Field Experiments in Marketing

Duncan Simester

MIT Sloan School of Management

January 2015

Field Experiments in Marketing

Marketing is a diverse field that draws from a rich array of disciplines and a broad assortment of empirical and theoretical methods. One of those disciplines is economics and one of the methods used to investigate economic questions is field experiments. The history of field experiments in the marketing literature is surprisingly long. Early examples include Curhan (1974) and Eskin and Baron (1977), who vary prices, newspaper advertising, and display variables in grocery stores. This chapter reviews the recent history of field experiments in marketing by identifying papers published in the last 20 years (between 1995 and 2014). We report how the number of papers published has increased during this period, and evaluate different explanations for this increase. We then group the papers into five topics and review the papers by topic. The chapter concludes by reflecting on the design of field experiments used in marketing, and proposing topics for future research

1. Papers that Report Field Experiments

We focus on the five leading marketing journals that publish papers with an economics focus. They include: the *Journal of Marketing* (JM), the *Journal of Marketing Research* (JMR), *Marketing Science* (MktSci), *Quantitative Marketing and Economics* (QME), and the marketing department of *Management Science* (MngSci).[1] To identify relevant papers we first had a research assistant read every issue of the journals published between 1995 and 2014. We then supplemented this initial list with a Web of Science topic search.

This process yielded over 300 papers. We then read and screened these papers to generate a final sample of 61 papers.[2] In this screening we restricted attention to studies in which the response measure represented a behavioral response in the "field". We excluded studies in which the response was a survey or perceptual measure, such as customer evaluations or purchase intentions. We also excluded studies where the responses were obtained in a laboratory setting, including studies where the response environment was created by the researchers.

The screening also restricted attention to studies that reported experiments, rather than just analysis of historical data. An experiment includes at least two experimental conditions, with exogenous variation introduced by the researchers. In most cases this exogenous variation results from variation across randomly selected customer samples. In other cases it involves rotation of treatments over time, or matched pairs of products in an auction setting.

The 61 papers actually report findings from a total of 86 field experiments, with 18 (30%) of the papers reporting multiple field experiments. This includes twelve papers that report two experiments, five papers that report three experiments and one paper that reports findings from four experiments.

---

[1] The two major marketing journals not included in this list are *Journal of Consumer Research* (JCR) and the *Journal of Consumer Psychology* (JCP). Papers published in these two journals have a strongly psychological perspective.
[2] Although we believe that the list of 61 papers is extensive, it is unlikely to be complete. We recognize that we will have overlooked some papers that report field experiments, and apologize to these authors.

Date of Publication

Our first investigation was to compare the dates the papers were published. The publication dates are summarized in Figure 1. Papers reporting field experiments were rare between 1995 and 1999, with just three examples identified. The frequency has since increased sharply; between 2010 and 2014 a total of 37 field experiment papers were identified.

Figure 1: Publications by Year
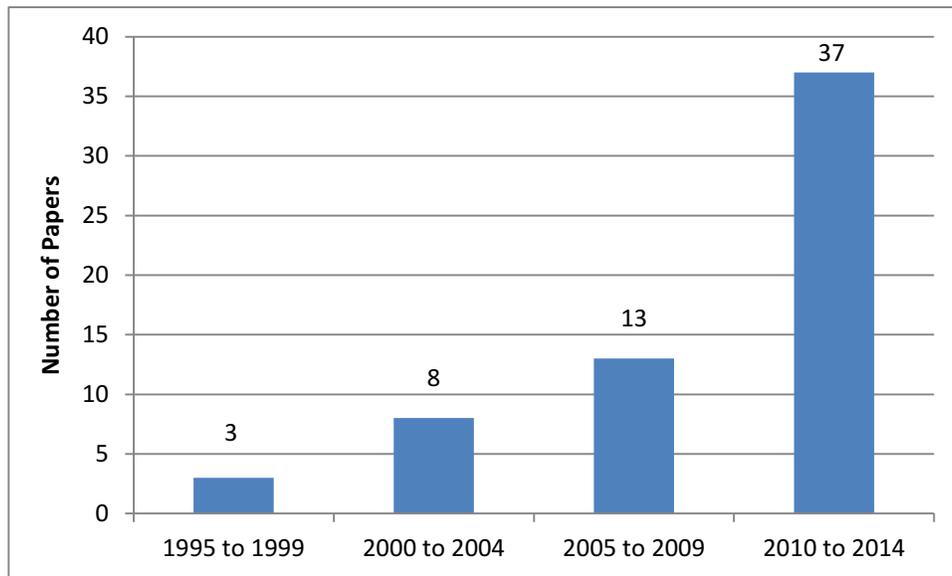
This figure reports the number of papers published by 5-year period. The sample size is 61 (papers).

There are likely to be multiple reasons for this trend. One explanation is that the field has become a lot more concerned about endogeneity when interpreting results estimated using historical data. Before 2000, the marketing literature contained numerous econometric papers studying historical supermarket scanner data.[3] Many of these early papers paid little attention to the endogeneity of the independent variables. This has since changed. The quantitative marketing field now pays considerable attention to the limitations inherent in interpreting endogenous variables. The focus on endogeneity can partly be attributed to a 1999 paper that highlighted these limitations (Villas-Boas and Winer 1999). The change also coincides with the development of structural models in the new empirical IO literature. Publication in a leading quantitative marketing journal now requires that the authors justify the source of their identification. Many papers using econometrics methods now use structural models to address this concern. Field experiments provide researchers with an alternative mechanism to overcome this hurdle. The increased prominence of concern about endogeneity may have contributed to the sharp increase in the number of published field experiments.

---

[3] This literature traces its origins to Guadagni and Little (1983), which demonstrated how to apply multinomial logit to scanner data.

A second explanation for the growth in the number of published field experiments is that it has become easier to conduct field experiments. The development of the Internet coincides with the growth in the number of reported experiments. It is now possible to conduct field experiments using eBay and Google AdWords, without requiring active participation from a cooperating company. Even for studies that do require cooperation, this cooperation is often much easier to obtain in an Internet setting because the cost of conducting the experiments is relatively low, and because many firms already conduct their own experiments. Firms that conduct experiments as part of their own operations have implicitly revealed that they understand the value of field experiments, and are likely to have invested in infrastructure to support their implementation.[4]

We can evaluate this explanation by investigating whether the format used to implement experimental variation has changed over time. In particular, we grouped the studies into three categories according to the type of experimental manipulation:

Physical   in-person interactions in a home, at a workplace, or in a bricks and mortar retail store (including shelf-signage or product assortments)

Direct Mail   telephone, catalog or other direct mail

Internet   search or display advertising, eBay, email, Twitter, website content, or another computer or Internet interaction

In Figure 2 we report how use of these manipulations has changed over time. The total sample size is 60 observations.[5]

In the five years between 2010 and 2014 there were nineteen papers published that used the Internet to implement experimental variation, representing more than half of the field experiment paper published during this period. Nineteen papers in this five year period contrasts with just three papers in the previous fifteen years. We conclude that it is likely the Internet contributed to the growth in field experiment papers across the data period.

However, it is notable that we also see growth in the number of papers implementing experiments through physical manipulations in-home, at the workplace, or in physical retail stores. There were twelve papers reporting physical manipulations between 2010 and 2014, compared to a total of just nine papers in the previous fifteen years. It seems that the ease of conducting experiments on the Internet is not a complete explanation for the growth in the number of published papers.

---

[4] We might expect that if it is easier to conduct field experiments in Internet settings then papers that use this setting are more likely to report findings from multiple experiments. It turns out that the reverse is true; papers reporting findings from Internet-based field experiments are actually less likely to report multiple studies.

