and the consistency of decisions based on them, are all enhanced. After five decades of development, most marketing strategies and tactics now have their own well-specified data and analytical requirements. Academic marketing research has developed methods that specifically tackle issues in areas such as pricing, advertising, promotions, sales force, sales management, competition, distribution, branding, segmentation, positioning, new product development, product portfolio, loyalty, and acquisition and retention. A number of marketing subfields have seen extensive development of analytical methods, so that a cohesive set of models and decision making tools is available, including CRM analytics, web analytics, and advertising-analytics. We next discuss analytics for three closely connected core domains in more detail: marketing/media mix optimization, personalization, and privacy and security.

4.1 Marketing Mix/Media Mix

Models to measure the performance of the firm's marketing mix, forecast its effects and optimize its elements date back to the 1960s. Some of these landmark developments were reviewed in section 2.2 (see Gatignon 1993, Leeflang et al. 2000 and Hanssens et al. 2001,

Hanssens 2014, and Rao 2014 for reviews). As new sources of data become available there are

increased opportunities for better and more detailed causal explanations as well as

recommendations for optimal actions at higher levels of specificity and granularity. This was the

case when scanner data became available (see Wittink et al. 1988), and new sources of digital

data will lead to similar developments. For example, digital data on competitive intelligence and

external trends can be used to understand the drivers of performance under the direct control of

the firm and disentangle them from the external factors such as competition, environmental,

economic and demographic factors and overall market trends. Similarly, field experiments

controlling for the impact of external factors are allowing online and offline retailers to calibrate

the effects of price and promotions on demand for their products and improve forecasts of their

impact (Muller 2014). Here, we focus on developments in marketing mix modeling in the era of

big data, which involve: (1) including information and metrics obtained from new digital data

sources to yield better explanations of the effects of marketing mix elements; (2) attributing

marketing mix effects to new touch points, allocating market resources across classic and new

media, and understanding and forecasting the simultaneous impact of marketing mix elements on

performance metrics; and (3) assessing causal effects of marketing control variables through

structural representations of consumer behavior, instrumental variables, and field experiments.

4.1.1 Incorporating New Data Sources

Research in marketing mix allocation significantly benefits from two specific

developments in data availability. The first is the increased availability of extensive customer-

level data from within firm environments – through direct surveys of customers, measuring

attitudes or satisfaction, or behavior of customers recorded in physical stores and on websites

and mobile apps. Hanssens et al. (2014) take advantage of one source of such data – consumer

mindset metrics – to better model marketing actions' impact on sales performance. They find that combining marketing mix and attitudinal metrics in VAR models improves both the prediction of sales and recommendations for marketing mix allocation. The second development involves using data collected on customers and prospects outside the firm environment, in addition to data that is available within the firm. This may alleviate the problem that activities of (potential) customers with competitors are unobservable in internal data, and may help to fully determine their path to purchase. For example, measures of online WOM (Word of Mouth) (Godes and Mayzlin 2004), online reviews (Chevalier and Mayzlin 2006), or clickstreams (Moe 2003) can be included in marketing mix models to provide better explanations and predictions of consumer choice and sales. Specifically, Moe (2003) uses clickstream data to categorize visits as buying, browsing, or searching visits based on observed navigational patterns, and shows that these different types of visits are associated with different purchase likelihoods. While significant strides have been made, future research should focus on establishing which specific metrics work and which do not, and how they can be best included in models of individual choice, aggregate sales, and market performance.

4.1.2 Attribution and Allocation to New Touch Points

Data from new channels and devices are contributing to the development of new ways in which better marketing mix decisions can be made. For example, while Prins and Verhoef (2007) have examined the synergies between direct marketing and mass communications, Risselada et al. (2014) take advantage of data from customers' social networks to understand the dynamic effects of direct marketing and social influence on the adoption of a high-technology product. Nitzan and Libai (2011) use data on over a million customers' individual social networks to understand how network neighborhoods influence the hazard of defection from a

service provider. Joo et al. (2014) find that television ads impact the number of related searches online, focusing on branded keywords as compared to generic keywords. Similarly, Liaukonyte et al. (2015), using large scale quasi-experimental data of TV advertising and online shopping frequency at two-minute windows, find that television advertising influences online shopping, and that the advertising content plays a key role. These studies highlight the role of cross-media effects in planning the marketing mix. In the context of new devices, Danaher et al. (2015) use panel data to examine the effectiveness of mobile coupon promotions. They find that location and time of delivery of coupons (relative to shopping time) influence redemption. Fong, Fang and Luo (2015) examine the effectiveness of locational targeting of mobile promotions using a randomized field experiment, and investigate targeting at the firm's own location (geo-fencing) versus a competitor's location (geo-conquesting). They find that competitive locational targeting produces increasing returns to the depth of promotional discounts.

The above discussion highlights convergence of different media (TV, Internet, mobile), and the resultant spillovers of marketing mix actions delivered through those media (see also Kannan and Li, 2016). Availability of individual-level path to purchase data – across multiple online channels such as display ads, affiliates, referrals, and search, across devices such as desktop, tablet, and smart-phones, or across online and offline touch-points – will create significant opportunities to understand and predict the impact of marketing actions at a very granular level. For one, it has thrust the attribution problem – assigning credit to each touch-point for the ultimate conversion – to the forefront. Li and Kannan (2014) propose a methodology to tackle that problem. Like marketing mix allocation, attribution involves a marketing resource allocation problem. But even if the attribution problem is completely solved, it is only an intermediate step towards predicting its effects on the entire customer journey, and

towards obtaining an optimal allocation of the entire marketing mix. Many challenges can be expected in this quest. The modeling has to accommodate spillovers across marketing actions, and has to reconcile more granular online and mobile data (e.g. derived from social networks) with more aggregate offline data, and reconcile the different planning cycles for different advertising channels.

In addition to the above, increased options for marketers to influence consumers, such as via firm-generated content in social media and content marketing, where firms become content-creators and publishers, has rendered the issue of understanding the individual effects of these options as part of the marketing mix important. Newer methods and techniques are needed to accurately measure their impact. For example, Johnson et al. (2015) measure the effect of display ads via a new methodology that facilitates identification of the treatment effects of ads in a randomized experiment. They show it to be better than Public Service Announcement (PSA) and Intent-to-Treat A/B tests, in addition to minimizing the costs of tests. Once such individual effects are measured, optimally allocating budgets across marketing/media mix elements becomes possible.

Albers (2012) provides guidelines on how practical decision aids for optimal marketing mix allocation can be developed. He points to the need to study managers' behavior to better determine the specification of supply-side models. One of the important payoffs of working in a data-rich environment lies in the creation of decision aids to better budget and better allocate investments across the marketing market mix, across different products, across market segments, and across customers. Hanssens (2014) provides a review of optimization algorithms that span single-period and multi-periods approaches, and are appropriate for monopolistic and competitive environments. Naik, Raman and Winer (2005) explicitly model the strategic

behavior of a firm that anticipates how competitors will likely make future decisions and reasons backwards to deduce its own optimal decision in response. While most extant work focuses on allocating the budget on single products, Fischer et al. (2011) propose a heuristic approach to solve the dynamic marketing budget allocation problem for multiproduct and multi-segment (countries) firms. This approach, implemented at Bayer's, is an example of a modeling direction that solves pressing practical problems.

4.1.3 Assessing Causality of Marketing Mix Effects

Assessing causality in marketing mix models has received widespread attention in academia, but unfortunately has not yet received as much attention in industry. If a marketing control variable is endogenously determined but not accounted for in the model (e.g., because of missing variables or management actions dependent on sales outcomes), the DGM is not accurately captured. In that case, predictions of the effects of this marketing mix element will be biased (Rossi 2014). This problem may be alleviated if exogenous instrumental variables (IV) that are related to the endogenous control variable can be found. First, the variety in big data might help to find better IVs, which is needed because IVs are often problematic. In the case of TV advertising, Shapiro (2014) exploits discontinuities in advertising spending in local Designated Market Areas. Regression discontinuity designs that exploit variations in a possibly endogenous treatment variable on either side of a threshold are not economical in their data usage and may, therefore, benefit from large data (Hartmann, Nair and Narayanan 2011). But, models with instrumental variables do not generally predict better out of sample (Ebbes, Papies and Van Heerde, 2011). Several instrument-free methods have been developed to help in situations where no valid instruments can be found (Ebbes et al. 2005; Park and Gupta 2012). These methods are suitable for automated application in large scale data-production

environments in industry, where searching for valid instruments on a case-by-case basis is often infeasible. Second, digital data environments allow for field experiments that enable one to assess the causal effects of marketing control variables. The work by Hui et al. (2013) and Andrews et al. (2015) was cited above in this context. Third, in structural modeling of demand and supply, new types of data can help in calibrating the specifications of the models more precisely and efficiently. An illustrative example is provided by Chung, Steenburgh and Sudhir (2014) who estimate a dynamic structural model of sales force response to a bonus-based compensation plan. Rather than assume what the discount factors are that are used by forward looking sales people as is done usually, they estimate them from field data using a combination of exclusion restrictions and a model specific to the institutional setting.

