



# The past, present, and future of measurement and methods in marketing analysis

Yu Ding<sup>1</sup>  · Wayne S. DeSarbo<sup>2</sup> · Dominique M. Hanssens<sup>3</sup> · Kamel Jedidi<sup>1</sup> · John G. Lynch Jr.<sup>4</sup> · Donald R. Lehmann<sup>1</sup>

Published online: 08 June 2020  
© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

The field of marketing has made significant strides over the past 50 years in understanding how methodological choices affect the validity of conclusions drawn from our research. This paper highlights some of these and is organized as follows: We first summarize essential concepts about measurement and the role of cumulating knowledge, then highlight data and analysis methods in terms of their past, present, and future. Lastly, we provide specific examples of the evolution of work on segmentation and brand equity. With relatively well-established methods for measuring constructs, analysis methods have evolved substantially. There have been significant changes in what is seen as the best way to analyze individual studies as well as accumulate knowledge across them via meta-analysis. Collaborations between academia and business can move marketing research forward. These will require the tradeoffs between model prediction and interpretation, and a balance between large-scale use of data and privacy concerns.

**Keywords** Measurement · Methods · Marketing Analysis · Meta-Analysis · Replicability · Customer Heterogeneity · Brand Equity

## 1 Constructs measurement

### 1.1 Past (pre-2000)

Measurement is about assigning numbers to a set of units reflecting their value on some underlying property. For some quantities, this is a simple and relatively objective matter—e.g., sales of Cheerios in the 12 months of 2019. However, for other variables like brand attitude or loyalty, there is an abstract underlying construct being measured,

---

✉ Yu Ding  
yu.ding@columbia.edu

and imperfection of the translation between conceptual variable and operational measures. The same holds when measuring firm-level variables like market orientation or brand equity.

Psychologists in the first half of the twentieth century rejected unobservable states, traits, and processes as proper objects of scientific study; they equated a concept with its operational definition. This rejection was a reaction to the lack of replicability of psychological research that relied on introspection as a method (Boring 1953).

A half-century ago, Jacoby (1978) deplored the lack of rigor in marketing measurement. Much has changed since then, in no small part due to the influence of two giants in social science research methods, Lee Cronbach and Donald Campbell.

Cronbach and Meehl (1955) legitimized the study of unobservable phenomena where no single operational definition could be held to be a “gold standard.” One could postulate some unobservable construct like brand loyalty or attitude and create a measure that imperfectly taps that underlying construct. One validates that measure and the claim that it taps “brand loyalty” if it correlates with measures of posited antecedents and consequences in some “nomological net.” Cronbach and Meehl argued:

Psychology works with crude, half-explicit formulations.... Yet the vague, avowedly incomplete network still gives the constructs whatever meaning they do have. Since the meaning of theoretical constructs is set forth by stating the laws in which they occur, our incomplete knowledge of the laws of nature produces a vagueness in our constructs....

Measurement can be imperfect due to random error—undermining “reliability”—or due to low “validity” caused by random error plus systematic contamination by other constructs. Cronbach (1951) proposed his alpha measure of reliability, reflecting internal consistency of different items in a multi-item test. Rust and Cooil (1994) generalized reliability formulae for qualitative and quantitative data.

Campbell argued that we could triangulate on truth by comparing results from multiple imperfect operationalizations of a given construct. Campbell and Fiske (1959) introduced “convergent validity”—measuring agreement between different methods for measuring the same construct—and “discriminant validity”—measuring discrimination of one’s measure of a new construct from measures of other related constructs. Heeler and Ray (1972) introduced these ideas to marketing. Later, Fornell and Larcker’s (1981) confirmatory factor analytic test for discriminant validity became widely used. Peter’s (1979) review paper on reliability and the paradigm of Churchill’s (1979) and Gerbing and Anderson (1988) for developing marketing scales were major influences.

Bollen and Lennox (1991) distinguished between “cause” indicators where the indicators are the causes of the latent construct and “effect” indicators where they reflect the underlying latent construct. Only “effect” indicators need to show internal consistency. For cause indicators – e.g. beliefs about a brand that determine one’s overall attitude toward the brand – there is no reason why beliefs about attribute 1 vs attribute 2 should correlate. Indeed, if they do, this may reflect “halo effects” (Beckwith and Lehmann 1975).

