Journal of Advertising

ISSN: 0091-3367 (Print) 1557-7805 (Online) Journal homepage: http://www.tandfonline.com/loi/ujoa20

A Primer on Using Behavioral Data for Testing Theories in Advertising Research

Yuping Liu-Thompkins & Edward C. Malthouse

To cite this article: Yuping Liu-Thompkins & Edward C. Malthouse (2017): A Primer on Using Behavioral Data for Testing Theories in Advertising Research, Journal of Advertising, DOI: 10.1080/00913367.2016.1252289

To link to this article: http://dx.doi.org/10.1080/00913367.2016.1252289

Published online: 25 Jan 2017.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=ujoa20

Download by: [Old Dominion University] **Date:** 25 January 2017, At: 10:25

Journal of Advertising, 0(0), 1–13
Copyright © 2017, American Academy of Advertising
ISSN: 0091-3367 print / 1557-7805 online
DOI: 10.1080/00913367.2016.1252289

A Primer on Using Behavioral Data for Testing Theories in Advertising Research

Yuping Liu-Thompkins
Old Dominion University, Norfolk, Virginia, USA

Edward C. Malthouse
Northwestern University, Evanston, Illinois, USA

Interactions with and between customers in digital, social, and mobile environments are commonly recorded, producing behavioral data that have the potential to advance advertising research. This article provides an accessible guide on how to leverage such data for advertising researchers who may have thus far relied mostly on lab experiment or survey data. Specifically, we suggest potential sources for behavioral data and present a process for analyzing and interpreting behavioral data. Each step of the process is discussed: exploring, understanding and preparing data; specifying and estimating models; and interpreting and presenting the results. Some fundamental issues with using multiple regression to analyze such data are covered, including standardization, outliers, transformations, multicollinearity, and the omitted variable bias. We also discuss issues that are especially problematic with using behavioral data in advertising research, including endogeneity, count data, data with many zeros, and grouped data. More advanced versions of regression that address these issues are surveyed, including instrumental variables, propensity scoring, generalized linear models, and mixed models. General advice for thinking about behavioral data is provided.

Consumers are increasingly interacting with advertisers and other consumers in digital environments where every behavior is recorded. Early examples include shopper panel data and TV set-top box data, where every purchase and every TV watching action of a consumer is recorded, and manufacturer and retail advertising activities can be overlaid. The advent of the Internet continued this trend, where every consumer's entire path to purchase—from the initial search for information, exposure to display ads and keyword advertising, to placing an item in a shopping cart and making the final purchase—can all be recorded through clickstream data. Social media further make consumer social interaction directly observable, and more recently mobile devices have added location data into the mix of observable advertising interactions. These interactions leave a trail of rich information about consumers' behaviors that can be used in advertising research to develop, refine, and test theories.

The goal of this article is to provide an introduction to the methodological and statistical issues involved in analyzing behavioral data for causal research. We do not intend to cover the advanced, untackled statistical challenges associated with using behavioral data. Instead, our target audience consists of advertising researchers who are familiar with linear regression but may have relied mostly on surveys and lab experiments to answer their research questions. We introduce these researchers to a complementary approach to research through behavioral data sources and alert them to the issues in using such data. Sample issues covered include data understanding and aggregation, determining the need for standardization and for normal distribution, handling outliers and missing values, dealing with endogeneity, count data analysis, handling group-structured data, validating and visualizing results, and so on. We identify the methods and techniques from econometrics, marketing, and statistics that are most likely to be useful in analyzing behavioral data for advertising research. Before we get into the specific issues, we first discuss the process of doing advertising research and how behavioral data can play a useful complementary role.

The authors contributed equally to the project, and the names are listed in alphabetical order.

Address correspondence to Edward C. Malthouse, Northwestern University, Industrial Engineering and Management Services, 1870 Campus Drive, Evanston, IL 60208. E-mail: ecm@northwestern.edu

Yuping Liu-Thompkins (PhD, Rutgers University) is a professor and the chair of marketing, Strome College of Business, Old Dominion University.

Edward C. Malthouse (PhD, Northwestern University) is the Theodore R. and Annie Laurie Sills professor of integrated marketing communications, professor of Industrial Engineering and Management Sciences, and research director, Spiegel Center on Digital and Database Marketing, Northwestern University.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ujoa.

PROCESS OF DOING ADVERTISING RESEARCH AND THE ROLE OF BEHAVIORAL DATA

Most existing advertising research begins with a problem, question, or hypothesis (e.g., Churchill and Iacobucci 2005, Figure 3.1).[1] Once the problem or hypothesis is formulated, the researchers then choose an appropriate research design and collect data to test the hypothesis. Empirical tests of hypotheses commonly culminate in some type of statistical hypothesis test, from which implications are drawn. In an analysis of articles published in advertising journals from 1980 to 2010, Kim et al. (2014) conclude that the most commonly used empirical methods in advertising journals are lab experiments and surveys. While lab experiments excel in internal validity and surveys are a relatively inexpensive way of capturing consumer opinions, these methods often rely heavily on self-reported measures of consumer responses, such as attitude and purchase intention.

There has been a long history of introspection about the use of self-reported measures and the traditional research process. For example, Peter's (1983) classic critique of the consumer research process points out that lab experiments often eliminate the "complex dynamic process of human behavior" and that there is a general lack of studying overt behavior as a dependent variable. He summarizes issues in establishing construct validity and concludes that "many of the validity problems which currently plague consumer research may be reduced by investigating overt behavior . . . it is clear that at least baseline data on consumer overt behavior are needed" (p. 388).

Although this reference was published more than 30 years ago and was focused on consumer research, the same issues exist today and carry over to advertising research. A large body of the advertising literature has been built on "laboratory experiments" using student or online consumer panel samples. This situation is understandable, because access to natural behavior has often been cumbersome and cost-prohibitive. However, with increasing trackability of ad exposure and product purchase behavior, as mentioned earlier, there are now great opportunities to investigate consumers' real-world reactions to advertising in areas such as search engine advertising, mobile advertising, long-term advertising effects, and cross-platform advertising issues.

Although behavioral data bring opportunities to advertising research, they also pose unique challenges and issues that are of minimal concern to lab experiment and survey data. For example, such behavioral data are often not collected directly by the researcher. Therefore, they may be "messy" and require extensive processing before they can be useful. As another example, the trackability of behavior often does not carry over to every domain and hence may cause extensive issues with missing data or omitted variables. This article aims to address some of these issues so that advertising researchers thinking about exploring behavioral data can be better informed in the choices they will be making. As an overall guideline, we recommend that researchers follow the process portrayed in Figure 1 to leverage behavioral data. With the exception of the research question and hypothesis formation step, which is mostly theory rather than methodology driven, we organize the rest of this article based on the flow of this research process.

