International Journal of Research in Marketing xxx (xxxx) xxx

Contents lists available at ScienceDirect

IJRM
International Journal of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

The evolving world of research in marketing and the blending of theory and data☆

Donald R. Lehmann

Columbia University, Marketing, 507 Uris Hall, New York, NY 10027, United States

ARTICLE INFO

Article history:
First received on July 25, 2019 and was under review for 2 months
Available online xxxx

Guest Editor: Gary L. Lilien

ABSTRACT

The world, and the way research is conducted in marketing, has changed over the last fifty years. For example, there has been an increased emphasis on theory and model testing. This paper briefly identifies some of the changes that have occurred. It then discusses the role of theory and presents a method for combining theory (intuition, expectations) with data to form a model. Finally, it identifies a number of concerns with the way research is typically conducted, most of which involve "formulaic" approaches that have developed. The purpose of the paper is not to present a correct approach. Rather it is to have all of us, but especially young scholars, think more critically (and imaginatively) about how to advance knowledge in marketing.

This paper begins with a brief discussion of how academic marketing has changed in the last 50 years. It next discusses the role of theory vs. data and outlines a specific procedure for formally blending the two in the context of developing a model linking multiple constructs. This is followed by a discussion of what I see as some issues with the way we conduct research and some positive observations about our field. The goal is to help researchers, particular young ones, think critically about the research conventions which have developed in marketing over the last half century.

☆ When I was informed I had been named an EMAC Distinguished Scholar I was surprised, honored, and unsure why I, and not several others, had been selected. The euphoria (and gratitude) began to wear off ever so slightly when I was informed that I should come to the meeting to accept the award (which actually became a positive as I visited Glasgow). I then was informed of the "creeping commitment": I was told I needed to make a presentation and later that I should write a paper for IJRM, i.e., it was no free lunch/conference. This paper is based partly on the morning talk I gave in Glasgow when many of those in the intended audience (young scholars) were (understandably) not yet at the conference or attending other sessions.

https://doi.org/10.1016/j.ijresmar.2019.12.001

Table 1
50 years of macro changes.

	Old World (1969)	New World (2019)
Travel Planning	How: Buy or Own Map; read, plan Who: Travel Agent	GPS; Verbal directions Self (Kayak, etc.)
Information	Own Books, Library, Inter-library loan Research librarian	Google
Communicate	Letter, Phone, In person	Email, Text, Facebook, …
Calculations	Memory/flash cards/slide rule	Computer/Phone
Analysis	Flow Chart, Program, Punch Cards, wait, simple statistics	Click an icon (Solver), data mining
Reading	Book, Newspaper	Online, Kindle
News Delay	Day (newspaper), Weeks (magazines)	Nano seconds
Report Writing	Index cards, hand write, multiple drafts, dictionary	Electronic Folders, Power Point Slides, Word, Spell-check
Learning	Patience, planning, in person	Immediate results/gratification, impersonal
Meeting	In Person	Teleconference, Skype, Dropbox
Intercontinental Travel	Jet Plane Boeing 707	Jet Plane Boeing 787

1. Preamble[1]

The 1960s were an interesting decade. In the broader society, an increasingly unpopular war sparked protests, a number of iconic groups emerged (Beatles, Rolling Stones, …), and space exploration became a realized goal. Business schools were entering a growth phase. Concern about what was taught in them centered on over-emphasis on specific industries and too little focus on research, theory, and rigor, driven partly by reports from the Ford and Carnegie Foundations. Ironically, this is the opposite of the current push for relevance. In the world of retailing, malls were emerging as go-to places with negative impacts on small business. Direct marketing was mail-based with Sears (which began selling watches and jewelry) catalog business the rough equivalent of Amazon online (which began selling books) today. Since then, the landscape has changed in terms of technology, the academic marketplace, and how research in marketing is conducted.

1.1. Technologies

The technologies which dominated various facets of life and are available to researchers have changed noticeably. TV consisted essentially of 3 network channels delivered over the air (with varying degrees of reception quality). Individual-level scanner data was not yet widely available (when it was, coffee data from Pittsfield, Massachusetts became a popular basis of analysis). In its place were sources such as SAMI monthly warehouse withdrawals (shipments to retail establishments). Computers were big, expensive, and far less powerful. The standard was to write Fortran programs, punch them into cards, carry them to a central location, and then wait overnight for the output (often to find out that something was wrong in card 15 which caused the program to abort). There were no Amazon Turks and finding participants (then called subjects) revolved around either central locations like waiting areas or enlisting church and similar groups. For analytical (i.e., math/algebra-based) papers, all proofs were done by hand, i.e., there were no Mathematica or similar programs and simulations were rarely done because of the strain they put on available computer power. (One month I managed to be the largest user of computer time in the university by running simulations). Content analysis could only be done manually; text analysis was in the distant future. While limiting, the lack of easy-to-collect data and computing power had the benefit of encouraging careful thinking before collecting data, something which currently is often lacking.

Other notable technology changes include those shown in Table 1. Interestingly, one area has not changed much over the last 50 years. While air travel has become more common (and informal), there isn't much difference between a Boeing 707 and a 787 except capacity (and often less leg room). Travel to conferences is still less than (and less) enjoyable.

[1] In 1969, I joined Columbia partly because it was a place I thought I could move from (to industry or another school) in spite of the fact that at the time, New York City was not in great shape, the area around Columbia (more precisely North, East, and South of campus – West is the Hudson River) was a bit sketchy (read dangerous) and I was (and I still am) more interested in sports and outdoor activities than plays and museums. Nonetheless, I ended up spending 50 years as an academic at the same school, using the same desk.

I had planned to accept the only job I was offered, programming (now called coding) in Fortran for GE in Pittsfield, MA when a professor at Union told me they were giving fellowships to go to graduate school in business including National Defense Education Act ones developed in large measure as a reaction to the launch of Sputnik by the Soviet Union. This seemed like a good deal to me so I applied to several, got into some, and settled on Purdue. This was extremely fortunate for me because it was the beginning of a decade where it (led by Frank Bass and Edgar (Mike) Pessemier) produced numerous top scholars. I started in finance but switched to marketing because a) being good in math (or at least having majored in it) was not a big advantage in finance and b) marketing seemed like an open field which allowed more flexibility in topic and method. One important factor which led me to the PhD program was that it was short: I spent the first year in the 11 month MSIA program and a total of two years and nine months at Purdue before receiving my degree.

The job market was strong (the benefit of being in the growth/takeoff phase of an industry) although not for jobs in industry. Being unsure what I wanted to do, I also applied for "real" jobs, completely unsuccessfully. The closest I got was during an interview with a "human resource" person at GE (where my father, grandfather, and great uncle worked). After explaining I wanted to have a "real" job (and be paid or fired based on my performance in it), he said there might be an opening in a small group being formed to explore alternative energy sources. My less-than-prescient response of "Gas is 20 cents a gallon, why do we need alternative energy sources?" effectively ended that option.

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

1.2. The academic marketplace[2]

When I "went on the job market", there were no expectations of having published or papers in the review process. What I had (which was typical) was overhead slides describing what I planned to do. The push to have papers and hopefully some published ones before going on the job market was well in the future. In addition, doctoral students attended few if any conferences prior to going on the job market. For those who attended the AMA Doctoral Consortium, it was often their first conference. Because it was held contiguously with the Summer AMA Conference, it made interviewing even more of an endurance contest.

1.3. Research approaches in marketing

Over time, research in the various subfields of marketing has tended to become specialized and, if not standardized, a bit "formulaic". This specialization has advantages for making it easier to compare work and for specialists/experts to communicate within their particular community, although not across communities due to the "shorthand"/terminology used. Unfortunately, it also confounds the method used with results, making it difficult to separate method effects from "real" ones.

A related problem is the tendency to classify and perform research based on method rather than topic. The focus on single methods discourages "triangulation" by multiple different methods and has led to a tendency to value methodology over substance. This trend has become more pronounced over time. At a high level, the lines between behavioral, quantitative, and strategy research have become sharper with members of each "tribe" talking almost exclusively to, writing for, and attending seminars of members of their group. This separation has led to the development of parallel recruiting processes, lessening the chance that a boundary-spanning PhD student will be hired. It also has led to the development of sub-groups defined by method rather than substance, e.g., experimenters, structural modelers, CCT researchers, and, recently, data analysts. Specializing in a method has value; one becomes more skilled in applying a method as they gain more experience in using it. However, defining oneself solely based on one method limits the extent and communicability of one's contributions and lessens the chance of finding innovative results. The consequences of the reliance on a single method are not good for either science or practice.