The measurement of constructs is inextricably linked to testing theories about how those constructs relate to each other and to observable, operational measures. A large literature in marketing developed using LISREL for theory testing (Bagozzi and Yi 1988; Fornell and Larcker 1981). Anderson and Gerbing (1988) introduced a two-step procedure for testing “structural equations” models of the relationships among latent constructs. First, using confirmatory factor analysis to test for the adequacy of the measurement model of the indicators of each construct. Second, testing hypothesized relationships among those constructs. Such models imply a testable pattern of partial relationships among constructs. This same idea underpins one of the most important developments in theory testing in the social sciences: “mediation analysis” tests how some independent variable  $X$  “indirectly” affects dependent variable  $Y$  via a mediator  $M$  (Judd and Kenny 1981; Baron and Kenny 1986).

### 1.2 Present (2000–2020)

“Reliability” reflects freedom from measurement error. Statisticians have shown how increased measurement error is produced when one discretizes continuous measures, as in “median splits” (Irwin and McClelland 2001).

If one’s indicators cause the underlying construct rather than reflecting it, internal consistency reliability is irrelevant. One must have a comprehensive set of cause indicators. This realization led to advances in “index” measures of constructs that are caused by their underlying indicators (Diamantopoulos and Winklhofer 2001; Jarvis et al. 2003).

With respect to validity, marketing scholars have continued to debate what constitutes convincing evidence of mediation (Zhao et al. 2010). The Baron and Kenny method claimed that in order to show mediation chain  $X \rightarrow M \rightarrow Y$ , one has to show:

- a) The independent variable  $X$  is significantly related to the mediator  $M$ ;
- b) The independent variable  $X$  is significantly related to the dependent variable  $Y$ ;
- c) When  $Y$  is predicted using both  $X$  and  $M$ , the mediator  $M$  should have a significant partial effect.

Zhao et al. and others have argued that (b) is not necessary, because the zero-order  $X$ - $Y$  relationship is the sum of the indirect  $X \rightarrow M \rightarrow Y$  channel and any “direct” effect of  $X$  on  $Y$ , perhaps with opposite sign. Moreover, (c) is weak evidence—correlational rather than causal—so any claim that the mediator  $M$  causes  $Y$  requires other evidence, often experimental, of the  $M \rightarrow Y$  link (Peters 2017).

### 1.3 Future (2021–2030)

Like other social sciences, marketing has relied heavily on deductive techniques where researchers specify a priori relationships among constructs and measures and test them. Without a prior specification, authors can fool themselves by running multiple models and believing the models that show significant results (Simmons et al. 2011). Increasingly, researchers debate the appropriate mix of induction and deduction (Lynch et al. 2012; Ledgerwood and Shiffrin 2018) depending on the aims of the research. Machine learning inductive methods are increasingly seen as acceptable and useful (Yarkoni and Westfall 2017).

In the context of construct identity and measurement, Larsen and Bong (2016) have shown that one can use text analysis and machine learning to determine whether two sets of authors are using different labels for the same constructs or similar labels for different constructs. Section 2 discusses how machine learning methods are improving our understanding of the replicability of findings.

## 2 Cumulating knowledge and replicability

### 2.1 Past (pre-2000)

Cronbach et al. (1972) gave the first systematic approach to assess generalizability across different settings, consumers, times, and contexts. Cook and Campbell (1979) distinguished between “internal validity,” “construct validity,” “external validity,” and “statistical conclusion validity” for experiments and quasi-experiments. Internal validity is whether one can confidently conclude that one’s operational independent variable caused differences in one’s operational dependent variable—saying nothing about the abstract labels attached to those operational variables. Construct validity is about the validity of conclusions about the abstract labeling of a cause-effect finding in terms of some higher-order constructs that are more general. External validity is about whether the treatment effects observed are general versus moderated by some unobserved or unanticipated background factors. Statistical conclusion validity is about freedom from type 1 and type 2 errors.