[5] In four papers the description of the treatment was insufficient to allow classification. Moreover, three papers either reported multiple studies using different types of variation, or used different types of variation across different treatments within the same study. These three papers were coded as appearing in multiple categories.

Figure 2: Publications by Type of Experimental Manipulation
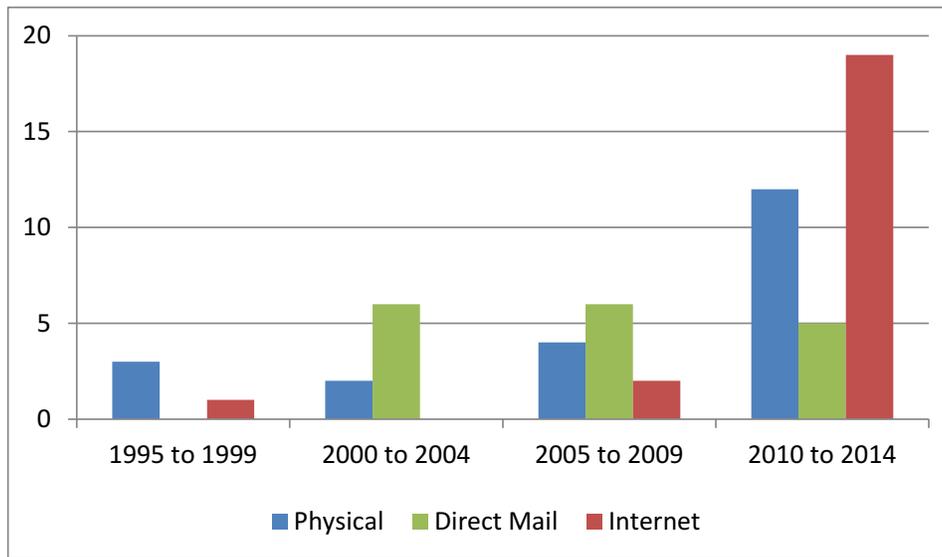
This figure reports the number of papers published using each type of experimental manipulation in each 5-year period. The sample size is 60.

A third explanation is that the top marketing journals are publishing more papers (not just more field experiment papers). In 2006 *MktSci* increased from four issues a year to six issues a year, and in 2008 *JM* and *JMR* made the same transition. In addition, *QME* began publication in 2003. The result is an increase in the total number of published papers. To investigate the extent to which this explains the increase in the frequency of field experiment papers, we counted the total number of papers published across the 20-year period (1995 to 2014) and report a histogram of the proportion published in each 5-year period. We do this separately for all papers, and for just those papers that report field experiments.[6] The findings are reported in Figure 3.

While there has been growth in the total number of papers published, this growth is a lot smaller than the growth in the number of field experiment papers. When counting all of the papers published between 1995 and 2014, we find that 31% were published in the last 5 years (2010-2014). In contrast, 64% of all field experiments papers were published in that period.

We conclude that the sharp growth in the number of marketing papers reporting field experiments can be attributed to several factors. The field itself has grown and so there has been growth in the number of papers published, including field experiment and non-field experiment papers. Second, the ease of conducting field experiments over the Internet may also have contributed to the phenomenon. Finally, the marketing field now pays more attention to the endogeneity of independent variables. Field experiments are an effective mechanism for resolving confounds due to endogeneity.

---

[6] The total number of papers in *MktSci*, *JM* and *JMR* are calculated using entries in the Web of Science. We excluded papers published in *MngSci* as the Web of Science does not identify which department editor accepted the paper. Unfortunately the Web of Science does not index the first few issues of *QME*, and so for this journal we manually counted the number of papers published in each issue.

Figure 3: Histogram of the Number of Papers Published in Each 5-Year Period
All Papers Compared to Field Experiment Papers
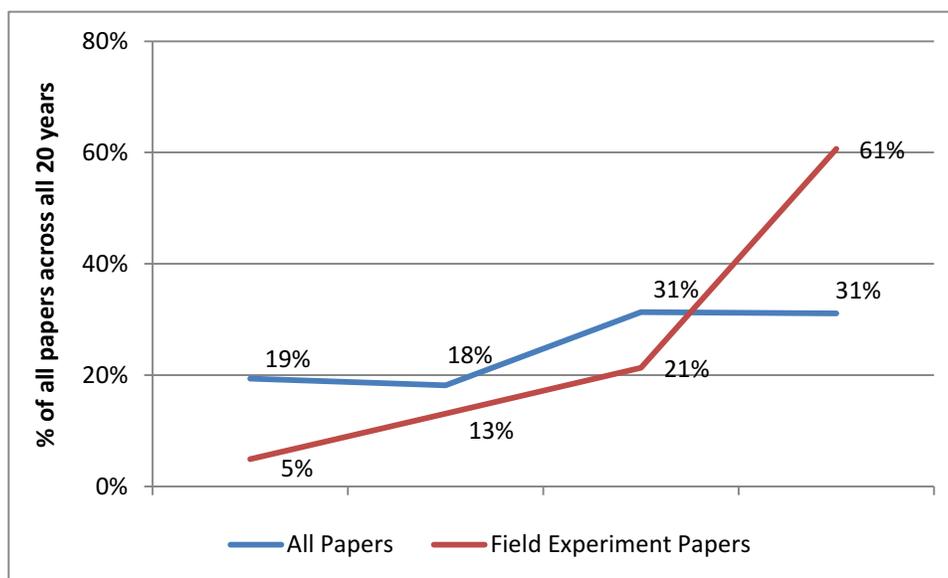
This figure reports a histogram of how many papers were published in each 5-year period. We separately report the findings for all papers, and for just those papers that report findings from a field experiment. The sample size is 3,250 for all papers, and 61 for the field experiment papers. The percentages within each curve add to 100%.

Choice of Topics

In the figure below we categorize the 61 papers into general topic areas. This categorization reveals that the papers are dominated by two topics: pricing and advertising. Further investigation also reveals clear differences in the distribution of topics over time. Prior to 2010, half of the papers (12 of 24) addressed pricing topics, with only two papers investigating advertising issues. However, this dominance has reversed, with more papers in the last five years studying advertising issues (13) than pricing issues (11).

Understandably, the recent focus on advertising has been dominated by research questions related to Internet advertising. Many of these papers study characteristics of Internet advertising that simply did not exist at the start of the data sample. For example, dynamic retargeting of advertising (targeting advertisements using individual browsing behaviors on other websites) was not possible earlier in the data period. Similarly, personalizing advertising using information posted on Facebook pages is a relatively recent advertising technique.