2. The role of "theory"

In contrast to the 1960s and 70s, the review process in marketing has increasingly insisted that articles have a "theory". Indeed, having a theory is often viewed as the sine qua non of good research. An obvious question is why. Even assuming the goal is to develop theory relevant to marketing, it does not follow that every paper has to have a theory (typically discussed by summarizing past research over multiple pages) as its starting point.

At its core, a theory is a story/explanation for some situation or phenomenon and is agnostic with respect to whether it precedes or follows data collection and analysis. No research is without some theory at least with respect to what variables will be explored, which fall into four categories.

1. Focal/key/theoretical (often manipulated)
2. Measured and analyzed (often as covariates)
3. Controlled for (e.g., by selection, matching, fixed effects)
4. Ignored

The fourth category is essentially infinite (making any study incomplete), meaning you can't (and it is not worthwhile to try to) fully enumerate it. Nevertheless, specifying the variables you considered and rejected (e.g., due to data availability or cost), helps both the author and the reader better understand study limitations (and provides directions for future research). When included early in a presentation or paper, this can help reduce questions and counter-arguing by seminar participants and reviewers.

Currently, theory is often operationalized as something published in a base discipline (typically psychology for behavioral and economics for quantitative research). These fields are certainly a reasonable place to look to for theories but doing so typically makes the research that results a replication (e.g., applying regulatory focus to financial decisions). As one of the original four editors of the Replication Corner at IJRM (and now JMB), I understand and appreciate the value of replications. Still it is important to recognize that this is what they are.

[2] At first, I planned to do an analytical dissertation based on media (TV) advertising scheduling, then a "hot" topic. At that time many agencies claimed to (and many actually did) use linear programming to allocate spending but I was having difficulty deciding exactly what to maximize. One day when I spoke to Professor Bass (he was typically referred to as Dr. Bass, hence my use of professor was somewhat informal), he told me two important things. First, if you start working on an analytical model, you don't know when (or if) you will be finished, but if you do something empirical, you could plan (PERT chart) it and know when you would be finished. Being married with a son who was eating and growing, that was pretty persuasive on its own. Second, he said he was planning to conduct a large scale survey with Market Facts (partly for the dissertation of Wayne Talarzyk) which I could join. Hence I converted to studying choice via a multi-attribute model. Given that I didn't want to waste the literature review I had done over the Christmas holiday on TV shows, I focused on viewer choices of TV shows.

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

The most irritating aspect of the demand for theory is the difficulty in getting novel work published. When reviewers insist on a theoretical basis (defined as something established by prior research), they effectively rule out innovative explorations. What makes work novel is that it doesn't have a clear predecessor. The absence of a clear predecessor results in sometimes convoluted efforts to fit what someone is trying to do into a particular "theory" and encourages the recycling of old topics (and methods) while at the same time highlighting something "novel and interesting" about the research.

Different research, and different researchers, at least implicitly view the role of theory in empirical research quite differently (See Table 2). To a person with a religious prior (i.e., one who is certain a particular theory is correct), data is mainly a nuisance, something to point to when it supports the theory and something to be ignored or discredited when it does not. A softer version of this position is a prior theory dominant viewpoint in which a person has a strong prior that a single theory is true. Because they believe only a single theory is "true", they collect data in order to either "confirm" or calibrate the theory. Alternative explanations are generally either lightly explored or ignored altogether. (The analytical quantitative equivalent is a person who has a single set of assumptions and takes as results whatever they lead to algebraically, i.e., does no serious assumption sensitivity testing).

At the other end of the spectrum is the pure data analyst. To a data miner, nothing is expected or unexpected; it all depends on the data and the relations among the variables in it. They emphasize prediction (pattern recognition) and "let the data speak for itself" with no prior thought about what will be found or desire to explain/generalize it. (In some ways this approach is similar to that of qualitative researchers who provide detailed accounts/"thick descriptions" of a behavior or pattern but make no attempt to generalize/create a model based on their efforts.)

Close to this approach is that of researchers who desire parsimony but are heavily data driven. They often use stepwise procedures (in spite of the local optimum problems they can encounter) with the stopping rule based on statistical significance, basically treating the cost of including a non-significant variable as infinite, or based on the size of the improvement in fit (e.g., stress, AIC).

The middle category most closely matches the mental model of many researchers (and experts) and is the most difficult to implement. It essentially has "weak priors" and requires exploring alternative models and variable operationalizations (e.g., via alternative model testing). It is philosophically "Bayesian" and requires treating research as a craft which both has certain soft "rules" and is flexible and open to discovery.

Requiring research to follow either of the extreme approaches in Table 2 rather than encouraging both, or combinations of them, is likely to slow, rather than facilitate, research progress. The approach presented in the next section provides an outline of a systematic framework to implement the middle option approach in the context of model development/creation.

3. Model development via combining theory/intuition and data

The emergence of big data, analytics, and AI provides the means to focus on prediction without considering theory, the last approach in Table 2 (of course one can post hoc develop theory based on the results). Thus, it seems worthwhile to see how to do some "analytics" and still include intuition and prior conceptions (theory) as an important component.

As discussed earlier, there are two distinct, and to some extent opposing, approaches to research and data analysis. One focuses on understanding/explaining and has an explanation/story/theory as both the goal and often the starting point. Theory testing essentially starts with a story/explanation and then exposes it to data/empirical testing. This is the TETE approach described by Bass (1995) which begins with theory (T), examines data to see if it is consistent with it (E), revises the theory when the empirical data is so strong that it makes the theory "untenable", and repeats the process.

The other approach begins with empirical data (E), creates a story/theory consistent with it, and repeats the process (ETET). Whereas in the first approach, theory is "king", in the latter, data is given more weight. The difference thus involves the relative weight one puts on data versus theory. Importantly, as the cycle is repeated many times the starting point (theory or data) becomes increasingly unimportant (i.e., the "prior" is overwhelmed by new data).

Recently, a third approach has emerged which is entirely prediction/data driven. The use of big data and data analytics have made it possible to make decisions or predictions without understanding the process driving them, i.e., "why" a prediction is being made. While at some point a purely data-driven approach may be sufficient, e.g., for routine repeated decisions (Bucklin, Lehmann, & Little, 1998), for many situations blending theory (or intuition) and data seems most appropriate (Blattberg &

Table 2
Relating theory and data.

Position	Consequences for Empirical Analyses
"Religious" prior; One theory is correct	No evidence can alter the prior (Data is irrelevant)
Theory first: research "tests" theories	Estimate one model, look for/expect confirmation
Soft prior	Willing to trade-off fit and prior conceptualization
Agnostic but desire parsimony	Data analysis with penalty for including variables
No theory/prior	Pure data analysis (Theory is irrelevant)

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

Hock, 1990), basically what Alba (2011) referred to as “bumbling around”. The purpose of this section is to provide the outline of an approach which can be largely automated that explicitly incorporates user priors about what relations exist, drawing partly upon some of my earlier (and little noticed) work (Lehmann, 1980, 1983). Its goal is to create a model describing the relations among a set of variables for which data is available or can be obtained.

3.1. Step 1

The first step is to assess the extent to which each variable is related to each other variable. Ignoring statistical niceties (e.g., dealing with categorical variables), the standard approach has been to rely on the Pearson correlation coefficient. The first modification here is to look at relations more broadly. Given the existence of non-linear relations, it can be helpful to use polynomials to identify how pairs of variables (A, B) are related to each other. A simple approach (Lehmann, 1974) is to run a cubic (3rd order polynomial) regression of A as a predictor of B and vice-versa and use the square root of the larger R^2 as the measure of the strength of their relation. Whichever regression result produces the larger correlation (A to B or B to A) suggests the direction of any causal relation between them. While for a monotonic relation, the difference is not very useful (and for a linear relation there isn't one), for U or inverted-U shaped relations, one direction may produce a noticeable correlation and the other (depending on the range of the data) close to zero, thus giving a strong indication of which causes which.

Another less immediately obvious consequence of using cubic relations is that this allows for even higher order polynomial (i.e., complex) relations between indirectly linked variables. For example, if B is a cubic function of A and C is a cubic function of B, then C can be a ninth order polynomial function of A. This means the approach can uncover subtle relations, albeit at the risk of being overly influenced by outliers and error.

3.2. Step 2

The second step is to create an overall matrix describing relations (similarities) among all variables (basically the larger of the two correlations described in step 1) and to get an overall sense of how they fit together. For many situations it is useful to display the relations among variables graphically. If one wants to graphically portray how the variables relate to each other, one can use standard multidimensional scaling algorithms which put similar (highly related) pairs of objects (here variables) close to each other and array the variables in an n-dimensional space (Lehmann, 1972).