Lynch (1982, 1999) argued that one could not “attain” external validity by following some particular method (e.g., using field rather than lab experiments). One can only “assess” external validity by testing for sensitivity to background factor  $\times$  treatment interactions reflecting a lack of generality of key treatment effects. Interactions between one’s treatment manipulations and background factors thought to be irrelevant often imply incompleteness of one’s theory.

A major development in marketing has been the use of meta-analysis. Meta-analyses began appearing in the social science literature (e.g., Glass 1976), where the primary focus was on establishing the statistical significance of relations among variables. By contrast, early work in marketing focused on the size of the effect (e.g., Farley et al. 1981; Farley and Lehmann 1986) and how it varied across situational and methodological variables (typically estimated in a regression analysis). The statistical significance of aggregated treatment is arguably unimportant (Farley et al. 1995). More important is whether effects vary across studies beyond what would be expected by chance. If so, can we explain those differences by variation in background factors? Farley et al. (1998) identify the most useful next study given those that have already been conducted.

### 2.2 Present (2000–2020)

The past decade has seen a rising concern with the lack of replicability of published findings, triggered by widely publicized cases involving either outright fraud or inappropriate statistics, running multiple analyses, and reporting the analysis that shows significant results (Simonsohn et al. 2014). This has led to large-scale attempts at

“exact” replication of studies. However, when multiple labs attempt to replicate the same famous study as closely as possible, their results often disagree. Klein and coauthors’ (2014) “Many Labs” project found considerable variation in results across labs that followed “identical” procedures. Therefore, failures to replicate a paper may not reflect spurious results in the paper or lack of statistical conclusion validity. The Replication Corner at *IRJM* (and subsequently *Journal of Marketing Behavior*) was founded on the premise validity that these may signal problems of external validity due to researchers’ poor understanding of background factor  $\times$  treatment interactions for most phenomena we study (Lynch et al. 2015). The Replication Corner publishes “conceptual” replications of important studies—testing robustness of conclusions to changes in operational definitions of key independent and dependent variables.

### 2.3 Future (2021–2030)

As a response to concerns that published results may be spurious, a growing number of researchers have called for “preregistration” of studies (Nosek and Lindsey 2018). The idea is to avoid the hidden escalation of type 1 errors that come if authors cherry pick which analyses to run, with what covariates, and with what sample size.

In marketing, preregistration has been winning adherents. Preregistration aims to improve the diagnosticity of statistical tests by solving statistical problems due to family-wise error. Skeptics suggest the benefits of preregistration are uncertain. Szollisi et al. (2019) point to the putative benefits of preregistration are to fix statistical problems of family-wise error. Alternatively, preregistration forces researchers, to think more deeply about theories, methods, and analyses. The latter argument is similar to Andreassen’s (1985) advocacy of “backward market research” where one sketches out dummy tables of possible outcomes and interpretations before collecting data. Szollisi et al. argue that neither promised benefit of preregistration is actually delivered, that “the diagnosticity of statistical tests depends entirely on how well statistical models map onto underlying theories, and so improving statistical techniques does little to improve theories when the mapping is weak. There is also little reason to expect that preregistration will spontaneously help researchers develop better theories (and, hence, better methods, and analyses).”

Some other fields have adopted the idea of “pre-accepted papers” where researchers initially submit their proposals and, if acceptable, get approved by a journal to publish their work regardless of the empirical results. Top business journals are increasingly experimenting with this approach (e.g., Ertimur et al. 2018). Hopefully, new procedures of publication will encourage investing effort in cumulating knowledge.

A very different approach eschews a priori models and relies on machine learning methods to uncover the complex boundary conditions that moderate independent  $\rightarrow$  dependent variable relationships, including “causal random forests” for estimating heterogeneous treatment effect (Wager and Athey 2018; Chen et al. 2020). When we have only a partial understanding of phenomena, it is difficult to identify complex contingent effects a priori. For example, a doctor may want to engage in “personalized medicine” but not know which treatments will be most effective for which patients. Machine learning techniques can predict the effect of the treatment for a given case more powerfully than standard a priori methods, with cross-validation techniques providing protection from errors from multiple hypothesis tests.