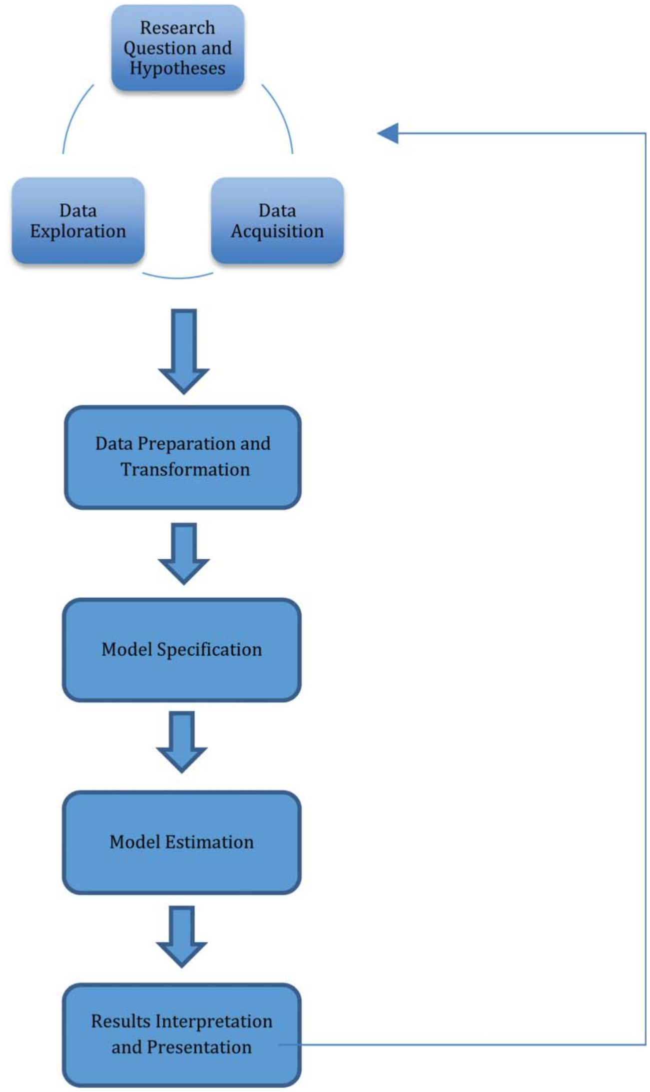

FIG. 1. Recommended workflow for modeling behavioral data.

DATA ACQUISITION, EXPLORATION, AND PREPARATION

Data Acquisition: Key Data Sources

Behavioral data relevant for advertising research can come from a variety of sources, such as the advertiser, public online spaces, and consumers themselves. The most common types of data include (1) transactions and marketing communication records advertisers track of their customers and campaigns; (2) website clickstream data; (3) consumer search data; (4) online social network data (network structure, social content, and information diffusion); (5) online word-of-mouth data (quantitative and qualitative); (6) other web-scrapable data[2] obtained from public websites, such as product and sales rank information from Amazon.com and social commerce data from Groupon and Living Social; and (7) data from mobile and wearable devices, such as location and health indicators. While the best way to obtain behavioral data may be to directly partner with advertisers or their service providers, some third-party organizations can facilitate such relationships or even provide behavioral data sets directly to researchers. Table 1 lists a few such sources primarily targeted toward academic research as a starting point for researchers.

Data Understanding

As previously noted, behavioral data are typically not collected directly by the researcher. Hence it is necessary to devote a substantial amount of time to understanding the data structure. Experienced data providers might supply a data dictionary with the data set, specifying the tables in the database and the columns in each table. This dictionary and conversations with the data provider are key to understanding the data. In examining and understanding the data, researchers should be especially aware of the following issues.

Time reference for each data field. The age field in the customers' table could have been captured at the time a customer first registered on the website and therefore would be outdated and not comparable across customers. If this is true, it will be necessary to merge information about age and the date of registration to derive the consumer's birth year to calculate an accurate current age. In other situations, outdated data, such as number of children and marriage status, cannot be easily derived from other data fields and need to be used with caution. In general, it is always worthwhile to talk with data managers about data-updating practices, which are often lax.

Time length of aggregated data fields. When a data field already represents some aggregation, it is important to know what the time period is for the aggregation. This can be an issue especially during the beginning and ending time periods, such as when a consumer first purchased from a firm or quit a firm. If a typical time period is, say, one month, the first month and the last month for a consumer may not represent a complete month, depending on the joining or quitting time. Consequently, any total amounts or transactions for those months are not comparable to the other complete months in between.

Computed fields. As the age example earlier suggests, it is often possible to compute values of a data field from other data fields in the same data set. For example, exposure to a specific TV commercial can be derived by matching the airing times of the commercial with the time periods and channel choices in a consumer's TV watching behavior. With some creativity, researchers can often construct useful new or seemingly missing variables from existing data. Creating such derived variables is a critical step in using behavioral data, and its importance should not be underappreciated. From time to time, some computed data fields are provided directly to researchers. In such situations, it is important to understand how the computation was done. For example, does the reported number of pages viewed during a visit include only unique pages or does it include duplicate pages too? Does the number of clicks for an ad exclude robot activities and/or multiple clicks from the same user?

Meaning of missing data. In some situations, missing data in a behavioral data set means that the data could have been captured but were not, in the true sense of missing data. However, in other situations, missing data means a default value of 0 (e.g., no pageview occurred). Hence, when null values are present, it is important to know which is the case for a null value and to adjust the data accordingly. Similarly, a *unary variable* is one that takes only a single value "yes" versus unknown. For example, a video website might know which of its customers watched some movie, but a customer who did not watch it on the website may either have watched it elsewhere or not at all.

Besides understanding behavioral data, researchers also need to devote substantial effort to data cleaning. The need for and approach to data cleaning can vary widely depending on the rigorousness of the original data collection and updating practices. We will not elaborate because each data set is unique in this perspective. Readers can refer to in-depth books on data cleaning to learn more (e.g., Osborne 2013). But it suffices to say that behavioral data often contain a lot of dirt and noise, and are rarely suitable for direct analysis without substantial cleaning and understanding. In our experience, it is common to spend a large amount (often the majority) of analysis time simply preparing the data.

Data Merging

Behavioral data often involve actions in multiple domains by the same consumer or household (e.g., TV tuning data combined with purchase data). This has been called single-source data in advertising research. With the proliferation of new media channels and tracking mechanisms, such single-source data now have much richer information that can span search behavior, display ad exposures and clicks, website visit actions, mobile actions, social media conversations, and online and offline purchase behavior. Such data offer the possibility of much more comprehensive studies of digital, social, and mobile advertising effectiveness, such as cross-platform effects.

From a tactical standpoint, unlike a single data file generated from a lab experiment or survey, behavioral data such as website navigation records or social media interactions typically come as multiple tables embedded within a relational database structure. For example, there may be one table listing the actions that occurred during each visit, another table showing each customer's profile, and a third table with the referring ad's information. These multiple tables will need to be joined into a single flat file to create the right input data for model estimation. The most common tools for doing this include SQL, Python, SPSS aggregate/merge, SAS proc summary/data step, and R plyr. These procedures use one or more "key" columns (e.g., consumer ID) to merge information across tables.