Establishing linkages among the variables can be done using the principles of path analysis. For example, if there are only three variables and one (B) fully mediates effect of one (A) on the another (C), a one-dimension solution will suffice, i.e., it will fully account for the correlations among them as shown here.

$$\mathbf{A} \longrightarrow \mathbf{B} \longrightarrow \mathbf{C}$$

When mediation is only partial (i.e., there is a direct link from A to B), and there is another driver of C (D), a two-dimensional solution will fit better:

$$\begin{tikzcd} \mathbf{A} \arrow[r] \arrow[rr, bend left] & \mathbf{B} \arrow[r] & \mathbf{C} \\ & \mathbf{D} \arrow[ur] & \end{tikzcd}$$

More complex relations may require higher dimensional spaces. One can use the improvement in fit (reduction in “stress”) as the number of dimensions increases as a way to develop the “optimal” number of dimensions.

3.3. Step 3

Step 3 explicitly links those pairs of variables that are most related. To do this, we first create a “penalty” function which consists of the sum of the squared correlations in step one (the R^2s of the polynomial regressions described earlier) across all possible pairs of variables. Initially no link is included in the model and thus the squared correlations are in essence unexplained “errors”. Next we sequentially allow for (add) the link of the not-yet-linked pair of variables that best helps improve prediction of the correlation matrix. Each time a link is added, the system of regressions implied by the links in the model are run and the change in the error term $\sum (R_{ij} - R_{ij})^2$ is calculated. Whichever new link creates the largest decrease in the error term is then added to the model.

3.4. Step 4

If step 3 is done without regard to theory/prior expectations, the procedure is purely analytical (i.e., whichever new link creates the largest decrease in the error term is added to the model as in stepwise regression). However, one can incorporate expectations by creating a penalty function that captures them. One way to create the penalty function is to ask experts or intended model users how much they do, or do not, want or expect to see each pair of variables linked in the model on a “cost” scale from +1 (having the variables linked is totally unacceptable) to 0 (indifference, the implicit assumption of a purely data-driven approach) to −1 (the link must be included). Essentially the −1 and +1 are “religious priors”, i.e., beliefs held with certainty and

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

unsusceptible to change regardless of whether the data indicates that a particular link does or does not belong in the model. In essence this input "pre-registers" intuition/expectations.

3.5. Step 5

The next step is to formally trade-off the fit criterion with the human expectation/intuition one. This can be accomplished by optimizing a function which trades off the two sources of error:

$$K \sum^{\text{All pairs}} (R_{ij}\text{-}R_{ij})^2 + (1\text{-}K) \sum^{\text{All pairs}} Z_{ij} D_{ij}$$

Where $D_{ij} = 1$ if link$_{ij}$ is included, 0 otherwise, and K captures the relative weight put on data versus prior intuition.

For unlinked (directly or indirectly) pairs, $R_{ij} = 0$.

Here, Z is the rating of the "cost" of i linking to j. One way to generate Zs is to ask a person (or persons) to indicate the cost of including each link on a − 1 (meaning the link should definitely be included) to +1 (meaning the link should definitely not be included) scale. This makes the cost of omitting a link which we are sure belongs in the model equal to the cost due to ignoring a relation between two perfectly correlated variables because $(R_{ij} - R_{ij})^2 = (0–1)^2 = 1$. Looked at differently, intuition, i.e., the non-zero Zs, are essentially informative priors in a Bayesian sense. If sufficient prior work exists, one could use results of a meta-analysis on the relations between pairs of variables to create the Z values based on significance levels (e.g., the percent of the time each relation was significant or a pooled *p*-value) or the magnitude of the relations (e.g., effect sizes measures such as Cohen's d).

If K = 1, the approach will be purely data driven, i.e., the person has no confidence in/opinion on what relations belong in the model. If K = 0, then only the opinion of the person (or persons) who supplied the costs of violating their priors matters, i.e., they are completely confident, thus effectively requiring considering only the model the person believes is correct (and estimating its parameters). Because K is typically between 0 and 1, one can either pick a K or run the analysis for a number of values of K and then select the solution that looks best (essentially allowing for a second way to incorporate judgment), similar to the approach used by Netzer and Srinivasan (2011) to combine results of a model and a heuristic. In order to limit the influence of overconfident users, one can constrain K to be "reasonable", e.g., at least 0.2. Alternatively, you can limit the total number of non-zero Zs.

Incorporating causal direction is more difficult. Searching for causal relations among a set of variables is an important problem in machine learning (e.g., see Spirtes et al., 2000, Mooij, Peters, Janzing, Zscheischler, & Schölkopf, 2016). Recently work has appeared suggesting the use of which direction fits better to determine causal direction (Janzing, 2008, Shalizi, 2017, Blobaum, Janzing, Washio, Shimizu, & Schoelkopf, 2018). Similarly, the usefulness of examining non-linear relations for determining causal directionality has been noted by work in the machine learning literature (e.g., Hoyer, Janzing, Mooij, Peters, & Schölkopf, 2009). Here we combine some of these concepts with prior conceptions (theories, intuition) of relationships to blend theory and data (Lehmann, 1983).

One can simply ask the user (or an expert) to specify the causal direction. Alternatively, we can rely on "clues" in the data to determine directionality. For example, if X_1 fits a cubic function of X_2 better than X_2 fits a cubic function of X_1, then it seems likely that X_2 leads to/causes X_1. In essence this uses Granger causation to make the choice of causal direction which is similar in spirit to using cross-lag correlations. This approach is related to reverse regression, approaches based on comparing the probabilities of $P(X_1|X_2)$ and $P(X_2|X_1)$, and work on directed graphs. What is different here is the use of cubic regressions and the possibility of including intuition in the form of making the Z matrix non-symmetric. The latter can be done by assigning a lower cost to the preferred direction, e.g., making $|Z_{ij}| < |Z_{ji}|$.

Importantly, the previous steps closely parallel what an expert does instinctively (and how science often progresses), adapting current theory to new facts and data in a Bayesian-like manner. This approach fundamentally differs from either pure "theory testing" which essentially declares a theory true/good depending on whether *p*-values are less than .05, or data-driven approaches which ignore intuition and, for the most part, prior results. Put differently, it mechanizes "bumbling around".

This approach has notable similarities to earlier work on model comparisons which used Bayesian methods to include (informative) priors (e.g., Allenby, Arora, & Ginter, 1995; Fornell & Rust, 1989; Lenk & Orme, 2009; Rust & Schmittlein, 1985; Rust, Simester, Brodie, & Nilikant, 1995). It differs from that work in that priors are used in the development of a model rather than in the comparison of already specified models.

3.6. Example

The approach is best illustrated with a small example. Table 3 shows the correlations among four variables. We use it to stepwise build a model by adding one link to the model at a time. Table 4 demonstrates how the model evolves for the pure data driven approach (K = 1). It begins with step 1 which adds a link between A and D because they have the highest correlation of any of the pairs of variables. Step 2 then adds a link between A and B. Note that Step 3 produces a bigger improvement in fit than step 2 by linking B to C because it not only accounts for the link between B and C but also partly captures the link between A and C operating via B, i.e., considers the model as a whole. Also, notice the large drop-off in the improvement in fit for any of the options in step 4. This suggests incorporating a minimal "cost" of adding a link, for example, the square of

Table 3
Example.

Correlations				
	A	B	C	D
A		.6	.4	.7
B			.5	.1
C				.2
D				

Table 4
Data only results.

	Link added	Improvement index ($\sum(\text{R-R})^2$)
Step 1	A-D	.49
Step 2	A-B	.36
Step 3	B-C	$.25 + .4^2 = .41$
	D-C(C-D)	$(.04)^2 = .0016$
Step 4	B-D	$(.01)^2 = .0001$
	A-C (Direct Link)	$.1^2 = .01$

the minimum size correlation which would be significant, would lead to stopping here. In "big data", of course, the sample size tends to be huge which makes a statistical significance based cutoff very easy to reach (essentially every non-zero correlation is significant).

Table 5 modifies the example by incorporating intuition/prior theory and weighting this and data equally (K = 0.5). In this case the priors are that A leads to B which in turn leads to C (costs = −0.5) while the person is indifferent with respect to whether A is linked to D, mildly prefers not to include the A to C link (cost = +0.1) and doesn't want to include B to D or C to D links (cost = +0.5). In this case, A to B is the first link added, followed by B to C in step two and A to D in step 3. This demonstrates how the priors (costs) can change the results.