Finally, models that account for forward looking behavior of consumers are important in developing marketing mix models that account for the fact that consumers may maximize their payoff over a finite or infinite horizon, rather than myopically. While the identification of these models benefits from increased variation in data of big volume and variety, they come with computational challenges that still need to be resolved. Liu, Montgomery and Srinivasan (2015) tackle this problem by building a model of consumers' financial planning decisions based on the assumption that they are forward looking and discount future revenues. The researchers estimate their model with parallel MCMC, which allows them to accommodate individual-level heterogeneity and to design targeted marketing strategies. This work is one of the first applications of a structural model on relatively big data and is a promising development because forward looking behavior is important to account for in marketing mix models, even those calibrated on field experiments.

4.2 Personalization

Personalization takes marketing mix allocation one step further in that it adapts the product or service offering and other elements of the marketing mix to the individual users' needs (Khan, Lewis and Singh 2009). There are three main methods of personalization. (1) Pull-personalization provides a personalized service when a customer explicitly requests it. An example is Dell, which allows customers to customize the computer they buy in terms of pre-specified product features. (2) Passive personalization displays personalized information about products or services in response to related customer activities, but the consumer has to act on that information. For example, Catalina Marketing Services, an industry leader of personalized coupons delivered at the check-out counter of brick-and-mortar retail stores, personalizes coupons based on shoppers purchase history recorded via their loyalty cards. Recommendation systems represent another example of this approach. (3) Push-personalization takes this one step further by sending a personalized product or service directly to customers without their explicit request. A good example is Pandora, which creates online or mobile personalized radio stations. The radio stations are individually tailored based on the users' initial music selections and similarities between song attributes extracted from the Music Genome database. For each of these types of personalization there are three possible levels of granularity: (1) mass personalization, in which all consumers get the same offering and/or marketing mix, personalized to their average taste; (2) segment level, in which groups of consumers with homogeneous preferences are identified, and the marketing mix is personalized in the same way for all consumers in one segment; and (3) individual level, in which each consumer receives offerings and/or elements of the marketing mix customized to his/her individual tastes and behaviors. However, the availability of big data with extensive individual level information does not necessarily make it desirable for companies to personalize at the most granular level. Big

data offers firms the opportunity to choose an optimal level of granularity for different elements

of the marketing mix, depending on the existence of economies of scale and ROI. For example, a

firm such as Ford Motor Company develops a global (mass) brand image, personalizes product

and brand advertising to segments of customers, customizes sales effort, prices and promotions

at the individual level, and now personalizes in-car experiences using imaging technology.

4.2.1 Recommendation Systems

Recommendation systems are powerful personalization tools, with best-in-class

applications by Amazon and Netflix. There are two basic types of recommendation engines

based on content filtering or collaborative filtering, but there are also hybrid recommendation

systems that combine features of both types. Content filtering involves digital agents that make

recommendations based on the similarity between a customer's past preferences for products and

services. Collaborative filtering predicts a customer's preferences using those of similar

customers. Model-based systems use statistical methods to predict these preferences; the

marketing literature has predominantly focused on these (Ansari, Essegaier and Kohli 2000).

Research has shown that model-based systems outperform simpler recommendation engines, but

at the cost of a larger computational burden. It has also shown that because many consumers are

unwilling or unable to actively provide product ratings, much of the information in ratings-based

recommendation systems is missing. Adequately dealing with this missing information in

ubiquitous ratings-based recommendation systems can render recommendations much more

effective (Ying, Feinberg and Wedel 2006). In addition, most systems produce

recommendations for consumers based on their predicted preferences or choices, but not

necessarily based on their predicted responsiveness to the recommendations themselves. Taking

receptivity into account in making recommendations by utilizing ideas of response based segmentation can greatly increase their effectiveness (Bodapati 2007).

Conceptually, personalization consists of (a) learning consumer preferences, (b) adapting offerings to consumers, and (c) evaluating the effectiveness of the personalization. Some of the problems with ratings-based recommendation systems have prompted companies to use data obtained unobtrusively from customers as input for online and mobile personalization of services (e.g., Amazon). These three stages have long been used in Closed Loop Marketing (CLM) strategies. In digital environments, CLM can be fully automated in a continuous cycle, which gives rise to Adaptive Personalization Systems.

4.2.2 Adaptive Personalization

Adaptive Personalization Systems take personalization a step further by providing dynamically personalized services in real time (Steckel et al. 2005). For example, Groupon-Now personalizes daily deals for products and services from local or national retailers and delivers them by email or on mobile devices; as it collects more data on the individual subscriber, the deals are more accurately personalized. Another example is the buying and selling of online display ad impressions in Real-Time Bidding (RTB) auctions on ad-exchange platforms. These auctions are run fully automated in the less than one-tenth of a second it takes for a website to load. The winning ad is instantly displayed on the publisher's site. To construct autonomous bidding rules, advertisers (a) track consumers' browsing behavior across websites, (b) selectively expose segments defined on the basis of those behaviors to their online display ads, and (c) record consumers' click-through in response to their ads. This enables ad-placement to be targeted across consumers, time, ad-networks and websites at a very high level of granularity. Pandora's adaptive personalization was mentioned as another example above. Adaptive

Personalization thus takes marketing automation to the next stage. Rather than automating simple marketing decisions, it automates CLM's entire feedback loop. Automation offers the additional benefit of speeding up the personalization cycle dramatically. Adaptive Personalization Systems require minimal proactive user input and are mostly based on observed purchase, usage or clickstream data. They learn consumer tastes adaptively over time by tracking consumers' changing behaviors. From a user's viewpoint, these systems are easy to use: the user only interacts with the service while automatically usage data is recorded and the service adapted. Online and mobile Adaptive Personalization Systems implement fully automated CLM strategies, by collecting and analyzing data, predicting user behavior, personalizing services and evaluating the effectiveness of the recommendations, in a continuous and automated cycle.

Zhang and Krishnamurthi (2004) were among the first to develop an adaptive personalization approach. They personalize online promotional price discounts by using an integrated purchase incidence, quantity and timing model that forecasts consumers' response to promotional effort over time, and employ numerical profit-maximization to adaptively determine the timing and depth of personalized promotions. This application is conceptually similar to Catalina's services in offline stores. In an extension of this work, Zhang and Wedel (2009) investigate the profit implications of adaptive personalization online and offline, comparing three levels of granularity: mass, segment and individual. The results show that individual-level personalization is profitable, but mostly in the online channel. Chung, Rust and Wedel (2009) design and evaluate an adaptive personalization approach for mobile music. Their approach personalizes music using listening data, as well as the music attributes that are used as predictor variables. They develop a scalable real-time particle filtering algorithm (a dynamic MCMC algorithm) for personalization that runs on mobile devices. An element of surprise is brought in

through random recommendations, which prevents the system from zeroing in on a too narrow set of user tastes. The model is unobtrusive to the users and requires no user input other than the user listening to the songs automatically downloaded in the device. Field tests show that this system outperforms alternative algorithms similar to those of Pandora.

Hauser et al. (2009) develop a system for adaptive personalization of website design. They call the approach 'website-morphing': it involves matching the content and look and feel of the website to a fixed number of cognitive styles. The probability of each cognitive style segment is estimated for website visitors, based on initialization data that involves the respondents' click-stream and judgments of alternative web-page morphs. In a second loop the optimal morph-assignment is computed using dynamic programming, maximizing both expected immediate profit and discounted future profit obtained from the user making a purchase on the website. It balances the tradeoff between exploitation, which is presenting product options that best suit users' predicted preferences, and exploration, which is introducing surprise to help improve estimation. Morphing may substantially improve the expected profitability realized at the website. Similar ideas were applied to the morphing of banner ads (Urban et al. 2014), which are automatically placed on websites, matched to consumers based on their probabilities of segment membership to maximize click-through rates.

Adaptive personalization will grow with the advent of the Internet of Things and Natural User Interfaces, through which consumers interact with their digital devices via voice, gaze, facial expression, and motion control. As this data becomes available to marketers at massive scales it enables Automated Attention Analysis, which will potentially benefit marketing mix personalization in numerous ways.

4.3 Privacy and Security

As more customer data is collected and personalization advances, privacy and security have become critical issues for big data analytics in marketing. According to a recent survey (Dupre 2015), more than three quarters of consumers think that online advertisers have more information about them than they are comfortable with, and about half of them believe that websites ignore privacy laws. These perceptions are indicative of two realities. First, firms have been collecting data from multiple sources and fusing them to obtain better profiles of their customers. Easy availability of data from government sources (census, heath, employment, telephone metadata, facilitated by the "Open Data Plan", released by the White House in 2013) and decreasing costs of storing and processing data have led to large ROI on such endeavors (Rust et al. 2002). However, combining datasets has led to what is known as the "mosaic effect", yielding information on consumers that should be private, but is revealed in the integrated data (which is exploited for example by Spokeo). Second, privacy laws and security technology have not kept pace with data collection, storage, and processing technologies. This has resulted in an environment where high-profile security breaches and misuse of private consumer information are prevalent. In the last 10 years, over 5,000 major data breaches have been reported, the majority in the financial industry. According to research by IBM and the Ponemon Institute, the average cost of a data breach approaches $4 million, around $150 per stolen record. Examples of recent high profile data security breaches are those that hit Target, Sony Pictures Entertainment, Home Depot, and Ashley Madison. With cloud storage increasing, data breaches are predicted to become more common.