In the sections that follow we briefly survey the papers on each of these topics. We also summarize all of the papers, grouped by topic, in a table as an Appendix to this chapter.
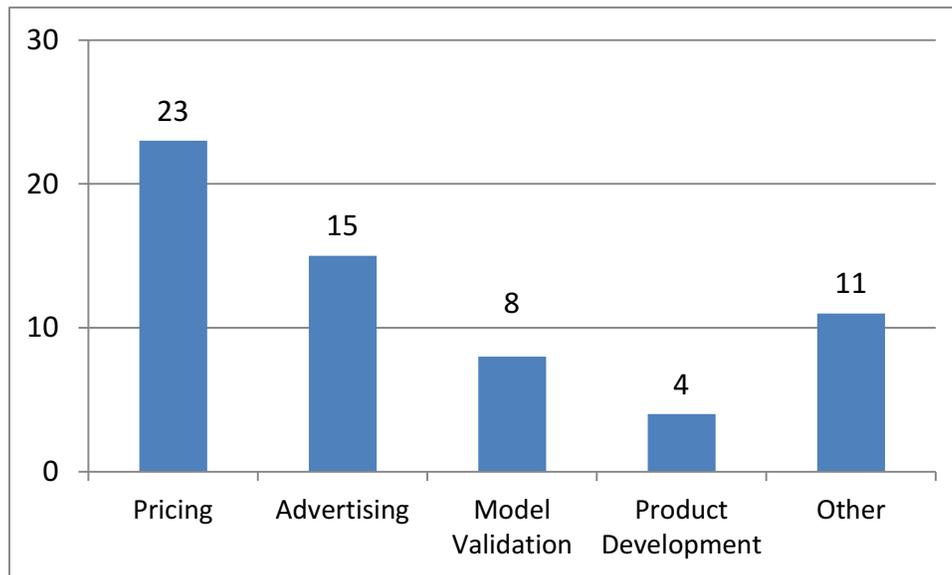
This figure reports the number of published papers by topic. The sample size is 61 papers.

2.   Pricing Topics

The 23 papers investigating pricing issues can be grouped into four sub-topics:

1. Who to target with discounts?
2. What types of discounts are most effective?
3. Price as a signal
4. Multi-part tariffs and other pricing schemes

The effectiveness of discounts has been a favorite research topic in the marketing field. This in part reflects both the pervasiveness of discounts, and recognition that they can have tremendous positive and negative impacts on unit sales and firm profits. For example, Ailawada, Harlam, César and Trounce (2007) study the unit sales and profit impact of every discount offered by the national pharmacy chain CVS in 2003. They report that more than 50% of the promotions are unprofitable because the loss of margin is not fully offset by the incremental sales. Using a very large-scale 13-week field test, in which they simply withheld promotions in 15 product categories in 400 stores, they demonstrated that CVS could improve its profits by over $50 million.

Who to Target with Discounts?

Three field experiments contribute to our understanding of who should receive discounts. Dholakia (2006) shows that sending a $5-off discount coupon to customers who previously paid full price can actually lead to less demand. This study was conducted using randomly selected customers of a large

automobile servicing firm.[7] Anderson and Simester (2001a) report that offering extended payment terms to customers of a prestigious jewelry catalog can lower demand. The authors interpret this effect as an adverse quality signal, with the installment billing signaling that the products are suitable for customers who are more sensitive to price than quality. In a later study with a different catalog retailer, the same research team report that deep discounts have more positive long-term externalities when targeted at new customers compared to existing customers (Anderson and Simester 2004). For the existing customers the long-term effects were negative; the deep discounts resulted in these customers accelerating their purchases and becoming more price sensitive. In contrast, deeper discounts increased future purchases by new customers, apparently due to more favorable expectations about future prices. Collectively these studies suggest that discounts are most effective when targeted at customers who are: newer, who have paid lower prices in the past, and who are relatively more concerned about price than quality.

Which Discounts are Most Effective?

Field experiments have been used to address not just who should receive discounts, but also what form the discounts should take. Collectively the literature now provides a rich answer to this question. In the oldest paper in our sample, Hoch and Dhar (1996) report the findings from two field experiments conducted at 86 supermarket stores in the Dominick's Finer Foods supermarket chain.[8] They showed that coupons lead to a 35% larger increase in unit sales than an equivalent discount in the shelf-price. Moreover, because redemption of coupons is incomplete, this leads to a 108% larger increase in profits. In two large-scale field experiments conducted at the (apparently) same 86 supermarkets, Wansink, Kent and Hoch (1998) illustrate two ways that retailers can increase the number of units that customers purchase. Imposing a quantity limit of 12 units per customers led to customers purchasing twice as many units (per buyer) compared to a quantity limit of 4 units. Framing discounts using multiple unit promotions (e.g. 2 for $1.50) instead of single unit promotions (75-cents each) also increased the number of units that customers purchased. Both of these effects are interpreted as examples of customers anchoring on retailer-provided cues to decide how many units to purchase. In another example of framing, Chen, Marmorstein, Tsiros and Rao (2012) show that describing a discount as "50% more free" is more effective than "33% off the regular price", even though both are economically equivalent. Their study is motivated by related evidence that customers tend to focus on the % magnitude of the discount, and neglect the base over which the percentage is calculated.

Ramanathan and Dhar (2010) present findings suggesting that the impact of different types of promotions may vary according to the customer's psychological orientation. Customers entering a Chicago grocery store were primed to think of either enjoying themselves or acting sensibly. Those primed to enjoy themselves purchased more items when they received coupons with longer expiration dates, and when the coupons were framed as "Get $x Off" rather than "Save $x". In contrast, customers

---

[7] Related findings from two field experiments are reported by Anderson and Simester (2010), which is published in an economics journal. Customers of a publishing catalog and an apparel catalog were less likely to place a subsequent order if they received a catalog containing lower prices than they had recently paid for the same item.
[8] This is the same retailer that provided the data for the now widely used Dominick's scan data panel.

primed to act sensibly purchased more items when they received coupons with shorter expiration dates and when the coupons read "Save $x".

There have been several studies investigating the impact of framing discounts as "free samples" or "free gifts". Bawa and Shoemaker (2004) use two field experiments to disentangle three effects of offering free samples: demand acceleration (temporal substitution), cannibalization, and demand expansion. They show that unlike coupons, free samples can generate positive long-term effects that can endure for as much as 12 months after the event. In a longitudinal study in the financial services sector, Haisley and Loewenstein (2011) investigated how offering customers unexpected gifts affected their subsequent deposit balances. They found that an increasing sequence of gifts ($35 then $100) had a more positive impact on subsequent balances than a decreasing sequence ($100 then $35). They described this as an "aversion to deterioration". Shu and Gneezy (2010) study how expiration dates on gift certificates affect redemption rates. Participants who received gift certificates with a longer expiration data had more favorable perceptions of the firm, but were actually less likely to redeem the certificates. They attribute this profitable coincidence to procrastination, and support this claim using a follow-up survey. Laran and Tsiros (2013) investigate how uncertainty about which free gift will be offered influence the effectiveness of these promotions. If customers were primed to think about their decisions, they responded more favorably if they knew what free gift they would receive. In contrast, if they were primed to make more emotional decisions then the uncertain gift was more effective.

We also highlight three particularly innovative studies. Hui, Inman, Huang and Suher (2013) investigate the impact of sending discounts through mobile phone technology. They show that sending coupons for products that are located further away from the customers' planned in-store shopping paths are effective at increasing unplanned purchases. Tsiros and Hardesty (2010) study how to most effectively *remove* discounts. They show that phasing out a discount gradually is more effective at raising unit sales than removing the discount all at once. They attribute this in part to higher customer expectations about future price levels. Gaurav, Cole and Tobacman (2011) study the adoption of rainfall insurance in rural Indian villages. The insurance is a relatively sophisticated financial product, and they find that training customers about the products is generally more effective than offering other forms of marketing promotions. The exception is a "MoneyBack" guarantee if the farmers never have cause to claim on the insurance, which works almost as effectively as the training programs.

Price as a Signal

A third group of pricing studies focuses on the signaling role of prices. Using data from an apparel catalog, Anderson and Simester (2001b) highlight the importance of rationing how often sale signs are used. They motivate this study as a test of an earlier paper that argues that sale signs serve a credible signaling role, enabling poorly informed customers to evaluate which prices are low relatively to other prices in the market (Anderson and Simester 1998). A key feature of that model is that the sale signs are self-regulating; any one sign becomes less credible the more often it is used. In a related study, Anderson and Simester (2003) measure the impact of 9-digit price endings (e.g. $1.99 or $49). Using a series of three field experiments conducted at two different women's apparel catalogs, they show that $9 price endings increase demand, and that this demand increase is stronger on newer items than for established items. However, the effect appears to be weakened when the $9 dollar ending is

accompanied by a "sale" sign. They interpret these findings as evidence that the price endings serve a similar signaling role to "sale" signs, revealing to customers which items are discounted compared to other prices in the market.