### 3 Sources of data and analysis methods

#### 3.1 Past (pre-2000)

In terms of collecting data, surveys were the primary source (often from a convenience sample). Sales data were generally aggregate and could be matched to advertising and other firm and competitor's aggregate data through substantial effort. Work in marketing included Lehmann and Hulbert's (1972) paper on the impact of the number of scale points used, and Beckwith and Lehmann's (1975) focus on the measurement of the halo effect of overall attitude on measures of specific components/attributes.

Multidimensional scaling (Cooper 1983), factor/cluster/conjoint analysis (Punj and Stewart 1983; Green and Srinivasan 1990), logit models (Guadagni and Little 1983), game theory (Moorthy 1985), structural equation models (Fornell and Larcker 1981; Jedidi et al. 1997), and stochastic and OR models (Gupta 1991; Mahajan et al. 1990) were widely used. See Wedel and Kannan (2016) for a detailed review of development in data and analysis.

#### 3.2 Present (2000–2020)

Data sources have expanded in their accessibility, type, and complexity, whereas pre-2000 sales data were mostly aggregated (e.g., SAMI warehouse withdrawals) or limited to a few categories (e.g., IRI's scanned data on coffee purchase from Pittsfield, Massachusetts); currently, more detailed consumer purchase data are available. The concept of "big data" has emerged and, along with data analytics, is accounting for a considerable portion of research in marketing. Similarly, the use of Mturk or other online data collection platforms has made individual-level data collection much faster and cheaper (i.e., done in hours versus months; Goodman et al. 2013), although there are quality concerns with it (Sharpe Wessling et al. 2017). Moreover, consumers are creating data on their own (vs. providing data upon requests). Research utilizing user-generated content (Fader and Winer 2012; Goh et al. 2013) and the wisdom of the crowd (Prelec et al. 2017) has experienced rapid growth in the past decade. Another source of data is biometric (neuroscience based). Whereas pre-2000 research used galvanometers to measure skin data, information display boards, and MouseLab to assess information acquisition and use (Johnson et al. 1989), now eye-tracking technologies (Pieters and Wedel 2017) and fMRI data (Yoon et al. 2009) provide more details on eye movements and brain activities.

Whereas exploring unobserved consumer heterogeneity was a main research thrust in the 1980s and 1990s, addressing endogeneity issues stemming from omitted variables, simultaneity, and measurement error became a growing concern among marketing academics in the 2000s and 2010s. Several econometric methods have been applied to address empirical issues such as latent and non-latent instrumental variables (Ebbes et al. 2005; Van Heerde et al. 2013), control functions (Papies et al. 2017), Gaussian copula (Park and Gupta 2012), and structural models (Chintagunta et al. 2006). Recently, field experiments have become the gold standard for causal inference (Simester 2017).

Bayesian methods (Rossi and Allenby 2003), hierarchical linear models (Hofmann 1997), latent class models (DeSarbo and Cron 1988), and social network analysis (Van den Bulte and Wuyts 2007) have become common analysis approaches. In addition, we are witnessing the rapid adoption of machine learning methods in marketing research including

topic modeling/text mining (Blei et al. 2010; Netzer et al. 2012), variational and nonparametric Bayesian methods (Dew et al. 2019), deep learning (Timoshenko and Hauser 2019), and reinforcement learning via multi-armed bandits (Schwartz et al. 2017).

### 3.3 Future (2021–2030)

Methods for analyzing spoken language, text, pictures, and other complicated data expand the ability to capture data on a broader scale as do programs for scraping the web and social media sites. They risk overwhelming researchers with a mass of data and, as T.S. Elliot suggested, losing knowledge and information in a “forest” of big data. They also raise issues concerning how to deal with the substantial amount of missing data.

There is also likely to be increased use of virtual reality to more accurately represent (once the novelty wears off, and people are accustomed to using it) the “path to purchase,” among other things. See discussions on marketing strategy (Sozuer et al. 2020), innovation research (Lee et al. 2020), and consumer research (Malter et al. 2020) in this issue.