Two trends are likely to emerge, changing the status-quo. First, governments will increasingly enact strict privacy laws to protect their citizens. This will limit how big data and analytics can be used for marketing purposes. The European Union, which already has stricter

privacy laws, is considering expanding the so-called "right to be forgotten" to any company that

collects personal individual customer data (Dwoskin 2015). Similar but less restrictive laws

could soon be enacted in the US. Goldfarb and Tucker (2009) show that privacy regulation that

restricts the use of personal data may make online display ads less effective, and imposes a cost

especially on younger and smaller online firms that rely on ad revenues (Campbell, Goldfarb and

Tucker 2015). Second, firms are increasingly likely to self-police. Most companies nowadays

communicate privacy policies to their customers. For one, respecting customers' privacy is good

business practice and helps to build relationships with customers. Research by Tucker (2014)

supports this. In a field experiment, she shows that when a website gave consumers more control

over their personal information, the click-through on personalized ads doubled. Johnson et al.

(2015) in comparing the effects of opt-out, opt-in and tracking ban policies on the display ad

industry find that the opt-out policy has the least negative impact on publisher revenues and

advertiser surplus. Increasingly, managers are expected to have a better understanding of new

technologies and protocols to protect data security. In addition, marketing automation (as in for

example Adaptive Personalization) will prevent human intrusion and give customers greater

confidence that their privacy is protected. Importantly, firms will need to ensure that sensitive

customer information is distributed across separated systems, that data is anonymized, and that

access to customers' private information is restricted within the organization. With security

breaches becoming common, there is an emerging view that one cannot completely render one's

systems breach-safe. In addition to taking measures to protect data, firms should have data

breach response plans in place.

The implication of all the above for marketing analytics is that there will be increased

emphasis on data minimization and data anonymization (see also Verhoef, Kooge and Walk

2016). Data minimization requires marketers to limit the type and amount of data they collect and retain, and dispose of the data they no longer need. Data can be rendered anonymous using procedures such as *k*-anonymization (each record is indistinguishable from at least *k*-1 others), removing personally identifiable information, recoding, swapping or randomizing data, or irreversibly encrypting data fields to convert data into a nonhuman readable form. These methods, although protecting privacy, may however not act as a deterrent to data breaches (Miller and Tucker 2011).

Due to data minimization, less individual-level data may become available for analytics development in academic and applied research. More and more data will become available in aggregated form only. Research in marketing analytics needs to develop procedures to accommodate minimized and anonymized data without degrading diagnostic and predictive power, and analytical methods that preserve anonymity. For example, the FTC (Federal Trade Commission) requires data providers such as Experian or Claritas to protect the privacy of individual consumers by aggregating individual-level data at the zip code level. Direct marketers rely on these data, but traditionally ignore the anonymized nature of zip-code level information when developing their targeted marketing campaigns. Steenburg, Ainslie and Engebretson (2003) show how to take advantage of "massively categorical" zip-code data through a hierarchical Bayesian model. The model allows one to combine data from several sources at different levels of aggregation. Further, if the models used for predictive analytics are a-priori known and have associated sufficient statistics or posterior distributions (e.g. means, variances, cross-products), those can be retained rather than the original data to allow for analysis without loss of information. Methods for analyzing aggregate data that accommodate inferences on unobserved consumer heterogeneity may provide solutions in some cases. Missing data

imputation methods can be used to obtain consumer-level insights from aggregate data (e.g., Musalem, Bradlow and Raju 2008), or for data in which fields, variables or records have been suppressed. Imputation methods are also useful when only a portion of customers opt-in for sharing their information, as data augmentation can impute missing data from those customers choosing not to opt-in. Future research in this area needs to focus on how customers' privacy can be protected in the use of rich marketing data while maximizing the utility that can be derived from it by developing models and algorithms that can preserve or ensure consumer privacy.

4.4. Synthesis and Future Research Directions

Ongoing developments in the analytics of big data (see Table 1) involve: (1) the inclusion of data obtained from external digital data sources with offline data to improve explanations and predictions of the effects of the marketing mix; (2) attribution of marketing mix effects through a better understanding the simultaneous impact of marketing mix elements while accommodating their different planning cycles; (3) the characterization of the entire path to purchase across offline and online channels and multiple devices, and dynamic allocation of recourses to individual touch points within that path, (4) assessing causal effects of marketing control variables through structural representations of consumer behavior, instrumental variables, instrument-free methods, and field experiments, and (5) personalization of the marketing mix in fully automated closed loop cycles. Future research should build on this and focus on the following topics and questions (Table 2).

Big Marketing Data:

1. How can the fusion of data generated within the firm with data generated outside the firm take advantage of meta-data on the context of customer interactions? How can this be done in a way that enables real-time analytics and real-time decisions?

2. What new methodologies and technologies will facilitate the integration of "small stats on big data" with "big stats on small data" approaches? What are key trade-offs that need to be made to estimate realistic models that are sufficient approximations?

3. How can field experiments be used to generate big observational data in order to obtain valid estimates of marketing effects, fast enough to enable operational efficiency without holding up marketing processes?

4. How can machine learning methods be combined with econometric methods to facilitate estimation of causal effects from big data at high speeds? What specific conditions determine where in the continuum of machine learning to theory-based models these new methods should be designed?

5. What are viable data analysis strategies and approaches for diagnostic, predictive and prescriptive modeling of large scale unstructured data?

6. How can deep learning and cognitive computing techniques be extended for analyzing and interpreting unstructured marketing data? How can creative elements of the marketing mix be incorporated in predictive and prescriptive techniques?

Marketing Mix:

1. How can granular online, mobile data be aligned with more aggregate offline data to understand the path to purchase and facilitate behavioral targeting? How can meta-data of contexts and unstructured data on creatives be incorporated in the analysis of path to purchase data?

2. How can ROI modeling more accurately identify and quantify the simultaneous financial impact of online and offline marketing activities?

3. What new techniques and methods can accurately measure the synergy, carryover and spillover across media and devices using integrated path to purchase data?

4. How can attribution across media, channels and devices account for strategic behavior of consumers and endogeneity in targeting?

5. How can different planning cycles for different marketing instruments be incorporated in marketing mix optimization models?

Personalization:

1. What content should be personalized, at which level of granularity, and at what frequency? How can content be tailored to individuals using individual level insights and automated campaign management?

2. How can individual-level insights be derived from big data using faster and less computationally expensive techniques to give readings of customers' pulse in real-time?

3. How can firms personalize the mix of touchpoints (across channels, devices and points in the purchase funnel) for each customer in closed loop cycles so that their experience is consistently excellent?

4. What role can cognitive systems, general AI and Automated Attention Analysis systems play in delivering personalized customer experiences?

Security and Privacy:

1. What techniques can be used to reduce the backlash to intrusion, as more personalization increases the chances that it may backfire?

2. What new methodologies need to be developed to give customers more control in personalizing their own experiences and enhance the efficacy of data minimization techniques?

3. How can data, software, and modeling solutions be developed to enhance data security and privacy while maximizing personalized marketing opportunities?

To provide more detailed examples of what such future research may entail, consider point 4 under "Big Marketing Data" above. Both in academic and applied research, unstructured data such as videos, texts and images are used as input for predictive modeling by introducing structure by deriving numerical data – bag of words methods for textual information, tags and descriptors for images and videos, and so on. However, in addition to requiring context-specific dictionaries and supervised classification, none of these techniques quite captures the complete meaning contained in the data. For example, word counts in reviews or blogs ignore dependence between words and the syntax and logical sequence of sentences. Machine learning methods have already been used to detect specific languages and to provide meaningful summaries of text. They can thus be used to provide an interpretation of textual data. The Google cloud

machine learning solutions for computer vision make it possible to interpret the content of images, classifying images into categories, and detecting text and objects, including faces, flowers, animals, houses and logos in images. In addition, the emotional expression of faces in the image can be classified. This means that an interpretation of the image based on meaningful relations between these objects is possible. These methods can be used to analyze and interpret, for example, earned social media content on platforms such as Facebook, Twitter and Instagram. These interpretations of text and images in, for example, online news, product reviews, recommendations, shares, reposts or social media mentions, can be used to understand online conversations around products and services. This can be used to make predictions about their success, provide recommendations to consumers, and customize content of commercial communication on online platforms, including owned social media content, keywords, and targeted advertising (thus touching on points 1 and 2 under "Personalization" above). To accomplish this, researchers need to understand how to include rich interpretations of images and text into predictive models. Exactly how the link between deep learning model output and marketing models can be forged, and what the interpretations are that are produced, and if they render marketing models more effective is the topic of that future research.