In the extreme case, a theorist might only want A to B and B to C links, i.e., believe A influences C fully mediated by B (with costs = −1). In this case, all the other links would be excluded because the cost (+1) would generally overwhelm any fit advantage derived from including them. Thus, the direct effect of A to C would not be explored and the unexpected link of A to D would not be discovered.

Importantly, the cost matrix (Z) can be asymmetric. Assume a researcher (based on theory and/or prior research) thought A impacted C mediated by B and that D would be difficult to measure (e.g., data were unavailable on it). They could then have a cost matrix (Z) like Table 6. The matrix is asymmetric such that it captures different in strengths the preference for a model in which A leads to B rather than the opposite. Here, the cost of having C lead to A is high (cost = 0.8) while the possibility of a direct link of A to C is mildly encouraged (cost = −0.1). These costs also "discourage" any model which includes links from or to D by assigning a positive cost of 0.2 for including it. The cost function thus provides flexibility in incorporating "priors" in the model. It also suggests that there is some advantage to "pre-registering" a Z matrix as well as sensitivity analysis to see how robust the resulting model is to different Zs and Ks.

3.7. Summary

To summarize, this paper has outlined an approach for formally combining intuition and/or theory or data cost considerations (via the Z matrix) with data fit/prediction ones (based on the "correlation" matrix where the correlations can be nonlinear). The intent here is not to fully develop and test this procedure which, when more fully developed (for example, by

Table 5
Results Including "Intuition"/Theory.
Assume K = 0.5 (Equally weight intuition and data).

	Cost	Benefit: Step 1	Step 2
A-B	-0.5	$.36 + .5 = .86$	0
A-C	+0.1	$(.16 - .1) = .06$	.06
A-D	0	.49	.49
B-C	-0.5	$.5 + .25 = .75$	$.75 + .16 = .91$
B-D	+0.5	$.01 - .5 = -.49$	−.49
C-D	+0.5	$.04 - .5 = -.46$	−.46
Link Added		A − B	B − C

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

Table 6
An asymmetric cost matrix.

		Effect			
		A	B	C	D
Cause	A		−.5	.1	0
	B	.8		−.5	.5
	C	.8	.5		.5
	D	.2	.2	.2	

accounting for heterogenicity/segments, a non-trivial problem) can be a useful tool to aid informed users. That is a work in progress.

In addition, it seems that often such tools are applied formulaically to ward off questions about what they do by researchers not fully aware what they do and do not do. Consistent with the overall theme of this paper to encourage less unquestioned use of "established" procedures, therefore, specific "code" is not provided here (although it is being developed). Nonetheless, my hope is that thinking about this approach will encourage a moderate position on the role of theory versus data.

4. General concerns

The previous section suggests that strong theory is not always available and that requiring it can discourage discovery. This section highlights a few other concerns with current research practice.

4.1. Formulaic analysis

As mentioned earlier, over time research (and researchers) in marketing has (have) tended to segment itself (themselves), first at a high level (behavioral, quantitative, managerial/strategic) and then into more tightly defined subfields. As a result each sub-field has gravitated toward a particular approach/"right way" of doing research. One consequence is that procedures have become standardized (which decreases generalizability by confounding method with content) and to some extent applied formulaically.

Increasingly analysis seems to be reported (and done by and limited to) the output of "apps". For example, Hayes has developed a number of useful analysis tools. Unfortunately, researchers increasingly report what they have done is "applied Hayes model ____". This practice makes one wonder if the users really know what the procedure is or does and any limitations it may have. It also deprives them of the potential discoveries that can emerge from employing multiple (often sequential) analyses. Similarly, I wonder how well the users of the metaphor program for meta-analysis appreciate the subjective aspects of variable selection and combination and how they influence the eventual results.

Part of my skepticism for standard solutions comes from two experiences. One involved my grandmother playing bridge with a leading expert. At the end of one hand she described the hand and noted he had played it differently than he had suggested in his book. His response, something like "did I write that?" suggests experts are not entirely routinized. The other experience was when I was trying to generate a multidimensional scaling picture which integrated foods, benefits, and parts of the body. I was unable to get a satisfactory (non-degenerative) solution. When Joe Kruskal was at Columbia to give a seminar, I asked for his help in using KYST (named for Kruskal, Young, Shepard, and Torgerson), the "go to" program at the time. He first asked to see the raw similarity data and then what the variables were. After a quick look, he said it looked like I had basically replicated the four basic food groups. When I said that seemed to be the case, he then responded something like "Then why do you want to scale them?" Here, the real expert was neither formulaic nor enamored of the most advanced (or their own) method. Since then I have noticed that most real experts are not formulaic. Rather it is their disciples/students who tend to have rigid views on the right way to do things.

4.2. The search for the "perfect" paper versus contributing to a future meta-analysis

The profession, and the increasingly directive and legalistic review process, continues to push for "more" theory, data, and analyses. The implicit assumption behind this is that having fewer more rigorous papers is better. Obviously directionally this push is correct up to a point. An important question, however, is whether at some point this decreases rather than helps advance the body of knowledge in an area.

There are at least four main problems with the search for a perfect paper approach. First, no paper is or can be "perfect". As the President of the Ford Foundation said, "we cannot let the perfect be the enemy of progress" (Walker, 2019). Empirical knowledge requires generalizations which can only be made with any real confidence based on conceptual (imperfect) replications and meta-analyses. Simply repeating the same procedure confounds any results with the method used, authors (if they are the same ones), etc. The field advances as the size of the effect (vs. effect size, i.e., correlation) of one variable on another becomes better calibrated. (In that regard, the trend toward a focus on numerous statistics – I^2, Q, fail-safe n versus regression coefficients in a

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

multivariate model of the size of the effects of the phenomenon of interest focuses too much attention on whether research is "right" versus what it tells us.)

The second concern is that enforcement of an (almost) perfect criterion tends to exclude truly innovative work (i.e., if something is really new, then no prior theory explains it). As the procedure described in the previous section suggests, less "formal" research has real value.

Another consequence of the perfect paper concept is on the review process. The review process increasingly resembles a legal proceeding, with the reviewers playing the role of prosecutors. First, expectation levels are overly high. While it would be nice if all papers made genuine theoretical and methodological contributions as well as had surprising findings, it is not reasonable to expect them to (e.g., if results match an existing theory, then they can't be surprising). A more reasonable criterion is whether they either contribute to a future meta-analysis (conceptual or quantitative) on the topic or are likely to spark future research. Second, it is often not clear that multiple rounds of reviews produce enough improvement to justify the effort and delay in publishing entailed. Remembering that the paper belongs to (and is identified by) the author is a good place to start, i.e., reviewer comments should be considered as suggestions which editors (and authors) are not always required to follow.

A fourth and related concern is that the pressure to produce perfect papers has contributed to many of the current concerns about research practices. For example, if based on the design of a study we would be interested in the results of it, then results don't have to be novel or statistically significant to be useful. This view would make the incentive to "p-hack" or hide studies small at most and free the profession from searches for possible p-hacks and the need to pre-register studies. It would instead put the emphasis on design (vs. results) and replications. (In the extreme, papers could be reviewed based solely on their design.)

4.3. Obsession over statistical significance (p value)

One consequence of the search for (nearly) perfect papers is the fixation on statistical significance. For most situations what we most want to know is how large the effect of something (X) is on something else (Y). This question has two components: the average effect (i.e., the size of the effect) and the range of possible effects (typically captured based on the standard deviation), i.e., the degree to which we can count on getting the average effect which can be portrayed with a "box and whisker" plot.

Currently the field seems overly focused on getting statistical significance at the (two-tailed) .05 level. This practice makes limited sense. First of all, the costs of deciding a treatment is ineffective when in fact it works are rarely equal to the costs of concluding a treatment is effective when in fact it doesn't work. Yet this is typically ignored (i.e., we do not minimize, say, the sum of the two costs times their probabilities as in optimal classification). Similarly, prior theory, logic, or intuition is often ignored since these are frequently directional in nature which implies a one-tail test.

To see the "folly" behind this obsession, consider Fig. 1. Case A has a much tighter correspondence between X and Y than case B and hence a lower significance level. On the other hand, the average size of the effect is much larger in case B. If you were an underdog, or a person dealing with a serious illness, which treatment would you choose? Case B also holds the possibility of finding variables that explain the deviations (e.g., omitted variables including individual differences), thus providing both practical results and opening the possibility of advanced theory development.

Effect size measures such as Cohen's d combine the size of the effect and its variation, e.g., by dividing the difference in treatment means by a measure of its variance or standard deviation. Thus it partially, but only partially, reflects the magnitude of the effect.