Gneezy, Gneezy and Lauga (2014) investigate the information that prices signal about quality. When customers see high prices they anticipate higher quality, and so react more adversely if quality is low. As a result, customers evaluate a low quality product with a high price more negatively than a low quality product with a low price. They establish this result using a field experiment at a small California vineyard.

Two other papers in this signaling stream have already been discussed as contributions to the discounting literature. Recall that Anderson and Simester's (2001a) evidence that extended payment terms lowered sales at a prestigious jewelry catalog was interpreted as evidence that targeting price sensitive customers can undermine the credibility of a retailer's quality claim. Similarly, the evidence that deeper discounts to first-time customers can make these customers more loyal also had a signaling interpretation (Anderson and Simester 2004), although in this case the signal was about future prices rather than quality.

Multi-Part Tariffs and Other Pricing Schemes

The fourth stream of pricing-related field experiments addresses multi-period and multi-part tariffs together with other innovative pricing schemes. Lambrecht and Tucker (2012) report one of the few field experiments conducted in a business-to-business setting. They collaborate with a web hosting provider in the United Kingdom to investigate the impact of varying both the monthly price and the framing of "hassle costs" in the first period of a multi-period contract. Rather than evaluating the entire contract term, they show that customers in this market evaluate the outcome in each distinct period. As a result, they show that if customers incur hassle costs in the first period, then it is more effective to lower the first period price and defer some of this revenue until later periods.

Anderson and Simester (2008) show that price differentiation can lower demand if it is perceived to be unfair. Their field experiment was conducted in a women's apparel catalog that sells plus-sizes. Because the cost of producing very large garments is higher than the cost of producing smaller garments, the retailer sought to charge higher prices on its largest sizes. The results revealed an asymmetry; there was no change in demand at the smaller sizes, but demand fell sharply for the larger sizes (even after controlling for the absolute price itself). They interpret these findings as evidence that price differentiation can lower demand if it is perceived to be unfair. However, customers actually have to experience the unfairness; it is not sufficient to see other customers treated unfairly.

Two papers focus on multi-part tariff schemes. Danaher (2002) uses the exogenous variation introduced through a field experiment in the telecommunications market to develop an optimal pricing model for a 2-part tariff. The scheme includes both a monthly access fee, together with a per-minute usage rate for cellular phone service. The results highlight the importance of accounting for both usage and customer retention. Solely focusing on usage and ignoring attrition will substantially under-estimate the sensitivity of revenue to price. Yao, Mela, Chiang and Chen (2012) also study cellular phone pricing, although in their case they use a field experiment to recover estimates of customer discount rates (rather than

revenue price sensitivity). Customers' plan choices reveal much faster consumer discounting of future periods than we normally assume.

Perhaps the most unusual pricing scheme represented in the literature is a "pay-what-you-want" (PWYW) scheme. Kim, Natter and Spann report three field experiments conducted in German restaurants and a movie theatre. The firms rotated PWYW schemes and their standard pricing schemes on different days. In general, customers chose to pay lower prices than the prices the firms regularly charged. However, this was not always the case. Customers in one of the experiments chose to pay an average of $1.94 for a hot beverage, compared to the regular price of $1.75. In another experiment, customers on average chose to pay less for a buffet lunch ($6.44) than the regular price ($7.99), but the PWYW scheme attracted enough extra demand to generate significantly higher daily revenue.

3.   Advertising Topics

The fifteen papers that focus on advertising related issues can be grouped into three broad sub-topics:

1.   Does advertising impact purchasing?
2.   Which advertising messages are most effective?
3.   Optimizing paid search advertising

Is there an Impact on Purchasing?

Four papers investigate whether advertising can impact purchasing, including two papers that reach a relatively negative conclusion. Lewis and Reiley (2014) study whether display advertisements at Yahoo! caused offline purchases at a department store. Despite a controlled randomized environment and a sample size of 1.6 million customers, they are only just able to establish a statistically significant effect. Lambrecht and Tucker (2013) study the effects of "dynamic retargeting" of advertisements. Dynamic retargeting describes the widely used practice of using browsing history from other websites to select which advertising content to display. They find that this practice does not increase advertising effectiveness. The exception is when the browsing history indicates that a customer is relatively close to making a purchase. It is only then that dynamic retargeting is effective.

Another paper with a key finding that is effectively a null result focuses on the impact of digital video recorders (DVRs) on advertising effectiveness. The introduction of DVRs led to concern that television advertising would become less effective because customers could filter out advertising. In this very large-scale scale study, the authors found no evidence that this was the case. Instead the authors concluded that filtering out advertisements occurs relatively infrequently. In contrast to the two previous papers, this could be considered a relatively positive paper about the impact of advertising. It suggests that the widespread consumer adoption of DVR technology will not diminish the impact of TV advertising.

The fourth paper in this sequence investigates the impact of competitive advertising. Competing catalog retailers share the names and addresses of their best customers on a reciprocal basis in order to lower the costs of prospecting for new customers. Using a randomized field experiment, Anderson and Simester (2013) measure how allowing competitors to target your customers with competing catalogs

impacts a firm's own sales to those customers. Although we might expect a negative effect due to substitution, they actually find that for many customers the competitors' advertising increased purchases from the original firm. This positive effect was particularly strong in product categories in which customers learn product sizes that are firm-specific (such as footwear). This leads the authors to attribute their findings to the importance of product standards, customer learning and switching costs.

Which Messages Are Most Effective?

An additional nine papers focus on improving messaging to increase advertising response rates. They include several papers that address targeting or personalization of advertising messages. For example, Tucker (2014) reports a field experiment conducted on Facebook that studies how customers react to advertisements that are personalized based on a customer's posted personal information (see also Lambrecht and Tucker 2013, which is discussed above). Fortuitously, the experiment coincided with a widely publicized change in Facebook's privacy policies, which gave users more control over their personalization settings. The findings reveal that giving customers the option of controlling their personalization settings greatly improved the performance of targeted advertisements. Schumann, von Wangenheim and Groene (2014) study how to mitigate adverse customer reactions to targeted Internet advertising. They show that reciprocity appeals ("Our service is free of charge to you – targeted advertising helps us fund it") are more effective than relevance claims ("You will see more interesting and less irrelevant advertisements in the future").

Three studies investigate how messaging should be customized to different customer segments using dimensions other than prior browsing behavior. Berger and Schwarz (2011) study messages that prompt customers to recommend products to other customers (word-of-mouth). They show that messages that link a brand to a product cue are more effective when customers do not already have strong associations between the brand and the cue. Using a Google AdWords field experiment conducted in Israel, Kronrod, Grinstein and Wathieu (2012) demonstrate that the use of assertive messages ("You must save the Mediterranean") are more effective when customers already believe in the cause, but are less effective for more general causes, for which customers' preferences are weaker.