As a second example, take point 3, under "Marketing Mix". On the one hand, integrated marketing mix models need to accommodate expenditures on media vehicles within the classes of TV, radio, print, outdoor, owned and paid social media, at granular, spatial and temporal levels, and measure its direct and indirect effects on WOM on earned social media, mindset metrics, and sales, accounting for endogeneity of marketing actions and computing ROI. This requires large-scale models with a time-series structure and multitudes of direct and indirect effects. Such models need to be comprehensive and enable attribution, and the quantification of

carryovers and spillovers across these media classes and vehicles. They need to accommodate different levels of temporal and spatial granularity, and levels of aggregation of customers. Further, current studies on attribution modeling scratch the surface of information available in customer touches of websites, display ads, search ads, and so on, because they code each touchpoint on a single dimension. Each touchpoint can be described with multiple variables. For example, a website has an associated collection of meta data describing the design, content and layout of the website, ad placements, and so on all of which could have impact on customer's behavior when they touch the website. Right now, the marketing literature tackles some of the issues, albeit tackles them in a piecemeal fashion. What is needed is a comprehensive approach to marketing mix and attribution modeling that integrates these various components of the marketing mix and addresses all these issues simultaneously. Data is no longer a limitation for doing so, although data from various sources will need to be combined. Close collaborations between academics and companies are likely needed to ensure availability of data and computational resources. New techniques need to be developed, which might combine such techniques as VAR modeling, HB and choice models, variable selection and reduction and data fusion. Future research needs to investigate which data sources and which models are suitable.

5. Implementing Big Data and Analytics in Organizations.

Organizations use analytics in their decision making in all functional areas: not only marketing and sales, but also supply chain management, finance and human resources. This is exemplified by Walmart, a pioneer in the use of big data analytics for operations, which relies heavily on it in HR and marketing. Integration of data-driven decisions across different functional areas is an aspiration for many companies. While managing Big Data analytics involves technology, organizational structure, as well as skilled and trained analysts, the primary

precondition for its successful implementation in companies is a culture of evidence-based decision making. This culture is best summed up with a quote widely attributed to Edwards W. Deming: "*In God we trust; all others must bring data.*" In such a culture, company executives acknowledge the necessity to organize big data analytics and give data/analytics managers responsibility and authority to utilize resources to store and maintain databases, develop and/or acquire software, build descriptive, predictive, and normative models, and deploy them (Grossman and Siegel 2014). In those successful companies, big data analytics champions are typically found in the boardroom (CFO, CMO), and analytics are used to drive business innovation rather than merely improve operations[3].

In an organization in which analytics is fully centralized, such as is the case for Netflix, initiatives are generated, prioritized, coordinated and overseen in the boardroom. Despite the tremendous success of Netflix, the highly specialized nature of marketing analytics that varies dramatically across marketing domains, frequently demands a decentralized or hybrid infrastructure. This provides the flexibility needed for rapid experimentation and innovation and is more conducive to stimulating a nimble and effective deployment of analytics to a wide variety of marketing problems. A decentralized organization facilitates cross-functional collaboration and co-creation, through communication of analysts and marketing managers in the company. It enables the analysts to identify relevant new data sources and opportunities for analytics, and --in interaction with marketing managers-- allows them to tailor their models and algorithms to the specific demands of marketing problems. However, this organizational structure requires deep distributed analytics capabilities and an emphasis on collaboration.

---

[3] See: Building an Analytics-Based Organization at http://www.atkearney.com/ last accessed January 2016.

AT&T is among the companies that follow this model by hosting data analytics within its business units.

One of the main challenges of decentralization is to achieve a critical mass of analysts that enables continuing development of broad and deep expertise across the organization, and flexible and fast response to emerging issues, without excessive overhead or bureaucracy. Therefore, a hybrid organizational model is often effective. Here, a centralized unit is responsible for information technology (IT), software, and creating and maintaining databases. Marketing analysts can draw upon the expertise of such a central unit when needed. Google takes this approach, where business units make their own decisions but collaborate with a central unit on selected initiatives. In some cases, especially for smaller companies and ones that are at the beginning of the learning curve, outsourcing of one or more of these centralized functions is a viable and cost-effective option.

Taking the role of the centralized unit one step further is an organization that forms an independent big data center of excellence (CoE) within the company, overseen by a chief analytics officer (CAO). The marketing and other units pursue initiatives overseen by the CoE. Amazon and Linkedin employ this organization, for example, and it seems to be the model most widely adopted by big data companies. It provides synergies and economies of scale because it facilitates sharing of data and solutions across business units, and supports and coordinates their initiatives. A problem of managing marketing budgets is the "silo" effect. Often, in large marketing organizations, investments in each of the marketing instruments such as branding, search marketing, e-mail marketing, etc., are managed by different teams with their own budgets. This can lead to each silo trying to optimize its own spending without taking a more global view. With more focus on integrated marketing communications, multi-channel marketing, and

influencing the entire path to purchase, the data analytics function best resides within a central unit or CoE, which prevents the silo-effect by taking a more global view of marketing budgets with direct reporting to the CMO.

Even a decentralized or hybrid analytics infrastructure, however, does not preclude the need for data and analytics governance. Analytics governance functions, residing in centralized units, or CoE's, prioritize opportunities, obtain resources, ensure access to data and software, facilitate the deployment of models, develop necessary expertise, ensure accountability and coordinate team effort. The teams in question comprise of *marketing and management functions* which identify and prioritize opportunities and implement data driven solutions and decisions; *analytics engineers* who determine data, software and analytics needs, organize applications and processes, and document standards and best practices; *data science and data management functions* which ensure that data are accurate, up to date, complete and consistent; and *legal and compliance functions* which oversee data security, privacy policies, and compliance. The CAO may promote the development of repeatable processes and solutions to gain efficiency and economies of scale across decentralized analytics teams.

To summarize, organizations that aim to extract value from big data analytics should (1) have a culture and leaders that recognize the importance of data, analytics and data-driven decision making, (2) a governance structure that prevents silos and facilitates integrating data and analytics into the organization's overall strategy and processes in such a way that value is generated for the company, and (3) have a critical mass of marketing analysts that collectively have both sufficiently deep expertise in analytics as well as substantive marketing knowledge. Almost every company currently faces the challenge to hire the right talent to accomplish this. An ample supply of marketing analysts with a cross-functional skill set, proficient in technology,

data science, analytics and with up to date domain expertise, is urgently needed, as are people with management skills and knowledge of business strategy to put together and lead those teams. We reflect on the implications for business education in the final section of this paper.

6.   Conclusion: Implications for Education

This article has reviewed the history of data and analytics, highlighted recent developments in the key domains of marketing mix, personalization, and privacy and security, and identified potential organizational barriers and opportunities towards successful implementation of analytics of rich marketing data in companies. Table 1 summarizes the state of the art, Table 2 summarizes future research priorities. In this section we round out our discourse with a discussion of the implications for the skill-set required for analysts.

In the emerging big data environment, marketing analysts will be working increasingly at the interface of statistics/econometrics, computer science and marketing. Their skill set will need to be both broad and deep. This poses obvious challenges that are compounded by the fact that various sub-domains of marketing, such as advertising, promotions, product development, and branding have different data and analytics requirements, and one-size-fits-all analytical solutions are neither desirable nor likely to be effective. Analysts therefore need to have sufficiently deep knowledge of marketing modeling techniques for predicting marketing response, marketing mix optimization and personalization. They must be well-versed in the application of estimation techniques such as maximum likelihood methods, Bayesian MCMC techniques, and machine learning methods, and be familiar with optimization techniques from OR. Moreover, they also need to possess soft skills and cutting-edge substantive knowledge in marketing, to ensure that they can communicate to decision makers the capability and limitations of analytical models for specific marketing purposes. This will maximize the support for and

impact of their decision recommendations. In many organizations marketing analysts will fulfill the role of intermediaries between marketing managers and IT personnel, or between marketing managers and outside suppliers of data and analytics capabilities, for which they need to have sufficient knowledge of both areas. Increasingly, routine marketing processes and decisions will become automated. This creates the challenge of how to embed these automated decisions in substantive knowledge and managerial intuition and oversight. Future marketers need to be well equipped to do that. Finally, the field will be in need of people with management skills and knowledge of business strategy, as well as sufficient familiarity with technology and analytics to oversee and manage teams and business units. A recent study by Gartner revealed that business leaders believe that the difficulty of finding talent with these skills is the main barrier towards implementing big data analytics.

These skill-set requirements also present a challenge for educators, and few people will be able to develop deep knowledge in all these areas early in their career. In organizations, these skill sets are most often cultivated through on-the-job training and collective team effort. Some students of analytics may specialize and develop deep expertise in substantive marketing, soft-skills and management, such that they can take up management positions and can oversee analysts, negotiate with outside suppliers of analytics, and help formulate problems and interpret and communicate results. At the other end of the spectrum are those who aspire to be marketing analytics engineers or data scientists, and seek to develop deep knowledge of the technical aspects of the field, including database management, programming and statistical/econometric modeling. Each of those will have a role to play in analytics teams in organizations. All those working in the field will need to continue updating their knowledge across a broad domain, through conferences and trainings, to stay abreast of the tidal wave of new developments.

Companies need to systematically invest in training and education of current employees, and hiring new ones with an up to date skill-set to fill specific niches in their teams. Walmart, for example organizes its own yearly analytics conference with hundreds of participants, and uses crowdsourcing to attract new talent.