The focus on *p* values has had a pernicious (bad) effect on research in marketing. Authors are motivated to choose topics which are likely to produce significant results, choose sample sizes designed to produce statistical significance given an expected effect size, and "hide" non-significant results from reviewers by finding the "right" measures and manipulations. Reviewers exacerbate this tendency by questioning non-significant results. The result has been (understandably) behavior by authors which values publication over discovery, leading to the current furor over p-hacking, and a cottage industry trying to find evidence of it. The obvious solution to this development is to focus more on the size of the effect and downplay (or in the extreme stop reporting) *p* values.

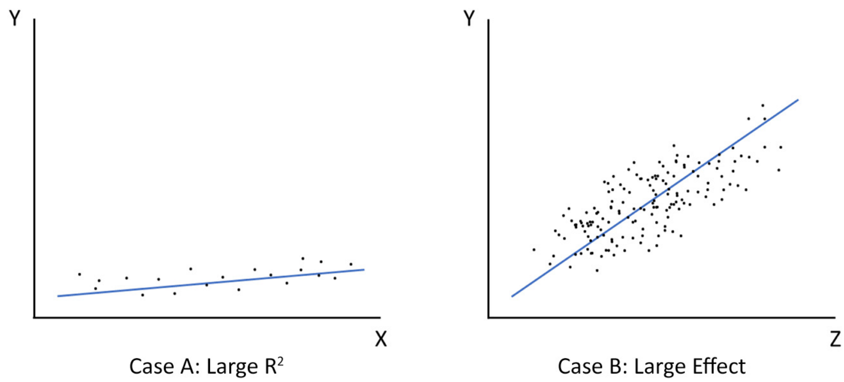

Fig. 1. Large R^2 versus Large Effect.

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

4.4. Replications and meta-analysis

Replications are important for establishing general knowledge. Unfortunately, much of the recent focus on replications is predicated on the assumption that things should exactly replicate, primarily in terms of statistical significance and if they don't, the first paper was "wrong". First, it is more important that the size of an effect is similar; whether it differs significantly from earlier results is a more worthwhile analysis versus whether an effect is non-zero based on a statistical significance test. Second, and more fundamental, perfect replications are impossible; if nothing else time, the position of the earth, etc. constantly change.

What makes sense is to view individual papers as inputs to a future meta-analysis which at some point will summarize them and identify contingencies ("boundary conditions"). Each paper then serves as an imperfect (conceptual) replication (Lynch, Bradlow, Huber, & Lehmann, 2015). Following this logic, subsequent studies can be designed in order to maximize the additional information they add by varying some, but not all, of their characteristics (Farley, Lehmann, & Mann, 1998).

4.5. Collateral damage from applying high-powered methods

Fortunately, empirical quantitative research has increasingly been reporting simple ("model-free") results. These in effect are simple correlations which can be compared with the relations (partial correlations) generated by multivariate methods, i.e., more complex methods. Because more complex methods make assumptions about the nature of the relations and are more sensitive in general, they are also less stable/more likely to be influenced by errors. Put differently, complex methods, especially those which are designed to infer causality, are more precise if their assumptions are correct. Because the assumptions are at best not typically fully tested, what makes sense is to see if the simple and more complex methods produce similar results. If they don't, it is prudent to identify which aspects led to the difference and to consider whether they are in fact correct/supportable.

A related issue for empirical researchers is the sometimes-overzealous pursuit of estimation precision without consideration of the purpose of the estimates. Model purposes range from conceptual organization to precise optimization, each requiring (and benefiting from) different levels of precision (Fig. 2). Given that greater precision imposes greater costs (e.g., in terms of effort, communicability, and, frequently, accepting additional assumptions), many/most model-based empirical analyses do not require the use of the most advanced estimation procedures. Employing unnecessarily complex estimation methods can both divert attention from more pressing concerns (e.g., conceptual rigor, consideration of how precise it is reasonable to expect results to be) and, in general, limits the audience for (and managers willing to act on) any results that emerge.

4.6. Experimental behavioral research

One tenet of behavioral research in marketing is that experiments test causality. While they certainly add evidence of it, they are not as foolproof as their advocates suggest. One obvious issue is generalizability. This is typically acknowledged as a sampling problem, in the past with using student samples (Ferber, 1977; Stem Jr. & Lamb, 1980) and now with using MTurks. There is also an issue with the design. With the exception of some field experiments, the setting at best abstracts from reality. Typically this means a) exposing participants to a stimulus which is hard to avoid attending to due to the absence of other stimuli likely to be present in the "real world" (e.g., phones ringing, people talking, multitasking) and b) measuring the response shortly after exposure to the stimulus. If nothing else this increases the measured size of the effect being examined and can cause people to attend to something they would totally ignore given a reasonable alternative.

Purpose	Numerical Precision Needed
Conceptual /Organizational	None
Exploration/Discovery/Directional Prediction	Low (Sign)
Description	Modest (Sign/One Decimal Point)
Optimization	High (6 Sigma)

Fig. 2. Model purpose and precision needed.

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

Beyond this obvious limitation, there is another important caveat. Assume you use A to manipulate X (for example, by asking participants to list 3 versus 10 reasons for something) and show X is (cor)related to Y. Just because you show via a manipulation check that A manipulates X doesn't mean A doesn't also change an essentially infinite list of other things. In my case being asked to list 10 things leads to fatigue and less care (more error) in answering subsequent questions as well as irritation. It is quite possible the "true" model is thus that A impacts both X and Z (they share a common cause) and that only Z impacts Y. In that case the relation of X to Y is caused by omitted variable bias due to the absence of Z.

A related "rule" is to use the manipulation rather than the measured value of a treatment in the analysis. This seems a bit strange given the measured value is used as a manipulation check to see if the manipulation had the desired effect. (If the measured value is useful for establishing the impact of the treatment, why does it become useless in assessing its impact?) The measured value is more likely to reflect a basic trait which is more stable than the transitory nature of the effect of most manipulations. A better way to decide is to compare the errors associated with the manipulation versus the measured value. In many studies while the manipulation is statistically significantly related to the desired construct, the variation in the variable it manipulates is large in relation to the difference in means between treatment groups and larger than the error variance in the measured value. The main reason to use the manipulation, then, is to control for other variables which could be correlated with the construct being manipulation (i.e., selection bias). However, rather than arguing for which is "correct", a better approach is to run the analysis separately for the manipulation and the measured value. When the results are essentially the same, this strengthens the result. When they are not, additional thinking is called for.

4.7. Analytical/"theory" papers

Analytical papers have a long tradition in marketing research beginning at least with Dorfman and Steiner (1954). The problem is that they a) require simplifying assumptions which are often chosen and altered to produce interpretable/tractable results, and b) are incomplete. (i.e., as Little noted, all models are wrong, some are useful).

Having worked on several analytical papers, I appreciate the problems which come from "complicating" them. Nonetheless, assuming a monopoly or a duopoly with symmetric (equal) competitors, managers who are "strategic" with an infinite time frame, etc. means the results strictly hold in a very limited number of cases. (As an aside, the common "it is easy to show" or "it is well known that" comments are not very convincing. If it is easy to show, then why not actually show/prove it? If it is well known, then what is the contribution?)

One reason for the popularity of analytical models (basically applications of algebra) in the 1960s was the absence of large scale data and computing power. This absence put a premium on simple models. While these can produce interesting results, the results may be best thought of as hypotheses rather than definitive results. As an alternative, it makes sense to focus on building more complete models, varying their assumptions and parameters and calculating results using, for example, agent-based simulations, and then using statistical models (e.g., regression) to summarize and learn how the results depend on them (i.e., doing a meta-analysis with the assumptions and parameter values as the "design" variables).

4.8. Big data, little insights; small data, big insights?

While they are related, there is a fundamental difference between prediction and explanation. For managerial decisions, the ability to predict the consequence of their actions or external events is largely sufficient. By contrast, for academics ("scientists"), understanding/explanation is often paramount.

Big data, and techniques to analyze/deal with it such as machine learning, have focused attention on prediction and uncovering complex relations among variables. In such cases, the analysis can overwhelm efforts at interpretation and attempts to provide a theory to explain the results, producing good predictions (at least in sample) but little insight into the processes which generated the data. Here, big data can lead to little insight. This result is reminiscent of T.S. Eliot's quote from "The Rock": "Where is the wisdom we have lost in knowledge. Where is the knowledge we have lost in information."

Put differently, the use of big data (which is not the fault of the data) and procedures to efficiently analyze them can overwhelm curiosity and the exploration which one is almost forced to apply when data and analytical tools are limited. In that sense it shares the same issues as the use of A.I. in general and opens the possibility that its use will perversely lead to researchers losing their "feel" for topics and less reflection or curiosity in exploring on them.