Data Aggregation

Frequently it is also necessary or desirable to aggregate the raw data. By aggregating we mean computing a summary across records for a single customer or over a time period, such as counting, summing, averaging, or finding the minimum. For instance, the researcher may be more interested in a consumer's total time spent on a website each week rather than how long each visit is. In this case, visits within the same week will need to be "rolled up" into one weekly observation. Important decisions here include the specific measure used for aggregation, the various ways that aggregation can be done (e.g., within individual across time, or across individuals within time), and the appropriate aggregation level. For example, one commonly used set of aggregation measures comes from the recency-frequency-monetary value (RFM) model, which was recently extended to also include the "clumpiness" of consumer purchases (see Zhang, Bradlow, and Small 2015). These aggregate measures can be applied to many different contexts by modifying the aggregation levels and entities. For example, instead of counting previous orders, a company can

TABLE 1
Some Third-Party Behavioral Data Sources

Data Source	Cost	Description	Link
Marketing Science Online Databases	Some free	The journal *Marketing Science* has been publishing data sets for open access to researchers. Out of the four data sets published so far, the first, from IRI, contains consumer transaction information as well as advertising information that can be used to answer advertising-related research questions.	http://pubsonline.informs.org/page/mksc/online-databases
Marketing EDGE Data Set Library	Free for Marketing EDGE members	Marketing EDGE is formerly the Direct Marketing Educational Foundation. Given the focus of the organization, most of the data sets pertain to advertising through direct mail and catalogs for both businesses and nonprofit organizations.	http://www.marketingedge.org/marketing-programs/data-set-library
Wharton Customer Analytics Initiative Research Opportunities	Free but involves competitive process	The Wharton Customer Analytics Initiative from the University of Pennsylvania offers competitive research opportunities that allow researchers to submit their research proposals year-round. Some of these research projects are directly related to advertising, such as one in 2015 that explored the relationship between online display ads and customer conversion. Once a proposal is accepted into a research opportunity, the research team is given free access to the data set and is asked to report its research findings to the corporate data sponsor after one year.	http://wcai.wharton.upenn.edu/for-researchers/research-opportunities/
University of Chicago Kilts Center for Marketing Datasets	Some free	The University of Chicago Kilts Center for Marketing offers a free ERIM database that has both TV viewing data and purchase data for a sample of households. The center offers a few other free and paid data sets that contain consumer transactions and retail promotion information only.	https://research.chicagobooth.edu/kilts/marketing-databases

(*Continued on next page*)

TABLE 1
Some Third-Party Behavioral Data Sources (*Continued*)

Data Source	Cost	Description	Link
Kaggle Datasets	Free	Kaggle Inc. hosts data science competitions in a wide variety of disciplines, some of which are related to marketing and advertising domains. Researchers can both upload and download data sets for free from the website.	https://www.kaggle.com/competitions
Yahoo! Research; Google Research; Microsoft Research	Free	Yahoo! Research provides a few advertising and marketing data sets, primarily on search engine advertising and web navigation behavior within the Yahoo! ecosystem. Researchers can request access by providing a short description of their research project. Google Research and Microsoft Research also offer public data sets to academic researchers, but their focus is toward data mining and machine learning.	Yahoo! Research: http://webscope.sandbox.yahoo.com/catalog.php?datatype = a Google Research: https://research.google.com/research-outreach.html#/research-outreach/research-datasets Microsoft Research: https://www.microsoft.com/en-us/research/academic-program/data-science-at-microsoft-research/
Marketing Science Institute Data Access	N/A	The Marketing Science Institute offers research support on its member companies' high-priority issues. Besides financial support, the organization can connect researchers to corporate partners who may have the data that researchers need for their projects. This is done through a standard proposal submission process.	http://msi.org/research/obtain-research-support/

also count visits to a website, number of viewing occasions, and so on. Likewise, instead of thinking of monetary value in terms of a dollar amount, one could also think of time, for example, the length of time on the company's website or the length of time viewing a particular video. It is also useful to compute RFM measures by category, for example, a supermarket may find it useful to know the number of purchases and total amount spent on organic products, ready-to-eat-products, baking supplies, and so on.

Researchers need to recognize the statistical issues associated with aggregate data. For example, relationships between variables based on aggregate data may not always hold at the disaggregate level (ecological correlations; see Clancy, Berger, and Magliozzi 2003; Robinson 1950). A correlation computed on averages or rates will overstate the strength of the correlation (Freedman, Pisani, and Purves 1998, pp. 147–51). At the same time, although efficient models and techniques for handling large behavioral data sets have made data aggregation less of a necessity, it is not always of value to analyze data at the most disaggregate level. Excellent references exist on this topic, such as Leeflang et al. (2000, ch. 14) and Malthouse (2013, sec. 5.3 and 6.1). Researchers should balance detail and efficiency while taking into consideration consumer and firm decision-making processes. Finally, when data are captured from different sources, researchers need to be aware of the possibility that not all sources are able to provide the most disaggregate data at the individual/household level. Researchers have started to address methodological challenges in integrating such data (Feit et al. 2013).

DATA DISTRIBUTION AND TRANSFORMATION

Once the data are understood, aggregated at the suitable levels, and joined, researchers still need to consider whether some of the data should be transformed before actual analysis. In this section we look at three commonly encountered issues in this step: the need for normal distributions, missing values, and outliers.

Do We Need Normal Distributions?

In answering this question, it is necessary to recognize that ordinary least squares (OLS) estimates do not require normal errors or normally distributed predictor variables. Furthermore, the central limit theorem (CLT) implies asymptotic normal distributions for large samples, which is often the case with behavioral data. Hence *t*-test and confidence interval construction are possible without normally distributed errors.

It might appear, then, that normality is not important for behavioral data. However, with the right-skewed data that advertising researchers frequently encounter, such as the number of ad exposures and the length of site visit duration, it is still often helpful to compute the log of a variable to make it more normal. Why? There are several reasons for predictor variables. First, log transformation can account for diminishing returns. This could be the case for ad expenditures, where the return on an extra advertising dollar may become smaller as total expenditure gets larger. Logging advertising expenditure can reflect that thought. Second, when the distribution of x is highly right skewed, a log transformation will reduce the influence of outliers. It will also symmetrize the distribution of x; without the log there will be little data in the right tail, producing estimates with wide confidence bands, while using log data will produce more uniform confidence bands across the range of x.

When it comes to transforming the dependent variable another issue is relevant: The variance of count and amount variables typically increases with their means, violating the assumption of constant variance. It is a good practice to take a log or square root as a variance-stabilizing transformation. Of course, predictions and tests are then done on the log units, rather than on the original data.

When to Standardize Data Before Analysis?

Behavioral data often measure multiple things on different scales and hence standardization may be necessary in many cases. Standardization usually involves transforming variables with $z = (x - \mu)/\sigma$, so that they have a mean of 0 and a standard deviation of 1. The units for x, μ, and σ cancel out, making Z scores "unitless." We also say that the unit of measure is the standard deviation. Variables are often standardized in this way before other analyses, such as regression, cluster analysis, and principal component analysis. Variables should usually be standardized whenever the units of the original variables are incommensurate, for example, one variable is measured in minutes and another in dollars. This is a simple way to find comparable units so that effects can be interpreted in relative terms.