The training and education of marketing analysts to develop this broad and deep skill set poses a challenge to academia. In many cases, people directly from programs in mathematics, statistics, econometrics or computer science may not become effective and successful marketing analysts. Instead, next to existing specializations in successful undergraduate and MBA programs at many Universities, recently created masters programs in marketing analytics at places like the Universities of Maryland and Rochester focus on developing these multi-disciplinary skill sets in students who already have a rigorous training in these basic disciplines. Similar programs are urgently needed and being developed elsewhere to meet to the increasing demand for marketing analysts worldwide. Also, our field may need to embraced the model of the Mathematics and Computer Science disciplines to educate Ph.D. students uniquely for industry functions (Sudhir 2016).

Finally, we emphasize opportunities for cross-fertilization of talent from academia and industry – such as practitioners benefiting from specialized classes developed by universities, and academics spending time within companies to get exposure to current problems and data. Such opportunities are becoming increasingly common and will benefit the field significantly in the near future, as big data analytics will continue to challenge and inspire academics and practitioners alike.

Table 1

Marketing Analytics: State-of-the-Art Approaches and Their Applications

Area of Focus	Developments and State-of-the-Art
Data: ■ Structured Data ■ Unstructured Data	a. A plethora of descriptive, diagnostic, predictive and prescriptive methods for analytics in many areas of marketing are available b. Approaches to deal with big data using Bayesian methods, data aggregation and data compression methods; sampling and variable selection methods; approximations and model simplifications; Efficient MCMC algorithms and parallel computing c. Field and Quasi experiments, IV and instrument-free approaches to endogeneity; regression discontinuity approaches d. Mostly descriptive and diagnostic analytical methods, predictive and prescriptive methods still play catch-up e. Text mining and Machine Learning approaches f. Incorporating structure through metrics for text, audio, image, and video data, eye tracking and face recognition and other neuro-data
Marketing Mix Modeling	a. Modeling effects of social networks, keyword search, online WOM, trending, and mobile/location, within the marketing mix b. Analysis of entire path to purchase; attribution modeling c. Incorporating specific institutional settings and contexts to enhance estimation of structural models and their policy simulations; better instrumental variables to address endogeneity; field and quasi experiments for causal effects
Personalization	a. Online and mobile personalization of the marketing mix b. Dealing with missing observations and incorporating receptivity into recommendations c. Adaptive personalization approaches – learning and adapting to users' changes in preferences in a continuous automated cycle
Security and Privacy	a. Research into the effects of privacy and security regulations and policies on consumer behavior and competition between firms b. Models to analyze minimized and anonymized data

Table 2
Marketing Analytics: Issues for Future Research

Area of Focus	Promising and Important Issues for Research
Data ■ Structured Data	a. Behavioral targeting with cross-device data; mobile, location based and social analytics b. Fusing data generated with firms with that generated external to the firm; integrating "small stats on big data" with "big stats on small data" approaches c. Combining machine learning approaches with econometric and theory-based methods for big data applications; computational solutions to marketing models for big data
■ Unstructured Data	d. Development of diagnostic, predictive and prescriptive approaches for analysis of large scale unstructured data e. Approaches to analyze unstructured social, geospatial, mobile data and combining them with structured data in big data contexts f. Using, evaluating and extending deep learning methods and cognitive computing to analyze unstructured marketing data
Marketing Mix Modeling	a. Aligning analysis of disaggregate data with that of aggregate data and including unstructured data in the analysis of the mix b. New techniques and methods to accurately measure impact of marketing instruments and their carryover and spillover across media and devices using integrated path to purchase data c. Dynamic, multi-time period and cross category optimization of the marketing mix d. Approaches to incorporate different planning cycles for different marketing instruments in media mix models
Personalization	a. Automated closed-loop marketing solutions for digital environments; fully automated marketing solutions b. Personalization and customization techniques using cognitive systems, General AI, and Automated Attention Analysis; personalization of content d. Mobile, location-based personalization of the marketing mix
Security and Privacy	a. Methods to produce and handle data minimization and data anonymization in assessing marketing mix effectiveness and personalization b. Distributed data solutions to enhance data security and privacy while maximizing personalized marketing opportunities

References

Adigüzel, Feray and Michel Wedel (2008), "Split Questionnaire Design for Massive Surveys," *Journal of Marketing Research*, 45 (5), 608-617.

Ailawadi, Kusum L., Jie Zhang, Aradhna Krishna, and Michael Kruger (2010), "When Wal-Mart Enters: How Incumbent Retailers React and How this Affects Their Sales Outcomes," *Journal of Marketing Research*, 47 (4), 577-593.

Albers, Sonke (2012), "Optimizable and implementable aggregate response modeling for marketing decision support," *International Journal of Research in Marketing*, 29(2), 111-122.

Allenby, Greg, Eric Bradlow, Edward George, Joh Liechty and Robert McCuloch (2014), "Perspectives on Bayesian Methods and Big Data," *Customer Needs and Solution*, 1 (September), 169-175.

Andrews, Rick L., Andrew Ainslie, Imran S. Currim (2002), "An Empirical Comparison of Logit Choice Models with Discrete Versus Continuous Representations of Heterogeneity," *Journal of Marketing Research*, 39 (4), 479-487.

Andrews, Michelle, Xueming Luo, Zeng Fang, and Anindya Ghose (2015), "Mobile Ad Effectiveness: Hyper-Contextual Targeting with Crowdedness," *Marketing Science*, forthcoming.

Ansari, Asim, Skander Essegaier, and Rajeev Kohli (2000), "Internet Recommendation Systems," *Journal of Marketing Research,* 37(3), 363-375.

——— and Carl F. Mela (2003) "E-customization," *Journal of Marketing Research*, 40 (2), 131-145.

Bartels, Robert (1988), *The History of Marketing Thought. 3d edition*. Columbus: Publishing Horizons.

Bass, Frank (1969), "A new product growth for model consumer durables," *Management Science,* 15 (5), 215–227.

Banko, Michele, and Eric Brill (2001). "Scaling to Very Very Large Corpora for Natural Language Disambiguation," *Proceedings of 39th Annual Meeting of the Association for Computational Linguistics*, pp. 26–33, Toulouse, France. Association for Computational Linguistics.

Berger, James O. (1985), "*Statistical Decision Theory and Bayesian Analysis*," Springer-Verlag, New York.

Berry, Steven, James Levinsohn and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841-890.

Bijmolt, Tammo H.A, Harald J van Heerde, Rik G.M. Pieters (2005), "New Empirical Generalizations on the Determinants of Price Elasticity," *Journal of Marketing Research*, 42, 141-156.

Bodapati, Anand (2008), "Recommender systems with purchase data," *Journal of Marketing Research*, 45 (1), 77-93.

Bradlow, Eric T., Bruce G. S Hardie and Peter S Fader (2002), "Bayesian Inference for the Negative Binomial Distribution via Polynomial Expansions," *Journal of Computational and Graphical Statistics*, 11 (1), 189-201.

Braun, Michael and Paul Damien (2015), "Scalable Rejection Sampling for Bayesian Hierarchical Models", *Marketing Science*, forthcoming.

Braun, Michael and Jon McAuliffe, (2010), "Variational Inference for Large-Scale Models of Discrete Choice," *Journal of the American Statistical Association*, 105(489), 324-335.

Brockwell, Anthony E. and Joseph B. Kadane (2005), "Identification of Regeneration Times in MCMC Simulation, With Application to Adaptive Schemes," *Journal of Computational and Graphical Statistics*, 14 (2), 436-458.

Bart J. Bronnenberg, Jean-Pierre Dubé and Matthew Gentzkow (2012) "The Evolution of Brand Preferences: Evidence from Consumer Migration," the *American Economic Review*, 102 (6), 2472-2508

Bucklin, Randy E. and Catharina Sismeiro (2009), "Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing," *Journal of Interactive Marketing*, 23 (1), 35-48.

Bult, Jan-Roelf and Tom Wansbeek (1995), "Optimal Selection for Direct Mail," *Marketing Science*, 14 (4), 378-394.

Campbell, James, Avi Goldfarb and Catherine Tucker (2015), "Privacy Regulation and Market Structure' *Journal of Economics and Management Strategy*, 24 (1), 47-73.

Casher, Jonathan D. (1969), *Marketing and the Computer*. Boston: D.H. Mark Publications.

Chevalier, Judith A., and Dina Mayzlin (2006), "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research,* 43 (3), 345-354.

Chintagunta, Pradeep (2002), "Investigating Category Pricing Behavior in a Retail Chain," *Journal of Marketing Research*, 39 (2), 141-154.

Chintagunta, Pradeep, Tulin Erdem, Peter E. Rossi, Michel Wedel (2006), "Structural Modeling in Marketing: A Review and Assessment, Marketing Science," 25 (6), 604-616,

Chung, Doug J., Thomas Steenburgh and K. Sudhir (2014), "Do Bonuses Enhance Sales Productivity? A Dynamic Structural Analysis of Bonus-Based Compensation Plans," *Marketing Science*, 33 (2), 165-187.

Chung, Tuck-Siong, Roland T. Rust and Michel Wedel (2009), "My mobile music: An adaptive personalization system for digital audio players," *Marketing Science*, 28(1), 52-68.

Coombs, Clyde (1950), "Psychological scaling without a unit of measurement," *Psychological Review*, 57, 148-158.