The output of data mining is, in effect, complex higher-order interactions. Identifying interactions is useful if they exist. Apparent interactions may be influenced by a small number of (or in the extreme one) observation(s). It would be helpful if they indicated how many observations' predictions were noticeably improved by their inclusion, i.e., explored whether they are driven by "outliers". Similarly, it would be useful to compare their results with simpler models (i.e., the relation between X and Y may be a fourth order polynomial but one which can be closely approximated by a second order one). To the extent the results can be approximated by simple models, both communication to managers and the likelihood someone will come up with a story (theory) to explain them is facilitated.

Small data, by contrast, of necessity encourage thinking/speculation about how the results generalize and what they mean, i.e., the theory/rule behind them. Early stage clinical trials in medicine fall in this category. They begin with a theory or hope that some therapy works and then initially try it on a small sample of individuals. The results on even a single patient (small data) can lead to a major insight. Similarly, qualitative research which examines small sample sizes in depth and pilot studies can produce fairly major realizations or speculations. Here, small data leads to big insight.

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

5. Observations on academic careers in marketing

The previous section focused on how to conduct research. This section, by contrast, addresses how we approach conducting it.

5.1. Research strategy

Conventional wisdom (and advice) is that a junior scholar should a) have a research strategy and b) focus on a single area. I have done neither; my approach is largely opportunistic and driven by interest and the interests of others (PhD students, co-authors). The first time I was asked for a research strategy was the day before the committee formed to evaluate my tenure case was scheduled to meet. When asked for my research strategy, I responded I didn't have one. The chair of the committee then said it would be good if I would give them one the next morning. Fortunately, I was able to give the appearance of having one.

Over time, most of my areas of focus have emerged "organically" (new products/diffusion, decision making, marketing metrics, meta-analysis, measurement and analysis methods, and brands) but not because I planned them. This suggests it is not necessary to obsess over a research strategy; rather, simply staying active will lead to a (post-hoc) one. Put differently, while having an area you can classify at least half your work in may help at the time a person is up for tenure, having multiple areas is good for long-run productivity. Forcing all your work to be in a narrow area is likely to lead to marginal contributions, and for many of us, burnout.

Probably the worst strategy for research (as it is for new products) is to follow trends (Goldenberg, Lehmann, & Mazursky, 2001) Even working papers usually appear two years after the authors started working on the topic which gives them a two-year lead on you. If you continually follow others, you will never lead them. Sometimes, as in sailing, you need to tack to clear air.

5.2. What to study

There have been numerous calls for more relevant research and to combine rigor and relevance. My view is that for those of us in business schools, we (by accepting the higher salaries and lower teaching loads) have the obligation to devote at least 50% of our research to issues of business and public policy.

One major influence on my research (and a good source of ideas for research) was the Marketing Science Institute. When I was Executive Director there, I read roughly 200 dissertation and research proposals and working papers and interacted with executives in 70 member companies each year. This was complemented by editorial positions (at the time I was editing Marketing Letters or co-editing IJRM with Stefan Stremersch). These experiences both gave me a broad perspective on the field and an overabundance of relevant issues to explore.

For many researchers, the choice of topics to study often evolves over four stages. The first stage is reading/literature based. New PhDs are naturally inclined to turn to the literature they read (in courses and in preparation for their qualifying exams, much of which the qualifying exam maddeningly never seems to cover) as a source for ideas. In the second stage, conferences and seminar presentations provide additional stimuli for generating research ideas. This stage is frequently followed by a third stage in which conversations with business and public policy figures as well as the design of teaching material and in-class discussions generate more, and typically more relevant, ideas. The final stage/source of ideas tends to be simply intrinsic interest.

In my case, I wrote a dissertation which was basically an application of a multiattribute model. This work was followed by a number of papers (e.g., Beckwith & Lehmann, 1975) which focused on methods and explorations of the Howard and Sheth (1969) model, i.e., stage two. Stage three focused on "real" topics including new products, brands, and metrics. Most recently, I have tended to explore "interesting" observations concerning, for example, decision satisfaction (Heitmann, Lehmann, & Herrmann, 2007), decision comfort (Parker, Lehmann, & Xie, 2016), and currently, the use of found time based on a lunch conversation about what to do with the time freed up that day by a meeting cancelation (Chung, Lee, Lehmann, & Tsae, 2019). Of course, mine was not a linear journey; fairly early on I explored "interesting" topics, e.g., decision delay (Greenleaf & Lehmann, 1995) and how people spend leisure time (Holbrook & Lehmann, 1981). Similarly, recently, I have explored substantive issues, e.g., the impact of having a marketer on the board of directors on firm growth (Whitler, Krause, & Lehmann, 2018). Perhaps most important, I have relied on almost 200 co-authors for both topic inspiration and research execution. The moral here is simple: if you keep your eyes and mind open, you will find more than enough research you want to conduct.

5.3. Mistakes/failures can lead somewhere

There is a natural tendency to "bury" (antithetically to the development of knowledge) or hide non or bad results. This tendency is bad science and can lead to lost opportunities to explore why the results were "bad". As an example, my paper on halo effects (Beckwith & Lehmann, 1975) came about because in computing importance values in a multiattribute model, I reversed the scale on the attribute weights and weighted the least important attribute most heavily and vice-versa. Surprisingly, to me, the predictive power (R^2) was largely unchanged. Remembering my committee member Mike Pessemier's comment on the survey I used that "you know all they are going to tell you is what they like", we set about using TSLS to identify the relative impact of beliefs on attitude and attitude on beliefs. They turned out to be about the same with larger impact of attitude (halo) occurring on the less important/more subjective attributes. The moral here is to not write off "bad" results before considering what could have made them bad or what you can learn from them.

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

5.4. Publishing is hard

My first three papers were accepted (one on the first round). Thus, it was a bit of a shock when the next several were rejected. Since then, I have observed a cyclic pattern: some good results are followed by a number of bad ones and vice-versa. The worst series of results I had were in the days of reviews by mail. One day, when I had four papers under review at four different journals, I went to my mailbox first thing that morning and found a rejection letter waiting for me. After a brief vent, I went back to work. At lunch, I checked again and found rejection number 2. Now a bit discouraged, I nonetheless continued working until I checked my mail before I left to go home and found, you guessed it, rejection 3. By this point, I was beyond frustrated and found the situation a bit amusing. At this point, my colleague John Farley cheerfully noted that tomorrow would be a better day. As usual, he was correct: I only received one rejection that day (number 4). In effect, my entire research pipeline had been rejected in just over 24 hours. Fortunately, I remembered my mother's advice that when things aren't going well: “just shut up and keep swimming”.

5.5. Not every paper will be a hit

While one might hope that each of their published papers will be influential and well cited, it won't happen. For example, early on, I produced several papers that I thought were really good. Nonetheless, after 40 years several of them had produced fewer than 10 ISI citations (Table 7). The phase “The world will little note nor long remember” (Lincoln, 1863) applies to almost everything we, or at least I, publish. Interestingly, the first paper in Table 7 (from 1974) which only had 7 citations is a core component of the model development procedure proposed earlier in the paper, thus demonstrating being ignored does not necessarily mean something is useless.

By comparison, some other articles received significantly more attention (cites). What is the difference between the two sets of papers? While the first set are mostly fairly technical and method focused, the second focus on substantive problems (and have simple titles, i.e., no colons or “cute” phrases). While extrapolating from an n of 1 is risky, this suggests working on “relevant” problems with good co-authors helps. For example, the most cited paper (Anderson, Fornell, & Lehmann, 1994) came about because I spent a sabbatical at the University of Michigan. At the time Gene Anderson and Claes Fornell were working on satisfaction data from Sweden (the predecessor to the American Customer Satisfaction Index). Having recently heard Erin Anderson give a seminar at Wharton on salesperson satisfaction which reported a low correlation (actually an inverted U relation) between satisfaction and performance, I asked what the value of satisfaction was. This led us to explore the impact of satisfaction on ROA which, serendipitously, turned out to be an early paper linking marketing to financial performance which helps explain its citations. Importantly, the paper was done out of curiosity and a focus on relevance, not a desire to create a “hit” paper.

Table 7
Some less and more cited papers.