However, when the original units are the same, such as all variables being on 7-point scales or all measured in dollars, standardization can obscure structure in the data. Standardizing involves dividing by σ, so the resulting range will be larger for variables with smaller σ than for those with larger σ. This implicitly increases the weight of variables with low variation. For example, consider clustering customers based on how much time they spent each month watching cable TV, browsing the Internet, and talking or texting on mobile devices. The range for how much people watch TV may be 0 to 150 hours, while web-browsing time may range from 0 to 50 hours. Standardizing would inflate the importance of online browsing and reduce the importance of TV, which is probably undesirable. Note that standardization does not affect the skewness of a variable, and, as discussed, one may want to log transform highly right-skewed variables prior to standardization.

One should not standardize variables before regression without careful consideration. First, as a rule, one should not standardize dummy variables. Let x_j be a 0–1 dummy variable where $\bar{x}_j$ is the fraction of 1 values. The standard deviation of x_j is $\sqrt{\bar{x}_j(1-\bar{x}_j)}$, which has a maximum value when $\bar{x}_j = .5$. Standardizing x_j would inflate the importance of variables with $\bar{x}_j$ close to 0 or 1, and reduce the importance of variables where approximately half the cases are 1. Second, in designed experiments the levels of x_j, and thus the standard deviation of x_j, are determined by the person designing the experiment, and so standardization should not be done. Third, in situations where a common unit is available, standardization should not be used. For example, suppose that the sales variable is regressed on the number of TV ads shown during a period of time and the number of sales representatives. Notice that the independent variables have different units, but a natural common unit would be cost in dollars. Standardization would imply that the magnitudes of the slopes would depend on the standard deviations of the number of ads and reps across geographic units.

How to Deal With Outliers?

Outliers are often present in behavioral data. To handle them, the first step is to understand why the values are outliers. There are two possible reasons: (1) the value is erroneous or (2) it is a correct but extreme value. For values that are mistakes, ideally the values should be corrected; but if not, they should be removed. For example, a search engine robot's visit to a frequently updated website can result in an unusually large number of site visits, pageviews, and ad impressions. Such "click" records should be identified and removed.

The more difficult situation is when the values are correct but extreme (e.g., a billionaire's income). The problem with such cases is that they can exert a strong influence on statistical estimates, for example, mixing one billionaire in with a sample of middle-class people causes everyone to be a millionaire, on the average. There are three possible actions in this case:

1. Form segments, and analyze billionaires versus everyone else or business customers versus individual consumers separately (see Malthouse 2013, ch. 2, for more information about segmentation approaches).
2. Transform the data by, for example, computing the log or square root. A disadvantage of this approach is that the results will be in the transformed units.
3. Use robust statistics such as a median or trimmed mean, which is the mean ignoring a certain percentage of observations at the extreme ends of the distribution (e.g., when the highest and lowest bidding prices for an ad keyword are dropped). An advantage is that they maintain the original units.

How to Handle Missing Values?

Missing values are common with behavioral data, and the values are usually not missing at random, where records with missing values are systematically different from those with values. For example, companies often obtain demographic information from third-party data providers by matching their house file with names, addresses, and so on, with the data providers' databases of demographics. But they would not be able to match all customers, producing missing values of, for example, age. An important question is why they have, or don't have, age information about a customer. It turns out that they were more likely to have information about individuals who were active "mail-order buyers." Individuals for whom they could not provide a match were less likely to make purchases, and it was not uncommon for whether a match could be made to be more predictive than the demographic information itself. In such cases, a good modeling strategy is to let missing be a separate dummy variable to model its effect on y. For example, let dummy agemiss equal 1 if age is missing and 0 if age is populated. Now set all missing values of age to 0, and regress y on both agemiss and age: $y = b_0 + b_1$ age $+ b_2$ agemiss. Notice that when age is missing, the equation reduces to $y = b_0 + b_2$, which is estimated with least squares to be the mean of y for cases where age is missing. When age is present the equation is $y = b_0 + b_1$ age, which is the regression of y on age estimated with cases where age is available.

MODEL SPECIFICATION

Model specification involves selecting predictor variables and their functional form to include in a model, such as a linear or logistic regression, structural equations model, and the like. Functional form means selecting transformations to allow for, for example, diminishing returns, inverted-*U*s, interactions, and so on. Additional issues arise when the data are grouped or the dependent variable is not normal, which is discussed later in this section. If the only objective is to make accurate predictions, such as predicting the likelihood that prospects will respond to an offer, there are many suitable statistical learning methods (e.g., see James et al. 2013). Such methods are also useful in exploratory analyses to understand what variables relate to an outcome. The findings from such explorations can inspire subsequent studies with causal research designs.

But advertising scholars have often been more interested in making causal inferences (Kim et al. 2014),[3] where the objective is to test whether some predictor variable or variables affect(s) the dependent variable based on theory, than in only making predictions, and model selection in these instances is much more complicated. When the objective is to make causal inferences, a good research design is usually a randomized, controlled experiment, where the causal factor(s) can be manipulated by the researcher to understand its (their) effect (s) on the criterion. In behavioral data advertising research, however, it is probably more common to have nonexperimental records of customer behaviors over time along with other static data sets, such as demographics from a third-party data provider. This section focuses on such situations by discussing omitted variables, multicollinearity, and endogeneity.

Omitted Variable Bias

Suppose that there is a single predictor variable x and a researcher estimates the model $y = \alpha + \beta x + e$, but the true model is $y = \alpha + \beta x + \gamma w + e$. In other words, variable w also affects y, but the research did not include it in the model. The omitted variable bias theorem states that the estimates of β will be biased, with the direction and magnitude of the bias depending on the correlation between x and w, and the sign of γ: $E(\gamma) = \beta + \gamma r_{xw} S_w / S_x$. It is easy to imagine behavioral data situations in advertising where this is an issue. For example, suppose that x is the level of social buzz across various products and y is purchase. Regressing purchase on buzz volume would seem to quantify the relationship between the two, but there could be an omitted variable, such as $w =$ mass advertising spend for the particular product. Suppose that mass advertising has a positive effect on purchase (i.e., $\gamma > 0$) and that brands with higher advertising also stimulate more social conversations (i.e., $r_{xw} > 0$). The omitted variable theorem tells us that we would overestimate the effect of buzz volume when ad spending is omitted; volume would do some of the explaining that should have been done by mass ad spend. The omitted variable bias can be extended for multiple predictors. In our example, one can think of many other causal factors that are related to both purchase and the volume of social buzz, for example, firm-generated content on social media, brand crisis, and even competitive activities. An important lesson is that even though the researcher might have a behavioral data set with many observations and variables, the variables that were not measured might be more important than the ones that were measured.

There is no simple, foolproof remedy for the omitted variable bias. A good starting point is to specify a conceptual framework that identifies the relevant causal variables, especially those that might be correlated with the main predictor variable(s) being studied. Constructing such a framework requires a solid theoretical understanding of the situation. Next, the researcher should attempt to get measurements of the relevant causal factors. In survey research, questions can be easily added, but in behavioral data research one may have to supplement existing data sets with new ones. In the social buzz example, one may have to obtain ad spend from another source, such as the Kantar Media data or the RedBooks data.