Davenport, Thomas H. (2006), "Competing on Analytics," *Harvard Business Review*, January 2006, 99-107.

DeKimpe, Marnik G. and Dominique M. Hanssens (1995), "The Persistence of Marketing Effects on Sales," *Marketing Science*, 14 (1), 1-21.

DeSarbo, Wayne S. and Vithala Rao (1986), "A Constrained Unfolding Methodology for Product Positioning," *Marketing Science*, 5 (1), 1-19.

Dev, Ryan and Asim Ansari (2015), "A Bayesian Semiparametric Framework for Understanding and Predicting Customer Base Dynamics," Working Paper, Columbia University.

Dorfman, Robert and Peter O. Steiner (1954), "Optimal Advertising and Optimal Quality," *American Economic Review*, 44 (December), 826-836.

Du, Rex Yuxing and Wagner A. Kamakura (2012), "Quantitative Trendspotting," *Journal of Marketing Research*, 49(4), 514-536.

Duncan, C.S. (1919), *Commercial Research*, New York: McMillan Co.

Dupre, Elyse (2015), "Privacy and Security Remain Tall Orders for Today's Marketers," *Direct Marketing News* (January 23).

Dwoskin, Elizabeth (2015), "EU Seeks to Tighten Data Privacy Laws," *The Wall Street Journal*, (March 15).

Ebbes, Peter , Huang, Z. and Rangaswamy Arvind (2015), "Sampling Designs for Recovering Local and Global Characteristics of Social Networks," *International Journal of Research in Marketing*, forthcoming.

Ebbes, Peter, Papies, Dominique and Van Heerde Harald (2011). "The sense and non-sense of holdout sample validation in the presence of endogeneity," *Marketing Science*, 30 (6), 1115-1122.

Ebbes, Peter, Michel Wedel, Ulf Böckenholt and Ton Steerneman (2005), "Solving and Testing for Regressor-Error (in)Dependence When no Instrumental Variables are Available: With New Evidence for the Effect of Education on Income", *Quantitative Marketing and Economics*, 3(4) , 365–392.

Elrod, Terry (1988), "Choice Map: Inferring a Product-Market Map from Panel Data," *Marketing Science*, 7 (1), 21-40.

Erdem, Tulin, and Michael P. Keane (1996), "Decision-making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15, 1–21.

Everson, Philip J. and Eric T. Bradlow (2002), "Bayesian Inference for the Beta-Binomial Distribution via Polynomial Expansions," *Journal of Computational and Graphical Statistics*, 11 (1), 202-207.

Fader, Peter and Bruce Hardie (2009), "Probability Models for Customer-Base Analysis," *Journal of Interactive Marketing*, 23 (1), 61-69.

Feit, Eleanor, Wang, P., Pengyuan Wang, Eric T. Bradlow, Peter S. Fader (2013), "Fusing Aggregate & Disaggregate Data with an Application to Multi-Platform Media Consumption," *Journal of Marketing Research*, 50 (3), 348-364.

Ferber, Robert (1949), *Statistical Techniques in Market Research*. New York: McGraw Hill Book Company.

Fischer, Marc, Sönke Albers, Nils Wagner, Monika Frie (2011), "Practice Prize Winner—Dynamic Marketing Budget Allocation across Countries, Products, and Marketing Activities," *Marketing Science,* 30(4), 568-585.

Fong, Nathan M., Zheng Fang, and Xueming Luo (2015), "Geo-Conquesting: Competitive Locational Targeting of Mobile Promotions," *Journal of Marketing Research*, 52(5), 726-735.

Gatignon, Hubert (1993), "Marketing Mix Models," in Handbooks in Operations Research and Management Science, Vol. 5: Marketing, J. Eliashherg and G.L. Liiien, eds. Amsterdam: North, 697-732.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, (2003), *Bayesian Data Analysis*, London: Chapman and Hall.

Geman, Stuart, Elie Bienenstock, and René Doursat (1992), "Neural Networks and the Bias/Variance Dilemma," *Neural Computation,* 4(1): 1-58.

Genkin, A., D.D. Lewis, D. Madigan, 2007. Large-scale Bayesian logistic regression for text categorization. Technometrics, 49, 291-304.

Gilula, Zvi, Robert E. McCulloch and Peter E. Rossi (2006), "A Direct Approach to Data Fusion," *Journal of Marketing Research*, 43(1), 73-83.

Godes, David and Dina Mayzlin (2004), "Using On-Line Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545-560.

Goldfarb, Avi and Catherine Tucker (2011), "Privacy Regulation and Online Advertising," *Management Science*, 57 (1), 57-71.

Goldgar, David E. (2001), "17 Major strengths and weaknesses of model-free methods," *Advances in Genetics*, 42, 241-251.

Green, Paul E. (1963), "Bayesian Decision Theory in Pricing Strategy," *Journal of Marketing*, 27 (1), 5-14.

Green, Paul E. (1969), "Multidimensional Scaling: An Introduction and Comparison of Non-Metric Unfolding Techniques," *Journal of Marketing Research*, 6 (3), 330-341.

Green, Paul E. and V. Srinivasan (1978), "Conjoint Analysis in Consumer Research: Issues and Outlook," *Journal of Consumer Research*, 5 (2), 103-123.

Grossman, Robert L. and Kevin P. Siegel (2014), "Organizational Models for Big Data and Analytics," *Journal of Organization Design*, 3(1), 20-25.

Guadagni, Peter M. and John D. C. Little (1982), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (3), 203-238.

Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (4), 342–355.

Hanssens, Dominique M., L. J. Parsons, and R. L. Schultz (2001), *Market Response Models*, Boston: Kluwer Academic Publishers.

———, Koen H. Pauwels, Shuba Srinivasan, Marc Vanhuele and Gokhan Yildirim (2014), "Consumer Attitude Metrics for Guiding Marketing Mix Decisions," *Marketing Science*, 33 (4), 534-550.

——— (2014), "Econometric Models," in *The History of Marketing Science*, Russ Winer and Scott Neslin (eds.), Singapore: World Scientific Publishing.

Hartmann, Wesley, Harikesh S. Nair and Sridhar Narayanan, (2011), "Identifying Causal Marketing Mix Effects Using a Regression Discontinuity Design," *Marketing Science*, 30 (6), 1079-1097.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2008), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Hauser, John R., Glen L. Urban, Guilherme Liberali and Michael Braun (2009), "Website Morphing," *Marketing Science*, 28 (2), 202-223.

Hinton, Geoffrey (2007), "Learning Multiple Layers of Representation," *Trends in Cognitive Sciences,* 11 (10), 428-434.

Huang, Dongling and Lan Luo (2015), "Consumer Preference Elicitation of Complex Products Using Fuzzy Support Vector Machine Active Learning," *Marketing Science*, forthcoming.

Hui, Sam K., J. Jeffrey Inman, Yanliu Huang and Jacob Suher (2013), "The Effect of In-Store Travel Distance on Unplanned Spending: Applications to Mobile Promotion Strategies," *Journal of Marketing*, 77 (2), 1-16.

Johnson, Garrett A. (2013), "The Impact of Privacy Policy on the Auction Market for Online Display Advertising," *Simon School Working Paper No. FR 13-26*, available at [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2333193].

Johnson, Garrett A., Randall A. Lewis and Elmar I Nubbemeyer (2015), "Ghost Ads: Improving the Economics of Measuring Ad Effectiveness," *Simon Business School Working Paper No. FR 15-21*, available at [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2620078].

Joo, Mingyu, Kenneth C. Wilbur, Bo Cowgill and Yi Zhu (2014), "Television Advertising and Online Search," *Management Science*, 60 (1), 56-73.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (4), 379-390.

——— and Michel Wedel (1995), "Life-Style Segmentation with Tailored Interviews," *Journal of Marketing Research*, 32(3), 308-317.

——— and ——— (1997), "Statistical Data Fusion for Cross-Tabulation," *Journal of Marketing Research*, 34 (4), 485-498.

———, ———, Fernando de Rosa and Jose Afonso Mazzon (2003), "Cross-selling through database marketing: a mixed data factor analyzer for data augmentation and prediction," *International Journal of Research in Marketing*, 20 (1), 45-65.

Kannan, P. K. and Hongshuang (Alice) Li (2016), "Digital Marketing: A Review, Framework and Research Agenda," *Robert H. Smith School of Business Working Paper*, available at [http://www.rhsmith.umd.edu/files/Documents/Departments/Marketing/kan-li-2016.pdf]

Kannan, P.K., and Gordon P. Wright (1991), "Modeling and Testing Structured Markets: A Nested Logit Approach," *Marketing Science*, 10 (1), 58-82.

Khan, Romana, Michael Lewis and Vishal Singh (2009), "Dynamic Customer Management and the Value of One-to-One Marketing," *Marketing Science*, 28(6) 1063-1079.

Landwehr, Jan R., Aparna A. Labroo and Andreas Herrmann (2011): "Gut Liking for the Ordinary: Incorporating Design Fluency Improves Automobile Sales Forecasts", *Marketing Science*, 30 (3), 416 – 429.