Some Rarely Cited Papers
Some Alternatives to Linear Factor Analysis For Variable Grouping Applied to Buyer Behavior Variables
Lehmann, JMR, 1974; 7 ISI citations
Exponents of Multiattribute Models
Lehmann and Gruber, JBR, 1980; 0
Objective and Strategy Determination
Hulbert, Lehmann, and Winer, JBR, 1983; 0
A Stochastic 3-Way Unfolding Model for Asymmetric Binary Data
DeSarbo, Lehmann, Holbrook, Havlena, and Gupta, Applied Psych Measurement, 1987; 3
Estimating Probabilistic Choice Models From Sparse Data
Steckel, Lehmann, and Corfman, Psych Bulletin, 1988; 4
PACM – A 2-Stage Procedure for Analyzing Structural Models
Lehmann and Gupta, Applied Psych Measurement, 1989; 1
A Combined Simply Scalable and Tree-Based Preference Model
Lehmann and Moore, JBR, 1991; 2
More Cited Papers
Customer Satisfaction, Market Share, and Profitability
Anderson, Fornell, and Lehmann (1994)
Brands and Branding
Keller and Lehmann (2006)
Valuing Customers
Gupta, Lehmann, and Stuart (2004)
A Meta-Analysis of Applications of Diffusion Models
Sultan, Farley, and Lehmann (1990)
The Long Term Impact of Promotion and Advertising on Consumer Brand Choice
Mela, Gupta, and Lehmann (1997)

5.6. Marketing PhD programs should not take 6 years

I went to Purdue and got an MSIA and PhD in two years and nine months. Through the 1980s, the norm was three years. Since then, the norm seems to have evolved to be five or six years plus not infrequently a postdoc. This has implications for who applies. Many people are unwilling to commit 6 years of below standard wages to enter what is now a mature industry where jobs and tenure are not easily attained. I know that I, who was married with a son, would have taken the job I was offered at GE as a computer programmer rather than enter a six-year PhD program.

There is also the issue of diminishing marginal returns. Several years ago, a student came to me at the end of his third year and said he was being advised by other faculty members to spend an extra year before going on the job market (i.e., take 5 years to get the PhD degree). I asked two questions. First, how much more did he think he was going to learn by staying an extra year rather than interacting with a new set of colleagues. The answer (which I agreed with) was not much. I then asked if he felt ready and wanted to go on the market. The response was yes. Then I asked him why he wouldn't. He decided to go on the market, got a good job, moved to another good school, got tenure, and is now an important scholar in marketing (who has had an extra year's salary).

I also sense that many students are a bit too comfortable being students. In the 1960s and 70s, people entering the field were largely drawn by its unstructured nature and the opportunity to explore it. Currently students (and faculty members) seem more content to focus on a particular area within it (often beginning as undergraduates) which makes them less in a hurry to "get on with it". Faculty, in part because of the assistance students provide, are often only too glad to accommodate them.

The goal of a PhD program is to train independent researchers. Most would agree that relatively little learning occurs after the fourth year in any particular program (i.e., there are diminishing marginal returns). Rather effort is largely devoted to building up a research portfolio and getting experience doing research i.e., preparing for the job market. In effect students become interns/post doc fellows. Why not formalize this so that a person is expected to defend a dissertation by the end of year four (but not necessarily formally file it with the university to retain student status) and is named a "doctor" at that point? Schools which granted the doctorate at this point would then be expected to offer two years of a post-doc position (with compensation) to their own graduates who would also be free to accept either such a position or a full-time job in academia or elsewhere. This approach would encourage faster progress toward the degree and provide a significant milestone on the path to becoming a professor. Interestingly, a similar process occurs in medicine where a person becomes a "doctor" in 3–4 years before serving as an intern and/or resident. Do we really need more time to create a marketing scholar than to train a medical doctor?

5.7. We take ourselves too seriously

Some of us work on topics which can have real impact (assuming anyone will listen to us rather than economists or consultants) in areas such as health, financial literacy, or bottom-of-the-pyramid consumers. For most of us, however, what we do at most will help some company increase its share 1%. Put differently, in general, we are not curing cancer, ending income equality, or contributing to world peace. That doesn't mean we shouldn't work hard at what we do – having made the decision to work on something means that, at least to us, at that time it is the most important thing in the world. We should honor that decision. On the other hand, we should recognize that it is not the most important thing in the world. Put simply, 100 years ago they didn't know we were coming, and 100 years from now they won't know we were here. It seems advisable to recognize this.

5.8. Excessive focus on extrinsic rewards is undesirable

Many doctoral students and young faculty members seem to have getting tenure at a particular institution as their primary goal and "play the game" with that in mind. (Indeed, many sessions at conferences are explicitly focused on how to do that.) For better or worse, my view has been and is that I would be happy being a carpenter, etc. I think this viewpoint freed me from the tension many young scholars unfortunately currently feel.

A related issue is the role of prizes and awards. I remember a talk Russ Belk gave reflecting on the ambivalence he felt when receiving one, something I also felt when I was named an IJRM Distinguished Scholar. Another aspect is the practice of giving cash rewards and plaques. Generally, the cash rewards would be better spent going to a charity for those who really need it. As for plaques etc., mine are largely piled in my downstairs from where, upon my demise, my children will undoubtedly throw them out. (As someone raised in a different era, the idea of "building my brand" and displaying evidence of my "excellence" is an anathema.)

5.9. Summary

In spite of the concerns/problems I have noted, there have been many positive changes in marketing. The number of researchers in the field has increased dramatically. Scholars from entire continents (e.g., China, India, and Korea in Asia) are becoming high quality researchers as are those from many other countries (e.g., Germany, Turkey, Iran). Students are better (or at least more extensively) trained, and what we know has increased substantially. This bodes well for the development of knowledge (if not for the chances for tenure).

Fortunately, there has also been a large increase in the number of journals in marketing. In the late 1960s when I was studying for my comprehensive exams, the journals in marketing were the *Journal of Retailing* (founded in 1925), *Journal of Marketing*

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

(1936), *Journal of Advertising Research* (1961), and *Journal of Marketing Research* (1964). This sparsity of journals made it relatively easy to study for comps. Importantly, proceedings "counted" and many fine pieces of work appeared there. The review process was also less like a legal proceeding than it is today and the methodological demands lower (partly because many of the currently popular methods had not yet been invented).

Each decade since then more journals have been launched, with an explosion of 61 new journals in the 1980s and 1990s (Lehmann, 2005). Since 2000, more journals have continued to be launched (e.g., *QME*, *JACR*, *Journal Marketing Behavior*, *AMS Review*, *Customer Needs and Solutions*) although the rate has slowed, largely due to the decreased profitability associated with them. At present, there are approximately 100 journals that can be classified as marketing-focused. This broad set of outlets should allow ample opportunities to publish.

In spite of the proliferation of journals and their special issues, however, the consensus is that it is harder to publish. My experience is consistent with that with an important caveat. For reasons having to do with both tenure and ego, there is pressure to publish in the "top" 3 or 4 journals. Logically, this pressure makes little sense given that we now live in a world where all publications are easily and equally accessible, e.g., by key word searches. We should evaluate research by reading it. Nonetheless, busy schedules and the use of increasingly complex methods, as well as university tenure committees, make it easier to rely on the implied endorsement top journals provide. As a consequence, the benefit of journal expansion has not reached its potential.

Another positive change has been the development of new ways to collect and analyze data. From easy sources of survey responses (e.g., MTurk) to enhanced econometric methods, useful advances have been made.

Of course, many of the advances have some downsides. For example, the ease of survey and experimental data collection has led to the proliferation of multistudy papers as the norm in consumer behavior. Combined with the obsession over p-values, this has led to a tendency to run multiple studies with reduced thought (versus in the "old days" when one had to recruit a church or other group to get respondents, a non-trivial cost) to get it (the manipulation effective or the result) "right". The result has been less careful thinking and more brute force research.

Similarly, advanced methods are less understood by many readers (and some users) and have to be in effect taken on faith as "right". In addition, there is a tendency to apply (and for reviewers to require) "state of the art" methods when simpler ones would suffice, resulting in research that can be understood by (and has impact on) fewer readers.

Nonetheless, many opportunities exist for new research and others appear almost daily. I expect the next 50 years will be as productive and exciting as the last 50 were.

6. Conclusion

I have no illusions that this article will cause a major change in the field (which may be a good thing), in part because in prior work and talks, I have made many of the same points to limited effect (e.g., Lehmann, 1996; Lehmann, McAlister, & Staelin, 2011). My hope is that at least somewhere someone will stop (or at least slow down) and think a bit more about what we (often automatically and/or with certainty) do. Viewing research as a "fun" pursuit aimed at discovery (versus work aimed at publication) will help. It will also help if we stop believing that everything we read or learned is "right". I published something in a "prestigious/A journal" that made sense and passed "rigorous" review. Unfortunately, I later learned that it was wrong (no, not off a bit – algebraically wrong). I suspect that there is a chance that I am not the only one. Thus, the reader is cautioned to think about what is written here and elsewhere rather than blindly applying it.