How to Reduce/Avoid Multicollinearity?

Multicollinearity is when two or more predictor variables are correlated, which makes it difficult to assess the extent to which the individual predictors affect the criterion. This issue is frequently present in behavioral data, where multiple related measures are recorded simultaneously about a consumer. While standard regression textbooks define multicollinearity and give diagnostics for detecting it (e.g., examining the correlation matrix of predictors and variance inflation factors), they usually do not differentiate between its different causes, nor discuss how to handle it beyond covering methods such as mean-centering for interactions (Aiken and West 1991) and variations that are not always suitable such as stepwise, ridge and principal components regression. Also see Leeflang et al. (2000, pp. 358–61) for related discussion. To remedy multicollinearity effectively, one must understand why it exists. We survey different causes of multicollinearity and point out solutions for each case in the behavioral data context.

There are three possible reasons for why predictor variables could be correlated: one x causes the other, and two subcases for when both xs are caused by another variable. If one x causes another, then the researcher should consider performing a mediation analysis. For example, a company may create a social media prompt in certain periods, which causes social media buzz, which in turn causes sales. Mediation analysis is often done in advertising research, and there are many excellent resources on this topic (e.g., see Hayes 2013; Zhao, Lynch, and Chen 2010).

When the x variables are correlated because of their relationship to another variable, there are additional considerations concerning whether the observed variables are decision variables or if they are reflections of a higher-order construct that is of theoretical interest. If the correlated variables are decision variables, then they should both be included. For example, price reductions may be associated with increases in ad spending because a brand may typically announce price reductions and not advertise as much when the product is not on promotion. Both decision variables should be included to explain sales; otherwise the effect of one variable will be overstated because of the omitted variable bias.

Alternatively, two predictor variables could be correlated because they are manifestations of some common construct. If the objective of the analysis is to make causal inferences (as opposed to only making predictions), it is usually best to develop theory about the construct, estimate and validate it with some sort of factor analysis (e.g., see MacKenzie, Podsakoff, and Podsakoff 2011), and use the construct estimates rather than individual variables in subsequent analyses. Readers are likely familiar with scale development, and the thought process and methods can be useful in behavioral data situations. For example, suppose that a cable TV provider can tally the amount of time that each household spends watching various types of content over an extended period, say baseball, football, basketball, soccer, hockey, and so on. Moreover, the (logged) variables are highly correlated. Depending on the specific research question, it is likely better to conceptualize a general sports interest variable rather than trying to theoretically justify specifically which sport variables to include, which would be a difficult task to defend. The omitted variable bias tells us that the slope estimates of, say baseball, will depend highly on the number of other sports variables included in the model.

Endogeneity Issues

Endogeneity is another issue that rarely causes an issue in lab experiments but frequently plagues the analysis of behavioral data. Imagine an analysis of TV tuning data and purchase records finds that those who watched a specific commercial are more likely to purchase the advertised product than consumers not exposed to the commercial. While one may think that the commercial is effective in promoting the product, it is possible that consumers who chose to watch the program in which the commercial was embedded are systematically different from those that did not watch the program (e.g., be more innovative, less price sensitive, more uniqueness-seeking). Such systematic differences could have led to the difference in purchase, not the commercial itself. When these underlying systematic differences are not included in the model (hence relegated to the error term), it causes an endogeneity issue. As researchers may not always have the ability to randomly assign consumers to treatment conditions, such self-selection-based endogeneity can be a serious concern. More broadly, self-selection based endogeneity is a special case of endogeneity. Endogeneity exists whenever an independent variable is correlated with the model error term, which may be caused by things such as measurement error and simultaneity. Left unaddressed, endogeneity will lead to inconsistent model estimates.

What should one do in such a situation? The first line of defense would be to make sure that all theoretically relevant variables have been considered and their effects controlled for. When possible, studies should also be designed to minimize systematic variations across conditions. Another line of defense, which lends itself to behavioral-data situations where consumer behaviors are monitored over time, is to incorporate pre-measures into the analysis. If those who self-select into a treatment are systematically different from the control, a pre-measure of the dependent variable can quantify the bias. A “difference in differences” analysis examines whether the change in the dependent variable (posttreatment minus pretreatment) differs between the treatment and control groups. See Allison (1990) for a survey of methods of analyzing such pre-post data. The pioneering single-source studies of Lodish et al. (1995) and Hu, Lodish, and Krieger (2007) are also instructive in analyzing before-after natural experiments. They use IRI BehaviorScan split-cable TV data, where subjects are from certain markets having two cable operators which can expose their consumers to different levels of advertising using a pre-post with control-group design (see appendix, Hu, Lodish, and Krieger 2007). The outcome sales measures are gathered through diaries, or later with scanners at grocery and drug stores. The authors employ several methods to make sure the groups receiving different levels of advertising are comparable, including controlling for other marketing mix variables that were not experimentally manipulated. Among their findings were that TV advertising weight tests were more effective for new products than for established ones. They also found substantial variability in the advertising effectiveness of each test, which indicates heterogeneity in the ad response function should not be ignored. See the discussion on random effects that follows.

Assuming necessary steps have been taken in the design process but endogeneity is still suspected, additional steps need to be taken to correct for the issue. We elaborate on two approaches here, instrumental variables and propensity scoring. Besides these two, there are other ways such as regression discontinuity design (Thistlethwaite and Campbell 1960) that can also address self-selection.

Instrumental variables. The basic idea of instrumental variables is to use other variables that are related to the suspected endogenous variable but not to the dependent variable. In the example mentioned, a possible instrumental variable could be the number of hours watching TV per week, which arguably influences the likelihood of being exposed to the commercial but not necessarily purchase of a specific advertised product. The steps for the instrumental variables approach are presented in the list that follows. Interested readers can consult Wooldridge (2002) and Leenheer et al. (2007) for more details and examples of this approach.

Step 1: Identify one or more instrumental variables (e.g., hours spent watching TV per week) related to the endogenous variable but not to the dependent variable. The instrumental variable(s) cannot already be in the main model.

Step 2: Regress the endogenous variable (e.g., exposure to commercial) on all predictor variables in the main model, as well as instrumental variables.

Step 3 (optional): If the main model contains an interaction term involving the endogenous variable, it is also necessary to do the same regression in Step 2 with the interaction term being the dependent variable.

Step 4: Run the main model as previously specified. But instead of using the actual values of the endogenous variable and its interaction term (if applicable) in the model, use the predicted values from the regressions in Step 2 and Step 3.

The instrumental variables approach can also be used to conduct the Hausman test of endogeneity to identify whether endogeneity is indeed an issue (see Wooldridge 2002, sec. 6.2.1). To do so, follow the same steps 1 through 3. Then run the original OLS main model but include the residuals from steps 2 and 3 into the model, and conduct an F test to see whether including these residual items result in a significantly better fit than the original OLS regression without the residual terms. A significant F test would suggest an endogeneity problem.

When to use this instrumental variables approach? It does require information on at least one other variable not in the main model that can serve as an instrumental variable. But finding an appropriate instrumental variable is not an easy task, and sometimes such data are simply not available. In that case, other approaches such as propensity scoring can be explored instead.