Beyond Kronrod, Grinstein and Wathieu (2012), two other studies investigate how to design messaging that encourages customers to engage in more environmentally friendly behavior. In an ambitious study, White and Simpson (2013) collaborate with a large metropolitan city to encourage residents to leave their grass clippings on the ground, rather than dispose of them through the municipal waste system (which sends the clippings to landfills). Using hangers placed on the front doors of resident's home, they tested six different messaging conditions, together with a Control condition. They then measured the reduction in waste. Varying the type of appeal after priming a focus on either individual or social benefits had a significant effect on residents' behavior. Spangenberg, Sprott, Grohman and Smith (2003) report two studies. In the first study they investigate how messages in outdoor advertising media (billboards) can increase recycling. In the second study they vary the content of messages inserted in a health club's monthly newsletters and billing statement, with the goal of increasing member visits to the club. They show that framing an appeal as a self-prediction increases its effectiveness. Specifically, in the fitness club study, the message "Fitness guilt?" was more effective at increasing member visits than the alternative treatment "Work out at [fitness club name]".

Two other examples of messaging tests were reported by Tucker and Zhang (2010 and 2011). Both of these studies investigate how reporting participation in a two-sided network increases participation in the network. The first example (Tucker and Zhang 2010) is another rare example of a study in a business-to-business market. The setting was a website that resembles craigslist.org, and the outcome measure was the number of sellers who chose to list on the site. The field experiment randomized whether to display the number of buys and/or sellers and (if so) the number of buyers and/or sellers to claim. They find that a large number of sellers can deter additional listings, unless only the number of sellers is revealed (not the number of buyers). Displaying many buyers will attract more sellers as long as the number of sellers is also revealed. They conclude that markets with more competitors can appear more attractive to entrants, as there must be sufficient demand to attract so many competitors (a positive network externality). The second study, Tucker and Zhang (2011), uses data from a website that lists wedding service vendors. They randomize whether the website reveals the previous number of clicks on a vendor. We might expect that revealing this information would increase the popularity of the vendors with the most clicks. Instead, they show that the same level of historical clicks results in a larger lift for brands with a niche market position, compared to a brand with a mainstream position. Customers appear to infer that these brands must provide high quality in order to overcome their narrow reach.

The final paper investigating the impact of advertising messages includes a pair of studies conducted using matched pairs of music CD auctions on eBay. Dholakia and Simonson (2005) study the impact of messaging that explicitly recommends that potential bidders "compare the price of this CD with the prices of similar CDs listed next to this one". They find that these messages lead to more cautious bidding behavior. The winning bidders tended to submit later bids, submit fewer bids, and avoid participating in simultaneous auctions. They interpret the findings as evidence that making competing options more salient makes customers more risk averse because opportunity costs become more prominent.

Optimizing Paid Search Advertising

Two papers by different teams investigate how to optimize paid search advertising. Yang and Ghose (2010) begin by using an historical panel dataset to investigate how the response to paid search advertising is affected by the presence of organic (unpaid) search results. They then test the robustness of the model predictions using an 8-week field experiment on Google. They confirmed that when paid search is present, so that customers can see both paid and organic search results, the combined click-through rate (CTR) was 5.1% higher compared to when there was no paid search present. The conversion rate (probability of a purchase) also increases by 11.7%.

Agarwal, Hosanagar and Smith (2011) investigate the importance of position in paid search advertising. Specifically, they ask whether being the first search result, second result, or a lower result, impacts both click-through rates (CTR) and conversion (purchases). As we might expect, the closer to the first position, the higher the CTR. However, the opposite is true for conversion rates. Appearing lower in the rank of search results can actually increase conversion rates because conversion is conditional on clicking. As a result, only the most motivated customers click on lower results, and these are the customers that are

more likely to convert. Given advertisers generally pay for clicks irrespective of conversions, it may be optimal for many advertisers to prefer lower positions in the search order.

4. Product-Related Topics

We identified seven papers that present field experiments addressing product-related topics. This includes four studies focused on market research and product development issues, and three papers focused on product sizes.

Market Research and Product Development

Two studies investigate the role that *mere measurement* can have on customer purchasing behavior. Chandon, Morwitz and Reinartz (2005) asked 251 customers of an online grocer about their purchase intentions. A second randomly selected set of 140 consumers served as the control (and were not asked any questions). The study revealed that the customers who were surveyed were significantly more likely to make a subsequent purchase, and were significantly more profitable for the retailer. Two years later, a different research team conducted a related study in the automotive services industry (Borle, Dholakia, Singh and Westbrook 2007). A telephone survey was administered to 3,773 randomly selected customers. Subsequent service visits by these customers were then compared with a randomly selected control sample of 1,227 customers. This comparison revealed that participating in the survey increased purchases and made customers more responsive to firm promotions.

The other two studies on this topic use field experiments to validate new market research methods. Urban, Hauser, Qualls, Weinberg, Bohlmann and Chicos (1997) validate a method for conducting research very early in the product development process. Forecasts of new product sales are more valuable the earlier they are available. The authors investigate the possibility of using multimedia representations of potential new products to provide earlier and less costly forecasts. They conduct two field experiments, which confirm that multimedia computer interactions can both realistically portray customers' information sources, and yield forecasts that are not significantly different from traditional methods. Neslin, Novak, Baker and Hoffman (2009) propose a model that is designed to maximize response rates to online market research panels. The model is dynamic, optimizing over a discrete number of studies in a finite period, where this finite period can be extended using a rolling horizon. They compare their model against the current managerial heuristic, and report significantly higher response rates.

Product Sizes and Bundling

Consumers' assessments of relative package volumes are often biased when packages have different shapes. Krider, Raghubir and Krishna (2001) investigate this bias by comparing how customers react to product packages that have the same volume but different shapes. In a university cafeteria that sells cream cheese to complement bagels, the authors compare demand for a ¾ oz round tub of cream cheese and a ¾ oz square tub of the same cream cheese. Customers were significantly more likely to purchase 2 tubs of cream cheese for their bagel on days that only the round tubs of cream cheese were available. This finding is consistent with other evidence in their paper that customers perceive round containers to be smaller than rectangular containers of the same volume.

Leszczyc and Haubl (2010) report findings from three field tests that investigate the profitability of product bundling. Using a series of eBay auctions of collectable postage stamps, they compare the profitability of bundling related and unrelated items. Their findings reveal that bundling substitutes or unrelated items in an auction is less profitable than selling them separately, but bundling is more profitable when the items are complementary. A third study related to product bundling has already been discussed as an example of a pricing-related paper. Recall that in an experiment conducted at 86 supermarket stores, Wansink, Kent and Hoch (1998) compared the effectiveness of multiple unit promotions (e.g. 2 for $1.50) versus single unit promotions (75-cents each). The multiple unit promotions increased the number of units that customers purchased by an average of 32%.

5. Model Validation

The marketing field has a long tradition of developing models that optimize marketing decisions. Traditionally these models are validated by measuring goodness of fit, either in-sample or with holdout samples. A limitation of this validation is that it generally relies on assumptions inherent in the models, and so errors in the assumptions are also introduced to the validation process.

Because of these limitations, researchers have begun using field experiments as a means of validating marketing models. Field experiments provide an almost ideal validation setting; different polices can be implemented in treatment and control settings and their outcomes compared. This provides a "model-free" basis for validation, together with a comprehensive test of all of the assumptions in the model.