## 4 Specific examples

### 4.1 Segmentation and heterogeneity

**Past** “Market segmentation involves viewing a heterogeneous aggregate market as decomposable into a number of smaller homogeneous markets in response to the heterogeneous preferences attributable to the desires of consumers for more precise satisfaction of their varying wants and needs” (Smith 1956). Since the pioneering work of Wendell Smith, market segmentation has become pervasive in both marketing academia and practice. In the 1970s and 1980s, a segment was treated as having homogeneous consumers which differed across other segments (e.g., Lehmann et al. 1982), despite the considerable movement of individuals between segments over time (Farley et al. 1987).

Statistical techniques such as cluster analysis, regression analysis, discriminant analysis, cross-tabulation, and ANOVA/MANOVA were employed depending on whether one was performing a priori (segments known) or post hoc (segments unknown) segmentation analysis (Green, 1977). Despite its importance and centrality to marketing, practitioners that attempted to implement segmentation faced several problems (DeSarbo et al. 2017) such as obtaining results that were incongruent with the business and difficult to interpret. Thus, it is not surprising that marketers often became dissatisfied with the results from such market segmentation analyses.

**Present** Every individual consumer is unique in terms of their demographic characteristics, preferences, responses to marketing stimuli, decision-making processes, etc. Similarly, on the B-to-B side, there is heterogeneity related to firmographics, operating variables, purchasing approaches, personal characteristics, etc. (Kotler et al. 2018). While aggregate market analyses tend to mask heterogeneity, individual-level analyses are tedious and often require massive amounts of data. To resolve burdensome data collection efforts, latent class and Bayesian procedures were generated for capturing consumer heterogeneity (Allenby and Rossi 1998; Wedel and Kamakura 2000). Roberts et al. (2014) reported that both



marketing academicians and practitioners ranked marketing segmentation as having the *most substantial impact* of any marketing tool.

Marketing researchers and practitioners formulated various criteria to describe the characteristics of quality segmentation results such as having distinctively behaving segments, being able to identify the members of each segment, and having the ability to reach target segments with a distinct marketing strategy (see DeSarbo et al. 2017). Such criteria provide a better basis for determining the overall quality of the segmentation than mere goodness-of-fit and led to the development of alternative segmentation methods accommodating multiple criteria, managerial a priori knowledge, constraints, differential end-use patterns, etc. (DeSarbo and Mahajan 1984; DeSarbo and Grisaffe 1998; Liu et al. 2010).

**Future** A representative of a major marketing research supplier recently stated: “We purposely try to dissuade our clients from undertaking segmentation studies given the growing dissatisfaction which typically accompanies the aftermath of such studies.” This may be because segmentation has been employed differently in different contexts with different requirements. Traditional multivariate methods (e.g., cluster analysis), latent class-based techniques, Bayesian methods, hybrid methods, and more recently introduced artificial intelligence algorithms (e.g., Hiziroglu 2013) do not typically accommodate all *the segment effectiveness* criteria discussed earlier. We see this as a major area for future methodological development if market segmentation is to continue to be a valued tool for practitioners. This requires the development of new methods that can accommodate multiple objectives, various constraints, dynamics, a priori managerial information known about the business application, normative implications, competitive policy simulations, predictive mechanisms, etc.

## 4.2 Brand and customer equity

**Past** In order to be relevant at the highest levels in the organization, marketing needs metrics that demonstrate its economic value to the organization. One such metric is the value of the brand. The term branding was first used in marketing by Shapiro (1983) and refers to endowing products and services with the power of a brand (Kotler et al. 2018). The resulting consumer perceptions of the brand are critical determinants of the brand’s commercial value, sometimes called brand equity. Ailawadi et al. (2003) utilized econometric methods on scanner data and showed that brand investments result in revenue gains from additional sales, higher price premiums, or both, i.e., strong brands command a *revenue premium*. See a review of brand research on this issue (Oh et al. 2020).

A second metric that is logically connected to marketing is customer equity. Customer equity is the total discounted lifetime values of all the firm’s customers. The first analytical treatment of customer equity was provided by Blattberg and Deighton (1996). The financial benefits of high customer equity are straightforward. However, one must adopt a probabilistic view, as customer equity is future oriented. A particularly appealing metric, the margin multiplier, was proposed by Gupta et al. (2004). This number projects a short-term metric such as profit margin into a longer-term metric that has greater strategic importance to the firm. Among other things, it motivates management to pay close attention to the loyalty of their existing customers (Oblander et al. 2020).