Propensity scoring. Another way to handle self-selection-based endogeneity is with propensity scoring models (e.g., see Stuart 2010). For example, Wang, Malthouse, and Krishnamurthi (2015) study the effects of the adoption of a grocery retailer's mobile app on future purchases at the retailer. Those customers who elected to adopt the app tended to be better customers to begin with than those who decided not to adopt it. Comparing adopters with nonadopters is flawed because the groups are different. The idea of propensity scoring models is to find a matched control group that is as similar as possible to the treatment group to strengthen internal validity.

If there were only one attribute on which the treatment and control groups systematically differ, then matching would be easy. For example, if the single attribute was age, the researcher could easily find a control with a similar age for each person who opts into treatment. (It is common for the pool of possible control subjects to be much larger than the treatment group.) But what if there were many attributes on which the groups differed? Matching on multiple attributes is much more difficult because it is not simple to know how much to weight differences on the different attributes. Propensity scoring solves this problem with a three-step process, which is implemented by the MatchIt package in R:

Step 1: Use logistic regression to predict whether a subject opts into treatment from a wide range of attributes that are known at the time of the decision. A difficulty with this step is justifying that all relevant predictors are included in the model. Stuart (2010, p. 5) advises being liberal in terms of including variables. The predicted probabilities of opting into treatment are the propensity scores.

Step 2: Matching is performed on the propensity scores. One simple way to do this is one-to-one matching without replacement, where each treated subject is matched to one control with a similar propensity score, and controls can only be assigned to at most one treated subject. Stuart (2010) discusses variations where multiple controls are selected for each treated subject, or the matching is done with replacement. The researcher should perform diagnostic checks to confirm that the two groups are similar.

Step 3: The treated cases and matched controls are used to study the effect of the treatment on the outcome measure. For example, the outcome may be regressed on the treatment indicator, controlling for all attributes included in the propensity scoring model.

This section has discussed methods of making causal inferences from observational data. In an influential scholarly discussion of the topic, Freedman and Berk agree that social scientists need to have "realistic aspirations about what can be learned from a given dataset, a far greater concern with data collection and research design, and the creative use of multiple datasets" (Berk 1991, p. 358). As a general guideline for making causal inferences with behavioral data, it is good to look for "shocks," such as the introduction of a mobile app in Wang, Malthouse, and Krishnamurthi (2015), a change in policy (e.g., see Kim et al. 2016), and so on. With such shocks, it is easy to form a matched control group and premeasures to strengthen internal validity. Without such a shock, where customers select into the treatment over an extended period of time, it is still possible to make causal inferences, but it will be more difficult and complicated to find controls and there will be even more confounding factors to consider, such as seasonality and changes in competitive activities. Berk also emphasized the importance of multiple data sets. A particular data set may lack data on key causal factors, and good research will often require stitching together data from multiple sources to control for confounds.

Models for Grouped Data Structure

Behavioral data often take on a grouped or nested data structure. For example, there may be many weekly observations for each consumer in the sample (i.e., weekly observations nested under consumers), or brand-level information may be nested under firms. OLS regression assumes model errors to be independently and identically distributed. With a nested data structure, this assumption may not be true because there could be consumer heterogeneity that causes the weekly observations to be related to each other within the same consumer and to differ systematically from those of another consumer. This violation of the assumption renders the estimated standard errors for the model coefficients incorrect and, as a result, makes hypothesis testing using these standard errors no longer valid. Researchers can look to two common ways of addressing the situation: using clustered robust standard errors and using hierarchical/multilevel models.

Clustered robust standard errors. The first approach is a fairly straightforward extension of the typical OLS model, as the coefficients are still estimated in the same fashion but the standard errors for the estimates (needed for hypothesis testing) are modified. The basic idea is that the error terms for the observations (in the previously mentioned scenario, the transactions) are clustered by the higher-level entities (i.e., consumers or groups of consumers), so that the errors for the observations of the same consumer are allowed to correlate with each other. This changes the calculation of the standard errors of the coefficient estimates, taking into consideration the clustered structure. Clustered robust standard errors, as the name suggests, are fairly robust to misspecifications and within-cluster correlations. Most statistics software packages have procedures or options for obtaining such estimates, such as the vce(cluster) option in STATA, the vcovHC() function in the plm package in R, and the cluster line in proc surveyreg in SAS. Although clustered robust standard errors carry few assumptions, they do require a sufficiently large number of clusters, recommending at least 50. This is typically not a problem with behavioral data.

Multilevel models. Another way of dealing with a grouped data structure is to use a multilevel model. Taking the example earlier, where transactions are nested within individual consumers, two levels of entities need to be modeled: transactions (level 1) and consumers (level 2). Suppose we believe that the amount of spending in each transaction is driven by the amount of advertising the consumer is exposed to and product price levels, besides individual differences. The level 1 equation would look very much like a regular OLS regression model, as follows:

$$\text{Spending}_{0i} = \beta_{0i} + \beta_{1i}\text{Advertising}_{ij} + \beta_{2i}\text{Price}_{ij} + e_{ij}$$

The key difference here is that rather than every consumer sharing the same intercept and slope estimates, the equation allows the baseline spending (the intercept) and/or the effects of advertising and price to differ across consumers. We can then model these beta coefficients using explanatory variables at the higher (consumer) level, such as age, gender and income. Following is a summary of the steps involved in implementing a multilevel model:

Step 1: Identify the structure in the data. Researchers need to determine the bottom-level entities (e.g., transactions, brands), the second-level entities (e.g., consumers, firms), and sometimes even higher-level entities (e.g., countries).

Step 2: Specify the first-level equation using explanatory variables that vary among the bottom-level entities (e.g., prices encountered in each transaction).

Step 3: Determine what effects/coefficients should vary at the second level. Do consumers' baseline spending differ from each other? If so, the intercept should be individual-specific. Do consumers have different receptiveness to advertising? If so, the slope for advertising should vary. Theoretically with enough observations, one could specify every coefficient in the first-level equation to vary across the second-level entity. But the tradeoff is a more complex model with higher computing cost and fewer degrees of freedom. Hence the determination made in this step should be based on theory and the data.

Step 4: Specify the second-level equations using explanatory variables that vary among the second-level entities (e.g., each consumer's advertising receptivity).

Step 5: Repeat this process for further levels of the model if needed. Most applications will have two or three nested layers.

Step 6: Model estimation, diagnostics, and interpretation. This is where the researcher can check whether some higher-level variations specified in the model are indeed necessary, whether there are missing variables, and so on. With an adequate model, one can then proceed to interpret the model estimates.

Multilevel models can be implemented using the HLM software specifically designed for such purposes, the nlme or lme4 packages in R, PROC MIXED in SAS, and SPSS Mixed Models. HLM tends to be the most intuitive, as it directly takes the equations at each level as the model specification. For most of the other packages, one needs to plug in the higher-level equations into the lower-level equations to create a single equation as model input. Excellent references exist for readers interested in reading further such as Snijders and Bosker (2012) and Galecki and Burzykowski (2013).