For example, Simester, Sun and Tsitsiklis (2006) propose a model for dynamically optimizing catalog (and other direct marketing) mailing decisions. Catalog firms regularly send catalogs to customers, and must decide who should receive each catalog. Traditionally these decisions have been made myopically; the companies send catalogs to customers who are most likely to respond to that catalog. The model proposed in this paper optimizes a sequence of mailing decisions over an infinite horizon. The model begins by proposing a method for dividing customers into discrete Markov states using a set of variables describing each customer's purchasing and mailing histories. Transition probabilities and rewards are then estimated for each strategy (mail or not mail) in each state space. Finally, a standard policy-iteration algorithm is used to calculate the optimal policy in each state. The authors then test their proposed method by varying catalog mailing decisions for 60,000 customers of an apparel catalog over a period of six months including 12 catalog mailing dates. The findings revealed that the model performed well for low-valued and moderately-valued customers. However, during the initial months of the experiment, the results were less favorable for the most valuable customers in the sample. Further investigation revealed an explanation for the poor initial outcome for these customers: in the training data there were too few occasions in which the firm had not mailed to these customers to provide a reliable estimate of the outcome.

These comparisons not only validate the proposed models, but have also proved valuable as a source of insights about the underlying phenomenon.[9]

In addition to Simester, Sun and Tsitsiklis (2006), we identified seven other papers published in the last twenty years that employ field experiments to validate optimization models. These include two pricing models, three advertising models and two new market research methods. All but one of these papers was published in 2006 or later, suggesting growing interest in this approach. Although some of the papers involve relatively small numbers of participants (see for example Urban et al. 1997; and Belloni, Lovett, Boulding and Staelin 2012), others include large numbers of participants and extended treatment periods. For example, Mantrala, Seetharaman, Kaul, Gopalakrishna and Stam (2006) compare the outcome when implementing an optimal pricing model in 200 experimental stores over a period of 8 weeks.

6. Other Topics

Our search revealed an additional 10 studies that do not fit easily within the previous four topics. Notably, these papers are all relatively recent, with the first published in 2008, and eight of them published in the last five years. We have grouped these papers into five topics:

1. Sales force optimization
2. Word-of-mouth and referrals
3. Participation in online communities
4. Encouraging positive behaviors
5. Other topics

Sales Force Optimization

Kumar, Venkatesan and Reinartz (2008) report findings from two large-scale field experiments. One experiment was conducted with a multinational business-to-business technology firm, while the second study was conducted with a firm in the telecommunications industry that sells to both businesses and retail consumers. The studies compared a "customer-focus" in which the timing of sales calls was coordinated with forecasts of customers' purchasing decisions. In this condition, sales calls were also coordinated across product categories, so that if a customer was expected to purchase in multiple categories assigned to different sales teams, the customer received a joint visit from both teams. This coordinated policy was compared with a standard policy, which lacked coordination. The two treatments were randomly assigned within matched pairs of customers. The findings revealed that coordination led to significantly higher profits and return-on-investment. Lim, Ahearne and Ham (2009) also investigate sales force optimization issues using randomized field experiments, although they focus on sales force incentives. Specifically, they compare the impact of different prize structures in sales contests. Their findings indicate that a sales contest should include multiple prize winners, and rank ordering prizes in contests with multiple winners does not increase sales or revenues.

---

[9] The findings in the Simester, Sun and Tsitsiklis (2006) field experiment led to a subsequent paper (Mannor et al. 2007) in which the authors documented the potential for positive bias when applying dynamic programming models to field data.

Word-of-Mouth and Referrals

In a widely-cited paper, Godes and Mayzlin (2009) investigate whether firm actions designed to encourage word-of-mouth (WOM) lead to higher sales. They find that WOM is more effective at increasing sales if it comes from relatively less loyal customers. They also report that WOM generates more sales if it comes from acquaintances, rather than close friends. Kumar, Peterson and Leone (2010) use four field experiments conducted with a financial services firm to identify which customers a firm should target when designing customer referral programs. They use the findings to validate an approach for computing the "customer referral value" for each customer.

Participation in an Online Community

In a study conducted in Germany involving eBay users, Algesheimer, Borle, Dholakia and Singh (2010) measure the impact of a program designed to increase participation in the firm's online community. Customers in the randomly assigned treatment group received multiple email messages inviting participation, while those in the control group did not receive these messages. Over the next year customers in the treatment group spent less and listed fewer items than customers in the control group.

Two years later an overlapping research team conducted a follow-up study again using German eBay users (Zhu, Dholakia, Chen and Algesheimer 2012). Email messages to a randomly selected treatment group were again used to invite participation in the firm's online community. The findings reveal that participants engage in more risk-seeking bidding behavior. This effect is only true when community members have strong ties to other members of the community. The findings are replicated in a second field experiment conducted with prosper.com.

Toubia and Stephen (2013) study why people contribute to the microblogging site Twitter (see also our earlier discussion of Berger and Schwarz 2011, who ask why customers contribute word-of-mouth). They experimentally manipulated the number of Twitter followers, and compared their posting activities to a randomly assigned controlled group. Their findings suggest that many users contribute because they care about how people perceive them, rather than because they derive intrinsic utility from the activity.

Encouraging Positive Behaviors

We earlier described two papers that study how advertising messages can lead to pro-social behavior (White and Simpson 2013 study waste reduction, and Spangenberg, Sprott, Grohman and Smith 2003 study recycling behavior). Two other papers have focused on encouraging positive behaviors. Raju, Rajagopal and Gilbride (2010) examine how to encourage children to choose healthier food options. The most effective intervention was to establish a "friendly healthful eating competition with students at the same grade level from other participating schools." Asking the children to sign a pledge to eat more fruits and vegetables, and providing direct incentives for healthy choices (e.g. pencils, stickers, key chains) also led to improved eating habits, even up to 10 weeks after the interventions ended.

Soman and Cheema (2011) compare methods for improving savings rates among low-income laborers in rural India. Financial planners visited the 146 families in the study for 15 weeks, and gave them a savings goal, which was placed in a sealed envelope. Experimental variations included the size of this goal, whether their pictures of their children were printed on the envelope, and whether the savings goal was

pooled into a single envelope or split between two envelopes. Although the size of the goal did not affect the savings rate, savings were significantly higher in the photograph condition, and when the savings were partitioned into multiple envelopes.

Other Topics

The remaining two papers study diverse topics. Dagger and Danaher (2014) report findings from field experiments designed to measure how remodeling a store impacts store demand. The first of these studies was conducted at an "equipment retailer and service provider" that retained its original retail space, while remodeling new space in the building to serve as a replacement retail space. The researchers randomly rotated the retail operations between the new and old spaces for a period of six weeks. They found that the remodeled space increased sales significantly more for new customers than for existing customers. They replicate their findings in a second study conducted in a large department store.

Haruvy and Leszczyc (2010) conducted a series of experiments using pairs of simultaneous auctions to measure the impact of a range of auction features, including: auction durations, shipping costs, the level of open reserve prices, and the use of secret reserve prices. The results reveal considerable price dispersion within the auction-pairs. They also confirm that the auction features had significant effects on final prices, which the authors attribute to the role of search costs.

7.   Designing Experiments and Future Topics

We conclude by reflecting on what is required to publish field experiments in the marketing literature. We also discuss the range of topics that have been studied, and identify topics that remain relatively under-studied.

Designing Experiments for the Marketing Literature

Field experiments published in marketing obviously share many of the same characteristics as field experiments published in economics journals. However, there are some distinguishing features. Publishing any empirical study in the marketing literature requires more than merely documenting an effect. Researchers are expected to also shed light on the mechanism that causes the effect. For example, it is not sufficient to just show that multiple unit promotions (e.g. 2 for $1.50) are more effective than single unit promotions (75-cents each). Wansink, Kent and Hoch (1998) were also expected to explain this result as an example of a more general phenomenon. They interpreted their finding as an example of "anchoring and adjustment".

This requirement can be a formidable obstacle for field experiments, as it is often not possible to interview customers, or to otherwise collect intermediate process measures that can reveal underlying causes. Instead, there are four approaches that researchers have generally used to investigate the cause of their effects, and they often use these approaches in combination.