**Present** How do brand and customer equity relate to each other? While the former is intrinsically related to the marketing concept, the latter has a special appeal to financial executives. Using data on the automotive market, Stahl et al. (2012) identified which elements of brand equity contribute to which aspects of customer equity. Their results make a strong case for the need to coordinate the hard (CRM) and the soft (branding) elements of marketing strategy.

The digital age allows improved tracking and prediction of the critical customer metrics of acquisition, retention, and consumption rates. At the same time, consumers now find it much easier to inspect product and service quality before buying. A meta-analysis by Floyd et al. (2014) reported an average sales elasticity of product review quality of 0.69, which is almost seven times the average advertising elasticity. These studies illustrate that the causal chain of product or service experience → customer satisfaction → customer loyalty → firm financial performance is becoming stronger than ever. In fact, Binder and Hanssens (2015) show that across 5000 global mergers and acquisitions between 2003 and 2013, the relative importance of the customer relations score of the acquired firm has grown relative to its brand value score. Similarly, an extensive meta-analysis by Edeling and Fischer (2016) reveals that the customer equity → firm value elasticity averages 0.72, compared with 0.33 for the brand value → firm value elasticity. Thus, while both brand equity and customer equity continue to be important, the latter is gaining in relative terms.

**Future** While several agencies now routinely report brand values of major firms worldwide, their valuations differ, which highlights the need for more precise and reliable measurement of brand values. On the customer equity side, the concept is more readily applied to relationship businesses (e.g., telecoms and financial institutions). However, it needs to be applicable to all organizations to be a universally used performance metric (McCarthy and Fader 2018). In parallel, businesses continue to contribute innovations in data collection and new models to realize their decision-making and application potential. For example, the emerging field of *conversational commerce* uses artificial intelligence (AI) tools to enhance customers' experiences with their suppliers, and at the same time, the digital data produced yield valuable information for firms to better serve their customers in the future. How conversational commerce and other AI innovations enhance customer equity is an important topic for future research.

There are also new challenges. At the societal level, one challenge is the increasing emphasis on privacy. At the managerial level, the abundance of digital data can lead to silo-ism in organizations. For example, the traditional role of VP Sales and Marketing is often split into separate VP Sales and VP Marketing positions. When that happens, marketing and sales compete for corporate resources, and each may inflate their contribution to firm performance in order to fetch higher budgets. Within marketing, there may be a separation of brand marketing and direct marketing. In addition, the direct marketing may be split into online and offline departments, and online marketing may be split into search, social, and mobile. Every split creates an opportunity for more specialized practice but also creates a challenge for management to oversee the total enterprise of marketing. For that, we need to develop performance metrics that properly re-aggregate the contributions of different silos. Marketing academics can contribute here by developing empirical generalizations about the expected impact of different

marketing initiatives that can serve as benchmarks for resource allocation (e.g., Hanssens 2018). Absent that, the familiar plight of the CMO will continue to ring through: “If I add up all the reported returns produced by the different marketing groups in my organization, I end up with a company that is three times the size of its current operations.” The extent to which brand equity and customer equity provide this re-aggregation function is another important area for future inquiry.

## 5 Conclusion

Thanks to many marketing researchers’ efforts, measurements, and methods in marketing analysis have been significantly improved over the last 50 years corresponding to the complex datasets researchers and practitioners have access to. Trading off model goodness-of-fit and predictions vs. model interpretations and generalizability, joint efforts among different tracks within marketing and different experts across disciplines, and between academia and business are needed to ensure the growth of marketing analysis methods. This will be difficult given the need to balance the large-scale use of data and privacy concerns.