Count Data Models

Although not frequently encountered in experimental data, behavioral data often involve a dependent variable that is a count of something, such as the number of website visits or the number of social media sharings in a certain time period. While such a variable looks similar to a continuous variable, it is not really continuous since the variable can only take nonnegative integer values (0, 1, 2, 3, etc.). In such situations, Poisson regression is typically used. It treats the count process as following a Poisson distribution and models the probability of arriving at a certain count as follows: $Y \sim \text{Poisson}(\lambda)$, where $\log(\lambda) = X'\beta$. Such models can be estimated using the glm() function in R, PROC GENMOD in SAS, and generalized linear models analysis in SPSS.

The interpretation of coefficients from a Poisson regression is less straightforward than the OLS regression. In Poisson regression, a one unit increase in an explanatory variable with coefficient β changes the expected count by $\exp(\beta)$ times.[4] When $\beta < 0$, the change factor $\exp(\beta)$ will be smaller than 1, indicating a decrease in expected count. In contrast, when $\beta > 0$, the change factor $\exp(\beta)$ will be larger than 1, suggesting an increase in expected count. Hence, the directional interpretation of the coefficients is still similar to OLS regression.

Dealing With Zeros in the Data

A frequently encountered situation in behavioral data is when there are a large number of zeros for a nonnegative outcome variable. For example, for many weeks, a consumer may not visit a website at all and hence will report a weekly total visit duration of 0 minutes. Although a few zeros here and there do not pose a problem, having many zeros can lead to biases in OLS estimates. Alternatively, some researchers may choose to omit such zero cases as having no observation, which is incorrect because the act of not visiting also contains useful information. Multiple approaches can deal with such "zero-abundant" situations, such as the various types of Tobit models, zero-inflated models, and the hurdle model. While the statistical treatments across these approaches differ, the core idea behind the different approaches is quite similar. Generally, these methods dissect the outcome into two parts. One can be called a membership part that determines whether an observation is inherently zero or nonzero/positive, and the other part models the actual quantity or amount of the outcome.

Given the core idea, the different models vary on whether the first part and the second part are driven by the same (e.g., the classic Tobit model) or different (e.g., Tobit II model and hurdle model) processes, whether the two parts are estimated sequentially (e.g., hurdle two-step model) or together (e.g., Tobit models and zero-inflated models), and whether they are suitable for a discrete outcome (e.g., zero-inflated models and hurdle models) or a continuous outcome (e.g., Tobit models), among other subtler differences. These approaches have been implemented in standard statistical packages. For example, the zero-inflated models and the hurdle model can be estimated using the zeroinfl() and hurdle () functions in the R pscl package, and PROC GENMOD in SAS; and Tobit models can be estimated using PROC QLIM in SAS, the vglm() function in the R VGAM package, and the select() function in the R sampleSelection package. More in-depth examination of models for data with excessive zeros can be found in Long (1997, ch. 7 and 8), and Cameron and Trivedi (2005, ch. 16 and 20). Sample applications of these approaches in the advertising and marketing literature that are not overly technical can be found in Danaher and Dagger (2013, Tobit II model), Hinz et al. (2011, hurdle model), Liu-Thompkins and Tam (2013, zero-inflated models), and Malthouse (2013, ch. 6, two-step models).

INTERPRETING AND PRESENTING THE RESULTS

Opportunities for Results Validation Using Behavioral Data

Many of diagnostic measures advertising researchers are familiar with (e.g., R-squared, residual analysis) are equally applicable in the behavioral data context. We do not elaborate on them here. With an overabundance of data, there may be additional opportunities for testing the generalizability of model findings in a behavioral data context. For example, data from a subsample of consumers can be put into a holdout sample that will not be used in model estimation. The researchers can then see how well the model findings work in this separate subsample by comparing the model estimates using this holdout sample with what is observed in this sample.[5] As another example, one or more observations for each consumer in a different time period may be treated as a holdout sample to test the robustness of the findings across time. While such out-of-sample validation approaches are typically used to test a model's ability to predict, they can also yield insight on the generalizability of causality related findings. Freedman (1991) states that "replication and prediction of new results provide a harsher and more useful validation regime than the statistical testing of many models on one data set. Fewer assumptions are needed, there is less chance for artifact, more kinds of variation can be explored, and alternative explanations can be ruled out" (p. 307). See James et al. (2013, sec. 5.1) for more discussion of cross validation.

Visualizations

Data visualization has become even more important with behavioral data, as dashboards and other reporting applications have proliferated. Brand managers now have more measurements available for monitoring and tracking brand-related phenomena. Visual displays can also be useful in presenting and understanding the results from a statistical model. It is beyond the scope of this article to summarize the vast literature on visualization. Two of the pioneering researchers in this field are Cleveland and Tufte, each of whom have multiple books and articles (see e.g., Cleveland 1985; Tufte 1983).

DISCUSSION

The proliferation of digital environments in which many consumer actions can be recorded implies that ad researchers can now have access to digital records of many consumer activities. Such data sources offer tremendous opportunities to complement traditional lab experiments and survey data, especially in areas such as addressable advertising, advertising performance evaluation, online display and search advertising, and mobile advertising. In the meantime, rich behavioral data also bring unique challenges and pitfalls. This article serves as an aggregation place for such issues and intends to help advertising researchers navigate this terrain. While we cannot delve into each issue in depth due to space constraint, we hope this article at least makes advertising researchers new to the behavioral data domain cognizant of possible issues, and we provide book and article references throughout for anyone interested in learning more about a specific topic.

Although we have focused on conducting causal research using behavioral data, such data can also be very useful in generating exploratory insight that perhaps no existing theory can deftly explain yet. For example, analyses of social media conversations about advertisers may reveal a new dynamic of consumer-brand relationship built on the continuous cocreation of brand stories, something that existing theories have not been able to fully understand and explain. Although we do not cover such insight-driven exploratory research, we note that induction-based research using behavioral data can be equally important in driving new areas of scientific discovery.

Of course, it would be remiss not to mention the opportunity of field experimentation combined with rich behavioral data. Our reading of the top advertising and marketing journals from the past two and a half years already reveals 10 such articles that feature at least one field experiment to address an advertising research question. Using randomization or sample matching (e.g., propensity score matching), researchers can benefit from both the rigor of an experimental design and the realism of actual behavior through field experiments. Besides answering intriguing new research questions, we believe field experiments combined with behavioral data offer great opportunities for replication studies of important past work done with small groups in a lab setting.

To us, the journey of utilizing behavioral data to answer advertising research questions is one too fascinating to miss. We encourage advertising researchers who have not dipped their toes in this water to add a few new tools in their toolbox (or a few new collaborators from other disciplines) and experience the fun and challenges of such data through their own research. Bon voyage!