Lee, T.Y., and Eric T. Bradlow (2011), "Automatic Construction of Conjoint Attributes and Levels From Online Customer Reviews", *Journal of Marketing Research*, 48 (5), 881-894.

Leeflang, Peter S.H., Dick R. Wittink, Michel Wedel and Alan Bultez (2000), *Building Models for Marketing Decisions*, Boston, Massachusetts: Kluwer Academic Publishers.

Li, Honshuang (Alice) and P. K. Kannan (2014), "Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment," *Journal of Marketing Research*, 51 (1), 40-56.

Liaukonyte, Jura, Thales Teixeira, and Kenneth C. Wilbur (2015), "Television Advertising and Online Shopping," *Marketing Science*, 34(3), 311-330.

Lodish, Leonard M. (1971), "CALLPLAN: An Interactive Salesman's Call Planning System," *Management Science*, 18, B-25 - B-40.

Lilien, Gary L. and Arvind Rangaswamy (2006), *Marketing Engineering*, CreateSpace Independent Publishing Platform.

Little, John D. C. (1970), "Models and Managers: The Concept of a Decision Calculus," *Management Science*, 16 (8), B-466–B-485.

Little, John D.C., and Len M. Lodish (1969), "A Media Planning Calculus," *Operations Research*, 17 (1), 1-35.

Liu, Xiao, Alan Montgomery, and Kannan Srinivasan (2015), "Overhaul Overdraft Fees: Creating Pricing and Product Design Strategies with Big Data," *Working Paper*, Carnegie Mellon University.

Louviére, Jordan J. and George Woodworth (1983), "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data," *Journal of Marketing Research*, 20 (4), 350-367.

Luce, R. Duncan and John W. Tukey (1964), "Simultaneous Conjoint Measurement: A New Scale Type of Fundamental Measurement," *Journal of Mathematical Psychology*, 1 (1), 1–27.

Mantrala, Murali K., Prabhakant Sinha and Andris A. Zoltners (1994), "Structuring a Multiproduct Sales Quota-Bonus Plan for a Heterogeneous Sales Force: A Practical Model-Based Approach," *Marketing Science*, 13(2), 121-144.

Massy, William F., David B. Montgomery and Donald G. Morrison (1970), *Stochastic Models of Buying Behavior*, Massachusetts: The M.I.T. Press.

McFadden, Daniel (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers in Econometrics*, 105-142, New York: Academic Press.

Miller, Amalia and Catherine Tucker (2011), "Encryption and Data Security," *Journal of Policy Analysis and Management*, 30 (3), 534-556.

Mittal, Vikas, Pankaj Kumar, and Michael Tsiros (1999). Attribute-Level Performance, Satisfaction, and Behavioral Intentions Over Time: A Consumption-System Approach, *Journal of Marketing*, 63(2), 88–101.

Moe, Wendy W. (2003), "Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream," *Journal of Consumer Psychology*, 13 (1&2), 29-40.

——— and Peter S. Fader (2004), "Dynamic Conversion Behavior at e-Commerce Sites," *Management Science*, 50 (3), 326-335.

——— and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (3), 444-456.

Montgomery, Alan L., Shibo Li, Kannan Srinivasan and John C. Liechty (2004), "Modeling Online Browsing and Path Analysis Using Clickstream Data," *Marketing Science*, 23 (4), 579-595.

Muller, Frans (2014), "Big Data: Impact and applications in Grocery Retail," Presentation at the 2014 Marketing and Innovation Symposium, Erasmus University, May 27-28, Rotterdam, Netherlands.

Musalem, Andrés, Eric T. Bradlow and Jagmohan S. Raju (2008), "Who's got the coupon? Estimating Consumer Preferences and Coupon Usage from Aggregate Information," *Journal of Marketing Research*, 45 (6), 715-730.

Naik, Prasad A., Michel Wedel, Lynd Bacon, Anand Bodapati, Eric Bradlow, Wagner Kamakura, Jeffrey Kreulen, Peter Lenk, David M. Madigan and Alan Montgomery (2008), "Challenges and Opportunities in High Dimensional Choice Data Analyses," *Marketing Letters*, 19 (3-4), 201-213.

———, Kalyan Raman and Russell S. Winer (2005), "Planning Marketing-Mix Strategies in the Presence of Interaction Effects," *Marketing Science*, 24 (1), 25-34.

———, and Tsai, Chi-Ling (2004), "Isotonic single-index model for high-dimensional database marketing," *Computational Statistics and Data Analysis*, 47(4), 175-190.

———, Wedel, Michel, & Kamakura, Wagner (2010), "Multi-index binary response model for analysis of large data," *Journal of Business and Economic Statistics*, 28 (1), 67-81.

Nakanishi, Masao and Lee G. Cooper (1974), "Parameter Estimation for a Multiplicative Competitive Interaction Model-Least Squares Approach," *Journal of Marketing Research*, 11 (3), 303-311.

Nam, Hyoryung and P. K. Kannan (2014), "Informational Value of Social Tagging Networks," *Journal of Marketing*, 78(4), 21 – 40.

Neiswanger, W., Wang, C., and Xing, E. (2014), "Asymptotically Exact, Embarrassingly Parallel MCMC," Proceedings of the 30th International Conference on Conference on Uncertainty in Artificial Intelligence.

Neslin, Scott A., (2014), "Customer Relationship Management," *The History of Marketing Science*, Russell S. Winer and Scott A. Neslin (Eds.), Hackensack, NJ: World Scientific Publishing Col, Pte. Ltd, 289-318.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg and Fresco Moshe (2012), "Mine Your Own Business: Market Structure Surveillance Through Text Mining," *Marketing Science*, 31(3), 521-543.

Nguyen A, Yosinski J, Clune J. (2015) "Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images," *Computer Vision and Pattern Recognition*, IEEE.

Nitzan, Irit and Barak Libai (2011), "Social Effects on Customer Retention," *Journal of Marketing,* 75(6), 24-38.

Nixon, H. K. (1924), "Attention and Interest in Advertising," *Archives of Psychology*, 72 (1), 5-67.

Oravecz, Z., Huentelman, M., and Vandekerckhove, J. (2015), "Sequential Bayesian updating for big data," In: M. Jones (Eds.), *Big Data in Cognitive Science: From Methods to Insights* (forthcoming). Sussex, UK: Psychology Press (Taylor & Francis).

Park, Sungho and Sachin Gupta (2012), "Handling Endogenous Regressors by Joint Estimation Using Copulas," *Marketing Science*, 31(4), 567–586.

Parsons, Leonard J. and Frank M. Bass, (1971), "Optimal Advertising Expenditure Implications of a Simultaneous-Equation Regression Analysis," *Operations Research*, 19 (3), 822-831.

Pieters, Rik, Michel Wedel and Rajeev Batra (2010), "The Stopping Power of Advertising:

Measures and Effects of Visual Complexity," *Journal of Marketing*, 74(5), 48-60.

Prins, Remco and Peter C. Verhoef (2007), "Marketing Communication Drivers of Adoption Timing of a New E-Service Among Existing Customers," *Journal of Marketing*, 71(2), 169-183

Raiffa, Howard, and Robert Schlaifer (1961), *Applied Statistical Decision Theory*, Boston: Clinton Press, Inc.

Rao, V. R. (2014), "Conjoint Analysis," in *The History of Marketing Science*, Russ Winer and Scott Neslin (eds.), Singapore: World Scientific Publishing.

Reilly, W.J. (1929), *Marketing Investigations*. New York: Ronal Press Company.

Ridgeway, George and David Madigan (2002), "Bayesian analysis of massive datasets via particle filters," in: *Proceedings of KDD-02, The Eighth International Conference on Knowledge Discovery and Data Mining*, 5-13.

Risselada, Hans, Peter C. Verhoef and Tammo H.A. Bijmolt. (2014), "Dynamic Effects of Social Influence and Direct Marketing on the Adoption of High-Technology Products", *Journal of Marketing,* 78(2), 52-68.

Roberts, John, Ujwal Kayande and Stefan Stremersch (2014) "From Academic Research to Marketing Practice: Exploring the Marketing Science Value Chain", *International Journal of Research in Marketing*, 31 (2), 127-140.

Rossi, Peter E. (2014), "Even the Rich Can Make Themselves Poor: a critical examination of the use of IV methods in marketing" (2014), *Marketing Science*, 33 (5), 655-672.

———— and Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22 (3), 304-328.

————, Robert E. McCulloch, and Greg M. Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15 (4), 321-340.

Rust, John (1987), "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55 (5), 999-1033.

Rust, Roland T., Katherine N. Lemon, Valarie A. Zeithaml (2004), "Return on Marketing: Using Customer Equity to focus Marketing Strategy," *Journal of Marketing*, 68 (1), 109-127.

————, P.K. Kannan and Na Peng (2002), "The Customer Economics of Privacy in E-Service,"*Journal of Academy of Marketing Science*, 30(4), 455-464.

Rutz, Oliver J., Michael Trusov and Randolph E. Bucklin, (2011), "Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?" *Marketing Science*, 30 (4), 646-665.

Scott, Steven L., Alexander W. Blocker, Fernando V. Bonassi, Hugh A. Chipman, Edward J. George, Robert E. McCulloch (2014), "Bayes and Big Data: The Consensus Monte Carlo Algorithm," University of Chicago Working Paper.

Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13 (1), 41–67.

Shapiro, B. T. (2014), "Positive Spillovers and Free Riding in Advertising of Pharmaceuticals: The Case of Antidepressants," *Working Paper*, Booth School of Business, University of Chicago.

Shaw, Arch W. (1916), *An Approach to Business Problems*, Cambridge: Harvard University Press.

Shaw, Robert (1987), *Database Marketing*, Gower Publishing Co.

Starch, Daniel (1923), *Principles of Advertising*. Chicago: A.W. Shaw Company.

Steckel, Joel, Russell Winer, Randolph E. Bucklin, Benedict Dellaert, Xavier Drèze, Gerald Häubl, Sandy Jap, John Little, Tom Meyvis, Alan Montgomery, and Arvind Rangaswamy (2005), "Choice in Interactive Environments," *Marketing Letters*, 16 (3/4), 309-320.

Stonborough, Thomas H.W. (1942), "Fixed Panels in Consumer Research," *Journal of Marketing*, 7 (2), 129-138.

Suchard, Marc A., Quanli Wang, Cliburn Chan, Jacob Frelinger, Andrew Cron and Mike West (2010), "Understanding GPU programming for Statistical Computation: Studies in Massively Parallel Massive Mixtures," *Journal of Computational and Graphical Statistics*, 19 (2), 419–438.

Sudhir, K (2016), "The Exploration-Exploitation Tradeoff and Efficiency in Knowledge Production," *Marketing Science*, 35(1), 1-9.

Steenburgh, Thomas, Andrew Ainslie and Peder Hans Engebretson (2003) "Massively Categorical Variables: Revealing the Information in Zip Codes," Marketing Science, 22 (1), 40-57.

Sweeney, Latannya (2002). "K-anonymity: a model for protecting privacy," *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems,* 10(5):557-570, 2002.

Teixeira, Thales, Michel Wedel, Rik Pieters (2010), "Moment-to-Moment Optimal Branding in TV Commercials: Preventing Avoidance by Pulsing," *Marketing Science*, 29 (5), 783-804.

Telpaz, Ariel, Ryan Webb and Dino Levy (2015), "Using EEG to Predict Consumers' Future Choices," *Journal of Marketing Research*, 52(4), 511-529.

Tibbits Matthew, Haran, Murali, Liechty John C. (2011), "Parallel multivariate slice sampling, " *Statistics and Computing*, 21(3): 415–430

Trusov, Michael and Liye Ma (2015), "Crumbs of the Cookie: User Profiling in Customer-Base Analysis and Behavioral Targeting," *Marketing Science*, forthcoming..

Tucker, Catherine (2014), "Social Networks, Personalized Advertising, and Privacy Controls," *Journal of Marketing Research*, 51 (5), 546-562.

Urban, Glen, Gui Liberali, Erin MacDonald, Robert Bordley, and John Hauser (2014) "Ad Morphing," *Marketing Science*, 33(1), 27-46.

Urban, Glen L. and John R. Hauser (2004), "'Listening-In' to Find and Explore New Combinations of Customer Needs," *Journal of Marketing*, 68 (2), 72-87.

Varian, Hal R. (2014), "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 28 (2), 3-27.

Verhoef, Peter C., Kooge, Edwin and Walk, Natasha (2016). *Creating Value with Big Data Analytics: Making Smarter Marketing Decisions*, London: Routledge.

Wang, P., Eric Bradlow, Ed George (2014), "Meta-Analyses Using Information Reweighting: An Application to Online Advertising," *Quantitative Marketing and Economics*, 12, 209 - 233.

Wedel, Michel and Wayne S. DeSarbo (1995), "A Mixture Likelihood Approach for Generalized Linear Models," *Journal of Classification*, 12 (1), 21-55.

——— and Rik Pieters (2000), "Eye Fixations on Advertisements and Memory for Brands: a Model and Findings," *Marketing Science*, 19 (4), 297-312.

White, Percival (1931), *Market Research Technique*. New York: Harper & Brothers.

Wilson, Melanie A., Edwin S. Iversen, Merlise A. Clyde, Scott C. Schmidler and Joellen M. Schildkraut (2010), "Bayesian Model Search and Multilevel Inference for SNP Association Studies," *The Annals of Applied Statistics*, 4 (3), 1342–1364.

Winer, Russell S., and Scott A Neslin (2015), *The History of Marketing Science*. New Jersey: World Scientific Publishing Col, Pte. Ltd.

Wittink, Dick R., Addona, M. J., Hawkes W. J. and Porter, J. C. (1988), "SCAN*PRO: The Estimation, Validation and Use of Promotional Effects Based on Scanner Data," In: Liber Amicorum in honor of Peter S.H. Leeflang, Wierenga, Jaap E., Peter C. Verhoef and Janny C. Hoekstra (eds.). Groningen: Rijksuniversiteit Groningen, 2011, p. 135-162.

Xiao, Li and Min Ding (2014), "Just the Faces: Exploring the Effects of Facial Features in Print Advertising," *Marketing Science*, 33(3), 338-352.

Ying, Yuan Ping, Fred Feinberg and Michel Wedel (2006), "Leveraging Missing Ratings to Improve Online Recommendation Systems," *Journal of Marketing Research*, 43 (3) 355-365.

Zhang, Jie and Lakshman Krishnamurthi (2004), "Customizing Promotions in Online Stores," *Marketing Science*, 23 (4), 561-578.

——— and Michel Wedel (2009), "The Effectiveness of Customized Promotions in Online and Offline Stores," *Journal of Marketing Research*, 46 (2), 190-206.

WEB APPENDIX

Marketing Analytic for Data-Rich Environments

Michel Wedel & P. K. Kannan

Techniques mentioned in the paper with links to Wikipedia pages with an explanation

1. (Additive) regression trees

https://en.wikipedia.org/wiki/Decision_tree_learning

2. Apache Cassandra

https://en.wikipedia.org/wiki/Apache_Cassandra

3. ASCII

https://en.wikipedia.org/wiki/ASCII

4. Bootstrapping

https://en.wikipedia.org/wiki/Bootstrapping_(statistics)

5. Bagging

https://en.wikipedia.org/wiki/Bootstrap_aggregating

6. Boosting

https://en.wikipedia.org/wiki/Boosting_(machine_learning)

7. C++

https://en.wikipedia.org/wiki/C%2B%2B

8. Cognitive systems/computing

https://en.wikipedia.org/wiki/Cognitive_computing

9. Deep Neural networks

https://en.wikipedia.org/wiki/Deep_learning

10. Deep mind

https://en.wikipedia.org/wiki/Google_DeepMind

11. MCMC

https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo

    12. Dremel

https://en.wikipedia.org/wiki/Dremel_(software)

    13. Hadoop

https://en.wikipedia.org/wiki/Apache_Hadoop

    14. Hierarchical Bayes model

https://en.wikipedia.org/wiki/Bayesian_hierarchical_modeling

    15. Hive

https://en.wikipedia.org/wiki/Apache_Hive

    16. Importance sampling/resampling

https://en.wikipedia.org/wiki/Importance_sampling

    17. Internet of Things

https://en.wikipedia.org/wiki/Internet_of_Things

    18. JAVA

https://en.wikipedia.org/wiki/Java_(programming_language)

    19. Lasso

https://en.wikipedia.org/wiki/Lasso_(statistics)

    20. Laplace prior

https://en.wikipedia.org/wiki/Laplace_distribution

    21. Machine learning

https://en.wikipedia.org/wiki/Machine_learning

    22. Maximum Likelihood

https://en.wikipedia.org/wiki/Maximum_likelihood

    23. MapReduce

https://en.wikipedia.org/wiki/MapReduce

    24. Meta-Analysis

https://en.wikipedia.org/wiki/Meta-analysis

    25. MySQL

https://en.wikipedia.org/wiki/MySQL

    26. Naïve Bayes

https://en.wikipedia.org/wiki/Naive_Bayes_classifier

27. Net Promoter Score

https://en.wikipedia.org/wiki/Net_Promoter

28. Neural Network

https://en.wikipedia.org/wiki/Artificial_neural_network

29. NoSQL

https://en.wikipedia.org/wiki/NoSQL

30. Particle filter

https://en.wikipedia.org/wiki/Particle_filter

31. Perl

https://en.wikipedia.org/wiki/Perl

32. Python

https://en.wikipedia.org/wiki/Python_(programming_language)

33. R

https://en.wikipedia.org/wiki/R_(programming_language)

34. Rejection sampling

https://en.wikipedia.org/wiki/Rejection_sampling

35. SQL

https://en.wikipedia.org/wiki/SQL

36. SAS

https://en.wikipedia.org/wiki/SAS_(software)

37. Spark

https://en.wikipedia.org/wiki/Apache_Spark

38. Stochastic Gradient Descent

https://en.wikipedia.org/wiki/Stochastic_gradient_descent

39. Unstructured Data

https://en.wikipedia.org/wiki/Unstructured_data

40. VAR model

https://en.wikipedia.org/wiki/Vector_autoregression

41. Variational Inference

https://en.wikipedia.org/wiki/Variational_Bayesian_methods