## References<sup>1</sup>

- Ailawadi, K. L., Lehmann, D. R., & Neslin, S. A. (2003). Revenue premium as an outcome measure of brand equity. *Journal of Marketing*, 67(4), 1–17.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6), 1173.
- Beckwith, N. E., & Lehmann, D. R. (1975). The importance of halo effects in multi-attribute attitude models. *Journal of Marketing Research*, 12(3), 265–275.
- Binder, C., & Hanssens, D. M. (2015). Why strong customer relationships trump powerful brands. *Harvard Business Review Online*, April 14.
- Boring, E. G. (1953). A history of introspection. *Psychological Bulletin*, 50(3), 169–189.
- Churchill Jr., G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), 64–73.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, 5(3), 249–282.
- DeSarbo, W. S., & Grisaffe, D. (1998). Combinatorial optimization approaches to constrained market segmentation: An application to industrial market segmentation. *Marketing Letters*, 9(2), 115–134.
- DeSarbo, W. S., & Mahajan, V. (1984). Constrained classification: The use of a priori information in cluster analysis. *Psychometrika*, 49(2), 187–215.
- DeSarbo, W. S., Chen, Q., & Blank, A. S. (2017). A parametric constrained segmentation methodology for application in sport marketing. *Customer Needs and Solutions*, 4(4), 37–55.
- Dew, R., Ansari, A., & Li, Y. (2019). Modeling dynamic heterogeneity using Gaussian Processes. *Journal of Marketing Research*, 57(1), 55–77.
- Edeling, A., & Fischer, M. (2016). Marketing’s impact on firm value: Generalizations from a meta-analysis. *Journal of Marketing Research*, 53(4), 515–534.

<sup>1</sup> Due to the word limit, we provide selected references in the paper. Please see the full reference list in the web appendix: <https://osf.io/4te29/>.

- Farley, J. U., & Lehmann, D. R. (1986). *Meta-Analysis in Marketing: Generalization of Response Models*. Lexington: Lexington Books.
- Farley, J. U., Lehmann, D. R., & Ryan, M. J. (1981). Generalizing from “imperfect” replication. *Journal of Business*, 54(4), 597–610.
- Farley, J. U., Winer, R. S., & Lehmann, D. R. (1987). Stability of membership in market segments identified with a disaggregate consumption model. *Journal of Business Research*, 15(4), 313–328.
- Farley, J. U., Lehmann, D. R., & Sawyer, A. (1995). Empirical marketing generalization using meta-analysis. *Marketing Science*, 14(3 supplement), G36–G46.
- Farley, J. U., Lehmann, D. R., & Mann, L. H. (1998). Designing the next study for maximum impact. *Journal of Marketing Research*, 35(4), 496–501.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.
- Guadagni, P. M., & Little, J. D. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2(3), 203–238.
- Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research*, 41(1), 7–18.
- Hanssens, D. M. (2018). The value of empirical generalizations in marketing. *Journal of the Academy of Marketing Science*, 46, 6–8.
- Heeler, R. M., & Ray, M. L. (1972). Measure validation in marketing. *Journal of Marketing Research*, 9(4), 361–370.
- Jacoby, J. (1978). Consumer Research: How valid and useful are all our consumer behavior research findings? A state of the art review. *Journal of Marketing*, 42(2), 87–96.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16(1), 39–59.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5(5), 602–619.
- Ledgerwood, A., & Shiffrin, R. (2018). *Would preregistration speed or slow the progress of science? A debate with Richard Shiffrin*. Incurably Nuanced blog post, Nov. 8.
- Lee, Byung Cheol, Moorman, C., Moreau, C. P., Stephen, A.T., & Lehmann, D.R. (2020). The Past, Present, and Future of Innovation Research. *Marketing Letters*, <https://doi.org/10.1007/s11002-020-09528-6>.
- Lehmann, D. R., & Hulbert, J. (1972). Are three-point scales always good enough? *Journal of Marketing Research*, 9(4), 444–446.
- Lehmann, D. R., Moore, W. L., & Elrod, T. (1982). The development of distinct choice process segments over time: A stochastic modeling approach. *Journal of Marketing*, 46(2), 48–59.
- Lynch Jr., J. G., Alba, J. W., Krishna, A., Morwitz, V. G., & Gürhan-Canli, Z. (2012). Knowledge creation in consumer research: Multiple routes, multiple criteria. *Journal of Consumer Psychology*, 22(4), 473–485.
- Lynch Jr., J. G. (1982). On the external validity of experiments in consumer research. *Journal of Consumer Research*, 9(3), 225–239.
- Lynch Jr., J. G. (1999). Theory and external validity. *Journal of the Academy of Marketing Science*, 27(3), 367–376.
- Lynch Jr., J. G., Bradlow, E. T., Huber, J. C., & Lehmann, D. R. (2015). Reflections on the replication corner: In praise of conceptual replications. *International Journal of Research in Marketing*, 32(4), 333–342.
- Mahajan, V., Muller, E., & Bass, F. M. (1990). New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, 54(1), 1–26.
- Malter, M. S., Holbrook, M. B., Kahn, E. B., Parker, J. R., & Lehmann, D. R. (2020). The Past, Present, and Future of Consumer Research. *Marketing Letters*, <https://doi.org/10.1007/s11002-020-09526-8>.
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521–543.
- Oblander, E. S., Gupta, S., Mela, C. F., Winer, R. S., & Lehmann, D. R. (2020). The past, present, and future of customer management. *Marketing Letters*, 31(2) Xx–Xx.
- Oh, T. T., Keller, K. L., Neslin, S.A., Reibstein, D.J., & Lehmann, D.R. (2020). The Past, Present, and Future of brand research. *Marketing Letters*, <https://doi.org/10.1007/s11002-020-09524-w>.
- Peter, J. P. (1979). Reliability: A review of psychometric basics and recent marketing practices. *Journal of Marketing Research*, 16(1), 6–17.
- Pieters, R. (2017). Meaningful mediation analysis: Plausible causal inference and informative communication. *Journal of Consumer Research*, 44(3), 692–716.
- Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research*, 20(2), 134–148.
- Rust, R.T. & Cooil B. (1994), “Reliability measures for qualitative data: Theory and implications,” *Journal of Marketing Research*, 31(2), 1-14.