NOTES

1 When the problem is ill-defined, exploratory research may precede descriptive or causal designs (Churchill and Iacobucci 2005, Figure 4.1).
2. Web scraping, also known as web data extraction, refers to techniques of using computer software to extract information from websites automatically. Sample web-scraping tools or services include import.io and Visual Scraper.
3. Figure 1 from Kim et al. (2014) shows that in 1980 fewer than 30% of advertising research papers in academic journals were theory driven, but by 2010 67% of articles were theory driven. The second part of the figure shows that these percentages are even higher for the *Journal of Advertising* than for the *Journal of Advertising Research*, but does not break out the percentages by journal over time. Thus, there is a strong trend in advertising toward theory-based research. Whether this trend will or should continue is an important discussion question.
4. Poisson regression uses a log "link" function on y. Suppose $\log(y) = a + bx$. A unit change in x implies a change from $\log(y)$ to $\log(y) + b$. Note that $\exp(\log(y) + b) = y * \exp(b)$.
5. When using holdout samples to validate model results, it is necessary to recognize the issues associated with such validation when there is endogeneity correction in the model, as illustrated in Ebbes, Papies, and van Heerde (2011). There, the authors recommend that an exogenous holdout sample be used if the purpose is to make consistent causal inference, and a nonendogeneity corrected model be used instead if the purpose is to make accurate predictions on other similarly endogenous samples.

REFERENCES

Aiken, Leona S., and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions*, Thousand Oaks, CA: Sage.

Allison, Paul (1990), "Change Scores as Dependent Variables in Regression Analysis," *Sociological Methodology*, 20, 93–114.

Berk, Richard (1991), "Toward a Methodology for Mere Mortals," *Sociological Methodology*, 21, 315–24.

Cameron, A. Colin, and Pravin Trivedi (2005), *Microeconometrics: Methods and Applications*, New York: Cambridge University Press.

Churchill, Gilbert A., Jr., and Dawn Iacobucci (2005), *Marketing Research: Methodological Foundations*, Thousand Oaks, CA: Sage.

Clancy, K. J., P. D. Berger, and T. L. Magliozzi (2003), "The Ecological Fallacy: Some Fundamental Research Misconceptions Corrected," *Journal of Advertising Research*, 43 (4), 370–80.

Cleveland, William (1985), *The Elements of Graphing Data*, Monterey, CA: Wadsworth Advanced Books and Software.

Danaher, Peter, and Tracey Dagger (2013), "Comparing the Relative Effectiveness of Advertising Channels: A Case Study of a Multimedia Blitz Campaign," *Journal of Marketing Research*, 50 (4), 517–34.

Ebbes, Peter, Dominik Papies, and Harald van Heerde (2011), "The Sense and Non-Sense of Holdout Sample Validation in the Presence of Endogeneity," *Marketing Science*, 30 (6), 1115–22.

Feit, Eleanor Mcdonnell, Pengyuan Wang, Eric T. Bradlow, and Peter S. Fader (2013), "Fusing Aggregate and Disaggregate Data with an Application to Multiplatform Media Consumption," *Journal of Marketing Research*, 50 (3), 348–64.

Freedman, David (1991), "Statistical Models and Shoe Leather," *Sociological Methodology*, 21 (2), 291–313.

———, Robert Pisani, and Roger Purves (1998), *Statistics*, New York: W.W. Norton.

Galecki, Andrzej, and Tomasz Burzykowski (2013), *Linear Mixed-Effects Models Using R: A Step-by-Step Approach*, New York: Springer Science+Business Media.

Hayes, Andrew (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, New York: Guilford Press.

Hinz, Oliver, Bernd Skiera, Christian Barrot, and Jan Becker (2011), "Seeding Strategies for Viral Marketing: An Empirical Comparison," *Journal of Marketing*, 75 (6), 55–71.

Hu, Y., L.M. Lodish, and A.M. Krieger (2007), "An Analysis of Real World TV Advertising Tests: A 15-Year Update," *Journal of Advertising Research*, 47 (3), 341–53.

James, Gareth, Danielle Witten, Trevor Hastie, and Robert Tibshirani (2013), *An Introduction to Statistical Learning*, New York: Springer.

Kim, Kyongseok, Jameson Hayes, J. Adam Avant, and Leonard Reid (2014), "Trends in Advertising Research: A Longitudinal Analysis of Leading Advertising, Marketing, and Communication Journals, 1980 to 2010," *Journal of Advertising*, 43 (3), 296–316.

Kim, Su Jung, Rebecca Wang, Ewa Maslowska, and Edward Malthouse (2016), "'Understanding a Fury in Your Words': The Effects of Posting and Viewing Electronic Negative Word-of-Mouth on Purchase Behaviors," *Computers in Human Behavior*, 54, 511–21.

Leeflang, Peter, Dick Wittink, Michel Wedel, and Philippe Naert (2000), *Building Models for Marketing Decisions*, Dordrecht, the Netherlands: Kluwer Academic.

Leenheer, Jorna, Harald van Heerde, Tammo Bijmolt, and Ale Smidts (2007), "Do Loyalty Programs Really Enhance Behavioral Loyalty? An Empirical Analysis Accounting for Self-Selecting Members," *International Journal of Research in Marketing*, 24 (1), 31–47.

Liu-Thompkins, Yuping, and Leona Tam (2013), "Not All Repeat Customers Are the Same: Designing Effective Cross-Selling Promotion on the Basis of Attitudinal Loyalty and Habit," *Journal of Marketing*, 77 (5), 21–36.

Lodish, Leonard M., Magid Abraham, Stuart Kalmenson, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens (1995), "How T.V. Advertising Works: A Meta-Analysis of 389 Real World Split Cable TV Advertising Experiments," *Journal of Marketing Research*, 32 (2), 125–39.

Long, J. Scott (1997), *Regression Models for Categorical and Limited Dependent Variables*, Thousand Oaks, CA: Sage.

MacKenzie, Scott, Philip Podsakoff, and Nathan Podsakoff (2011), "Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques," *MIS Quarterly*, 35 (2), 293–334.

Malthouse, Edward (2013), *Segmentation and Lifetime Value Models Using SAS*. Cary, NC: SAS Institute.

Osborne, Jason (2013), *Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do before and after Collecting Your Data*, Thousand Oaks, CA: Sage.

Peter, J. Paul (1983), "Some Philosophical and Methodological Issues in Consumer Research," in Shelby D. Hunt, *Marketing Theory: The Philosophy of Marketing Science*, Homewood, IL: R.D. Irwin, pp. 382–94.

Robinson, W. S. (1950), "Ecological Correlations and the Behavior of Individuals," *American Sociological Review*, 15, 351–57.

Snijders, Tom, and Roel Bosker (2012), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd ed., London: Sage.

Stuart, Elizabeth (2010), "Matching Methods for Causal Inference: A Review and a Look Forward," *Statistical Science*, 25 (1), 1–21.

Thistlethwaite, Donald, and Donald Campbell (1960), "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology*, 51 (6), 309–17.

Tufte, Edward (1983), *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press.

Wang, Rebecca, Edward Malthouse, and Lakshman Krishnamurthi (2015), "On the Go: How Mobile Shopping Affects Customer Purchase Behavior," *Journal of Retailing*, 91 (2), 217–34.

Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

Zhang, Yao, Eric T. Bradlow, and Dylan S. Small (2015), "Predicting Customer Value Using Clumpiness: From RFM to RFMC," *Marketing Science*, 34 (2), 195–208.

Zhao, X., John, Lynch, and Q. Chen (2010), "Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis," *Journal of Consumer Research*, 37 (2), 197–220.