First, many papers report interactions rather than just main effects. For example, Tucker and Zhang (2011) do not just report the main effect of revealing popularity information on a website. Instead, they compare these effects for niche versus mainstream brands. Similarly, Berger and Schwarz (2011)

compare the effect of advertising message on word-of-mouth for customers who have strong and weak associations between the brand and the message. For researchers seeking to publish field experiments in marketing, it is important to give careful thought to what interactions they can measure, and how those interactions will help clarify the cause of their effects. Insightful interactions are often viewed as the "clever" element of a study. Ideally the interactions are consistent with a proposed argument, and are difficult to reconcile with alternative explanations. Indeed, this is the benefit of reporting interactions; they are more likely to be immune from alternative explanations.

To estimate interactions, researchers sometimes use multiple field experiments (recall that 30% of the papers report results from multiple field experiments). Although the additional studies are occasionally positioned as replications (e.g. Danaher 2002), they are also often used to investigate interactions (e.g. Anderson and Simester 2003). Other studies report a single experiment, but include a large number of experimental treatments. For example, in Gaurav, Cole and Tobacman's (2011) study of demand for rainfall insurance among rural Indian farmers, they include 14 different experimental conditions. In general, studies either report multiple experiments or multiple treatments (and not both).[10]

Where a paper reports a single field experiment with just two conditions, it is common to complement the field experiment with one or more laboratory experiments. Examples include, Kronrod, Grinstein and Wathieu (2012), who add two laboratory experiments, and Krider, Raghubir and Krishna, who report five laboratory experiments. In papers that include multiple laboratory experiments, the field experiment often serves a less prominent role. In particular, the laboratory experiments may establish the main effect, replicate the effect, and investigate interactions, whereas the field experiment is relegated to confirming generalizability in a field setting.

A third approach used to investigate explanations is to combine the field experiment with a customer survey. For example, recall that Anderson and Simester (2001a) measured how customers react to an installment billing offer in a premium jewelry catalog. Customers did not purchase as much in the installment billing condition, which they interpreted as an adverse quality signal, revealing that the products are targeted at customers who are more sensitive to quality than price. To support this interpretation, they mailed catalogs similar to those used for the experimental treatments to other randomly selected customer samples, together with a short survey. Customers who received the installment billing version were more likely to express concern about product quality. One respondent wrote on the survey instrument: "It makes [catalog name] look tacky to have installment plans – kind of like Franklin Mint dolls" (at page 326). Shu and Gneezy (2010) provide a similar example of using a survey to validate an explanation for their experimental findings.

The fourth approach is to use field experiment data to estimate a structural model. For example, in a recent working paper Dubé, Luo and Fang (2016) report findings from two field experiments in which they sent SMS (text) messages to Chinese mobile phone users offering promotions on movie tickets. They randomly varied (a) the size of the discount, (b) whether the text indicated the service provider would donate to a local charity for every ticket purchased, and (c) the size of the donation (if any). Surprisingly, they show that customers are less responsive to larger donations when the discounts are

---

[10] The pair-wise correlation between the number of experiments and the (maximum) number of conditions is −0.14.

large. They estimate a structural model to provide evidence that this effect results from "self-signaling"; larger price discounts prompt consumers to infer that their purchases are no longer altruistic. The authors observe that without the structural model there is no obvious way to test this mechanism and disentangle it from alternative explanations. This approach of using a structural model to explain the findings in a field experiment is still relatively novel in the marketing literature. The structural modeling and field experiment literatures have grown at similar times, perhaps partly because they offer alternative solutions to resolving endogeneity concerns (see earlier discussion). This paper recognizes that the two approaches are not just alternatives; they can also complement each other. Given the rapid growth in interest in both methods, once the benefits of combining the methods are more widely recognized we should anticipate many more papers using this combination of methods.

Randomization

Randomization offers well-documented statistical advantages when comparing treatment and control conditions. However, randomization is not required to publish field experiments in the marketing literature. In 29% of the papers, experimental treatments were *not* assigned by randomization (in three papers it was unclear whether the studies used randomization). In some cases randomization could have led to contagion between the experimental treatments. For example, in Soman and Cheema's (2011) study of how to increase savings levels among rural Indian laborers, participants were assigned to conditions "according to geographic and social clusters to minimize the possibility of households from different treatment conditions meeting and discussing their participation" (at page S17). In other cases it is not obvious how randomization could be achieved. For example, in the studies involving matched pairs of auctions on eBay (Dholakia and Simonson 2005, Leszczyc and Haubl 2010, and Haruvy and Leszczyc 2011), the matched pairs serve as treatment and control samples. Randomizing the experimental treatments between products within a pair would not be meaningful, either because the products are identical, or because the experimental variation involves differences in the product bundles themselves.

In the absence of randomization, a common approach is to rotate treatments across time. For example, in Dagger and Danaher's (2014) study of a store remodeling, "the original retail environment served as the control, used in weeks 1, 4 and 5. The new environment, which represented the treatment condition, was in place for weeks 2, 3 and 6" (page 66).[11] Other examples include Yang and Ghose (2010), where the treatment was implemented by rotating sponsorship of key words in two week intervals, and Krider, Raghubir and Krishna (2001), who rotate across days the shape of cream cheese packages in a store.

Future Topics

This survey of the recent marketing literature reveals a remarkable diversity in field experiment topics. This mirrors the diversity of topics in the field generally. Firms have a wide range of levers that they can use to influence demand, and the marketing literature consequently has a wide range of topics to

---

[11] Dagger and Danaher (2014) did randomly assign the treatments to the six weeks, although it is common for rotation to occur without randomization.

choose from. Notably, the Internet has increased the range of levers available to firms, and this has further broadened the topics that are studied.

While the diversity of available topics makes the field attractive to researchers, it is also a weakness. Almost without exception the papers we have discussed raise new research questions, rather than building on previous research. Other fields benefit from greater focus in their research topics; such as the study of price rigidity in the monetary economics literature. This is not the case in marketing, which lacks broad agreement about the research questions that should be prioritized. As a result, there has not been a tradition of extending previous findings and comprehensively answering well-defined problems. Although there are some exceptions to this observation, many of the exceptions reflect a dedication to a research question by a single research team (or by teams with overlapping researchers).

Despite the diversity of topics, there are topics that are clearly under-represented. None of the papers investigate issues related to managing upstream or downstream channel partnerships. For example, none of the papers investigate the benefits of exclusivity in supply or distribution relationships, or the role of standardization to facilitate coordination. The absence of studies on these topics may in part reflect the difficulty of implementing experimental variation. Variation in channel relationships almost always requires conflict, which disrupts relationships that firms have often spent considerable time building.

There is also relatively little research studying business-to-business markets. Only five of the papers report field experiments in which firms target other businesses (this includes one study targeting rural farmers in India). This paucity of studies may reflect difficulties in generating sufficient sample in these markets. Experimentation may also be hindered by greater transparency, which could lead to business customers in the different experimental conditions becoming aware of the experimental variation (although this limitation would also seem relevant in several of the Internet studies). A third possibility is that the time required to implement change and observe outcomes could be longer in business-to-business settings than in consumer markets.

What is perhaps most surprising is the lack of papers investigating the feasibility of using field experiments to optimize marketing decisions. As we have already recognized, the marketing field has a long history of developing models to optimize marketing decisions. Indeed, eight of the papers that we reviewed report on field experiments designed to *validate* example of optimization models. However, field experiments are themselves an optimization method. By experimentally changing marketing variables and comparing the outcomes, firms could in principle use field experiments to improve profits by searching across the space of possible decisions.