References

Alba, J. W. (2011). In defense of bumbling. *Journal of Consumer Research*, *38*(April), 981–987.

Allenby, G. M., Arora, N., & Ginter, J. L. (1995). Incorporating prior knowledge into the analysis of conjoint studies. *Journal of Marketing Research*, *32*(2), 152–162.

Anderson, E. W., Fornell, C., & Lehmann, D. R. (1994). Customer satisfaction, market share, and profitability: Findings from Sweden. *Journal of Marketing*, *58*(3), 53–66.

Bass, Frank M. (1995) "Empirical Generalizations and Marketing Science: A Personal View." *Marketing Science* 14.3_Supplement, G6-G19.

Beckwith, N. E., & Lehmann, D. R. (1975). The importance of halo effects in multi-attribute attitude models. *Journal of Marketing Research*, *12*(August), 265–275.

Blattberg, R. C., & Hock, S. J. (1990). Database models and managerial intuition. *Management Science*, *36*(August), 887–899.

Blobaum, P., Janzing, D., Washio, T., Shimizu, S., & Schoelkopf, B. (2018). Cause-effect inference by comparing regression errors. *International Conference on Artificial Intelligence and Statistics*, 900–909.

Bucklin, R., Lehmann, D., & Little, J. (1998). From decision support to decision automation: A 2020 vision. *Marketing Letters*, *9*(August), 235–246.

Chung, J., Lee, L., Lehmann, D. R., & Tsae, C. (2019). Found time. *working paper*.

DeSarbo, Wayne S., et al. (1987) "A Stochastic Three-Way Unfolding Model for Asymmetric Binary Data." *Applied Psychological Measurement*, 11.4, 397–418.

Dorfman, R., & Steiner, P. O. (1954). Optimal advertising and optimal quality. *The American Economic Review*, 826–836.

Farley, John U., Donald R. Lehmann, and Lane H. Mann (1998) "Designing the Next Study for Maximum Impact." *Journal of Marketing Research*, *35*.4, 496–501.

Ferber, R. (1977). Research by convenience. *Journal of Consumer Research*, *4*, 57–58.

Fornell, C., & Rust, R. T. (1989). Incorporating prior theory in covariance structure analysis: A Bayesian approach. *Psychometrika*, *54*(2), 249–259.

Goldenberg, Jacob, Donald R. Lehmann, and David Mazursky (2001) "The Idea Itself and the Circumstances of Its Emergence as Predictors of New Product Success," *Management Science*, 47:1, January, 69–84.

Greenleaf, E., & Lehmann, D. R. (1995). Reasons for substantial delay in consumer decision making. *Journal of Consumer Research*, *22*(September), 186–199.

Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research*, *41*(1), 7–18.

Heitmann, M., Lehmann, D. R., & Herrmann, A. (2007). Choice goal attainment and decision and consumption satisfaction. *Journal of Marketing Research*, *44*(2), 234–250.

Holbrook, M. B., & Lehmann, D. R. (1981). Allocating discretionary time: Complementarity among activities. *Journal of Consumer Research*, *7*(4), 395–406.

Howard, J. A., & Sheth, J. N. (1969). The theory of buyer behavior. *New York*, *1969*, 63.

Hoyer, P. O., Janzing, D., Mooij, J. M., Peters, J., & Schölkopf, B. (2009). Nonlinear causal discovery with additive noise models. *Advances in neural information processing systems. Vol. 21.* (pp. 689–696).

Hulbert, J. M., Lehmann, D. R., & Winer, R. S. (1983). Objective and strategy determination: Some empirical results. *Journal of Business Research*, *11*(4), 427–438.

Please cite this article as: D.R. Lehmann, The evolving world of research in marketing and the blending of theory and data..., International Journal of Research in Marketing, https://doi.org/10.1016/j.ijresmar.2019.12.001

Janzing, D. (2008). *On causally asymmetric versions of Occam's razor and their relation to thermodynamics.* (arXiv preprint arXiv:0708.3411).
Keller, K. L., & Lehmann, D. R. (2006). Brands and branding: Research findings and future priorities. *Marketing Science*, *25*(6), 740–759.
Lehmann, D. R. (1972). Judged similarity and brand-switching data as similarity measures. *Journal of Marketing Research*, *9*(August), 331–334.
Lehmann, D. R. (1974). Some alternatives to linear factor analysis for variable grouping applied to buyer behavior variables. *Journal of Marketing Research*, *11*(May), 206–213.
Lehmann, D. R. (1980). Developing a posterior model from a weak prior with panel data. In R. P. Leone (Ed.), *Proceedings: market measurement and analysis* (pp. 29–38). Providence, RI: The Institute of Management Science.
Lehmann, D. R. (1983). An approach to blending weak prior models ad data. In W. R. Dillon, K. B. Monroe, & W. R. Dillon (Eds.), *Research methods and causal modeling in marketing* (pp. 105–110). Chicago: American Marketing Association.
Lehmann, D. R. (1996). Knowledge generalization and the convention of consumer research: A study in inconsistency. *(presidential address), in Kim P. Corfman and John G. Lynch, eds., Advances in Consumer Research, 23, Association for Consumer Research, 1–5.*
Lehmann, Donald R. (2005) "Journal Evolution and the Development of Marketing," *Journal of Public Policy and Marketing*, 24:1, Spring, 137–42.
Lehmann, D. R., & Gruber, R. (1980). Exponents of multiattribute models. *Journal of Business Research*, *8*(3), 361–370.
Lehmann, D. R., & Gupta, S. (1989). PACM: A two-stage procedure for analyzing structural models. *Applied Psychological Measurement*, *13*(3), 301–321.
Lehmann, Donald R., Leigh McAlister, and Richard Staelin (2011) "Sophistication in Research in Marketing," *Journal of Marketing*, 75:4, July, 155–65.
Lehmann, D. R., & Moore, W. L. (1991). A combined simply scalable and tree-based preference model. *Journal of Business Research*, *22*(4), 311–326.
Lenk, P., & Orme, B. (2009). The value of informative priors in Bayesian inference with sparse data. *Journal of Marketing Research*, *46*(6), 832–845.
Lincoln, Abraham (1863) "The Gettysburg address." The collected works of Abraham Lincoln. Ed. Roy P. Basler. New Brunswick, NJ: Rutgers UP. 1955.
Lynch, J. G., Jr., 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*(December), 333–342.
Mela, C. F., Gupta, S., & Lehmann, D. R. (1997). The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research*, *34*(2), 248–261.
Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *The Journal of Machine Learning Research*, *17*(1), 1103–1204.
Netzer, O., & Srinivasan, V. (2011). Adaptive self-explication of multiattribute preferences. *Journal of Marketing Research*, *48*(1), 140–156.
Parker, Jeffrey R., Donald R. Lehmann, and Yi Xie (2016) "Decision Comfort," *Journal of Consumer Research*, 43:1, June, 113–33.
Rust, R. T., & Schmittlein, D. C. (1985). A Bayesian cross-validated likelihood method for comparing alternative specifications of quantitative models. *Marketing Science*, *4*(1), 20–40.
Rust, R. T., Simester, D., Brodie, R. J., & Nilikant, V. (1995). Model selection criteria: An investigation of relative accuracy, posterior probabilities, and combinations of criteria. *Management Science*, *41*(2), 322–333.
Shalizi, C. (2017). *Advanced data analysis from an elementary point of view.* Cambridge University Press.
Spirtes, P., Glymour, C. N., Scheines, R., Heckerman, D., Meek, C., Cooper, G., & Richardson, T. (2000). *Causation, prediction, and search.* MIT press.
Steckel, J. H., Lehmann, D. R., & Corfman, K. P. (1988). Estimating probabilistic choice models from sparse data: A method and an application to groups. *Psychological Bulletin*, *103*(January), 131–139.
Stem, D. E., Jr., & Lamb, C. W. (1980). An evaluation of students as surrogates in marketing studies. *Advances in Consumer Research*, *7*(1), 796–799.
Sultan, F., Farley, J. U., & Lehmann, D. R. (1990). A Meta-analysis of applications of diffusion models. *Journal of Marketing Research*, *27*(1), 70–77.
Walker, D. (2019). *In defense of nuance.* September: Ford Foundation Newsletter.
Whitler, K. A., Krause, R., & Lehmann, D. R. (2018). When and how board members with marketing experience facilitate firm growth. *Journal of Marketing*, *82*(September), 86–105.