- Schwartz, E. M., Bradlow, E. T., & Fader, P. S. (2017). Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36(4), 500–522.
- Shapiro, C. (1983). Premiums for high quality products as returns to reputations. *The Quarterly Journal of Economics*, 98(4), 659–679.
- Sharpe Wessling, K., Huber, J., & Netzer, O. (2017). MTurk character misrepresentation: Assessment and solutions. *Journal of Consumer Research*, 44(1), 211–230.
- Smith, W. R. (1956). Product differentiation and market segmentation as alternative marketing strategies. *Journal of Marketing*, 21(1), 3–8.
- Sozuer, S., Carpenter, G. S., Kopalle, P. K., McAlister, L. M., & Lehmann, D. R. (2020). The Past, Present, and Future of Marketing Strategy. *Marketing Letters*, <https://doi.org/10.1007/s11002-020-09529-5>.
- Stahl, F., Heitmann, M., Lehmann, D. R., & Neslin, S. A. (2012). The impact of brand equity on customer acquisition, retention, and profit margin. *Journal of Marketing*, 76(4), 44–63.
- Wedel, M., & Kannan, P. K. (2016). Marketing analytics for data-rich environments. *Journal of Marketing*, 80(6), 97–121.
- Zhao, X., Lynch Jr., J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197–206.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Yu Ding<sup>1</sup> · Wayne S. DeSarbo<sup>2</sup> · Dominique M. Hanssens<sup>3</sup> · Kamel Jedidi<sup>1</sup> · John G. Lynch Jr.<sup>4</sup> · Donald R. Lehmann<sup>1</sup>

Wayne S. DeSarbo  
wsd6@psu.edu

Dominique M. Hanssens  
dominique.hanssens@anderson.ucla.edu

Kamel Jedidi  
kj7@columbia.edu

John G. Lynch, Jr  
john.g.lynch@colorado.edu

Donald R. Lehmann  
drl2@columbia.edu

<sup>1</sup> Graduate School of Business, Columbia University, 3022 Broadway, New York, NY 10027, USA

<sup>2</sup> Smeal College of Business, Pennsylvania State University, 210 Business Building, University Park, PA 16802, USA

<sup>3</sup> Anderson School of Management, University of California, Los Angeles, 110 Westwood Plaza, Los Angeles, CA 90095, USA

<sup>4</sup> Leeds School of Business, University of Colorado-Boulder, 995 Regent Dr, Boulder, CO 80309, USA