None of the empirical papers we surveyed investigate the feasibility of this approach. Perhaps the closest example is Danaher (2002), which uses the data from a field experiment implemented by a telecommunications company to develop a statistical model that optimizes a two-part tariff (pricing) scheme. However, in this example the field experiment is used as a source of data, while the optimization is pursued using more traditional methods.

Although we could find no empirical papers that study whether firms could use field experiments as a practical optimization method, there is a recent theoretical paper. Li, Rusmevichientong, Simester,

Tsitsiklis and Zoumpoulis (2015) investigate how many experiments are needed to set prices in a product category, as the size of the category grows. Setting prices across a product category requires estimating a large matrix of cross-product demand elasticities (because items in the same category may be substitutes or complements). They show that if the category has a favorable structure, the number of experiments needed may grow just logarithmically with the number of products. They conclude that firms may be able to obtain meaningful estimates using a practically feasible number of experiments, even in categories with a large number of products. To our knowledge, this is the only paper that formally investigates when it is feasible to optimize marketing variables using field experiments alone. However, this paper presents theoretical results using information theory; it does not report the results of any field experiments.

8. Conclusions

Field experiments are no longer a rarity in the marketing literature. The sharp growth in the number of papers, particularly in the last five years, means that they are now a mainstream method for undertaking empirical research. The growth in the number of papers is matched by tremendous growth in the number of authors who are participating in these studies. Between 1995 and 1999 just 11 authors authored or co-authored a paper using a field experiment (in our sample). Between 2010 and 2014 there were 75 authors represented.[12]

There appear to be at least three reasons for this growth in the number of field experiment papers. First, the marketing field has become increasingly concerned about the interpretation of endogenous independent variables. This concern does not arise if experimental treatments are exogenously manipulated. Second, the field has grown generally, and so there are more papers of all types being published. Third, the Internet has made it easier to implement field experiments, often without the need for cooperation from firms.

Our survey reveals that before 2010, field experiment papers in marketing were dominated by pricing topics. Since 2010, at least some of this focus has shifted from pricing to advertising, with many of the recent papers investigating topics related to Internet advertising. However, even within the pricing and advertising topics, there is a remarkable level of diversity in the research questions. Very few of the papers build on a previous field experiment paper. We recognize this as both a strength and weakness of the field.

An important criterion for publishing a field experiment in a top marketing journal is the ability to extend the findings beyond documenting a main effect, to also provide insights about the cause of that effect. Although explaining effects is obviously valued in economics journals, this objective receives even greater emphasis in the marketing field. Field experiments are not always well-suited to evaluating competing explanations. For this reason, many of the papers report interactions, rather than just main effects. Other researchers complement their field experiments with laboratory experiments or customer surveys, in order to provide support for their interpretations.

---

[12] Across all 20 years there are 123 unique authors or co-authors of papers in the sample.

Although an explanation for an effect is generally necessary, randomization is not. Almost 30% of the field experiments published in marketing do not employ randomization to assign experimental treatments. When treatments are not randomized, they are generally rotated over time. We also identified three papers that report field experiments using matched sets of products auctioned on eBay.

Although we have highlighted the diversity in topics, our review also identified several topics that are yet to receive attention. These include the problems of motivating and coordinating upstream and downstream channel relationships. There are also very few studies of business-to-business markets. Finally, we recognized that field experiments are themselves an optimization method; firms could use field experiments to improve profits by searching across the space of possible decisions. Although there is some initial theoretical work in this area, the feasibility of using field experiments in this role remains an important but under-studied topic.

9. References

Agarwal, Ashish, Kartik Hosanagar, and Michael D. Smith (2011), "Location, Location, Location: An Analysis of Profitability of position in online Advertising Markets," *Journal of Marketing Research*, Vol. XLVIII, December, 1057-1073.

Ailawadi, Kusm L., Bari A. Harlam, Jacques César, and David Trounce (2007), "Quantifying and Improving Promotion Effectiveness at CVS," *Marketing Science*, 26(4), 566-575.

Algesheimer, René, Sharad Borle, Utpal M. Dholakia, and Siddarth S. Singh (2010), "The Impact of Customer Community Participation on Customer Behaviors: An Empirical Investigation," *Marketing Science*, 29(4), 756-769.

Anderson, Eric T. and Duncan I. Simester (1998), "The Role Of Sale Signs" *Marketing Science*, 17(2), 139-155.

Anderson, Eric T. and Duncan I. Simester (2001a), "Research Note: Price Discrimination as a Signal: Why an Offer to Spread Payments May Hurt Demand," *Marketing Science*, 20(3), 315-327.

Anderson, Eric T. and Duncan I. Simester (2001b), "Are Sale Signs Less Effective When More Products Have Them?" *Marketing Science*, 20(2), 121-142.

Anderson, Eric T. and Duncan I. Simester (2003), "Effects of $9 Price Endings on Retail Sales: Evidence from Field Experiments," *Quantitative Marketing and Economics*, 1(1), 93-110.

Anderson, Eric T. and Duncan I. Simester (2004), "Long Run Effects of Promotion Depth on New Versus Established Customers: Three Field Studies," *Marketing Science*, 23(1), 4-20.

Anderson, Eric T. and Duncan I. Simester (2008), "Research Note: Does Demand Fall When Customers Perceive That Prices Are Unfair? The Case of Premium Pricing for Large Sizes," *Marketing Science*, 27(3), 492-500.

Anderson, Eric T. and Duncan I. Simester (2010), "Price Stickiness and Customer Antagonism," *Quarterly Journal of Economics*, 125(2), 729-765.

Anderson, Eric T. and Duncan I. Simester (2013), "Advertising in a Competitive Market: The Role of Product Standards, Customer Learning and Switching Costs," *Journal of Marketing Research*, 50(4), 489-504.

Bawa, Kapil, and Robert Shoemaker (2004), "The Effects of Free Sample Promotions on Incremental Brand Sales," *Marketing Science*, 23(3), 345-363.

Belloni, Alexandre, Mitchell J. Lovett, William Boulding, and Richard Staelin (2012), "Optimal Admission and Scholarship Decisions; Choosing Customized Marketing Offers to Attract a Desirable Mix of Customers," *Marketing Science*, 31(4), 621-636.

Berger, Jonah and Eric M. Schwarz (2011), "What Drives immediate and Ongoing Word of Mouth," *Journal of Marketing Research*, Vol. XLVIII, October, 869-880.

Borle, Sharad, Utpal M. Dholakia, Siddharth S. Singh, and Robert A. Westbrook (2007), "The Impact of Survey Participation on Subsequent Customer Behavior: An Empirical investigation," *Marketing Science*, 26(5), 711-726.

Bronnenberg, Bart J., Jean-Pierre Dubé, and Carl F. Mela (2010), "Do Digital Video Recorders Influence Sales?" *Journal of Marketing Research*, Vol. XLVII, 998-1010.

Chandon, Pierre, Vicki G. Morwitz, and Werner J. Reinartz (2005), "Do Intentions Really Predict

Behavior? Self-Generated Validity Effects in Survey Research," *Journal of Marketing*, 69, April, 1-14.

Chen, Haipeng (allan), Howard Marmorstein, Michael Tsiros, and Akshay R. Rao (2012), "When More is Less: The Impact of Base Value Neglect on Consumer Preferences for Bonus Packs over Price Discounts," *Journal of Marketing*, 76, July, 64-77.

Curhan, Ronald C. (1974), "The Effects of Merchandising and Temporary Promotional Activities on the Sales of Fresh Fruits and Vegetables in Supermarkets," *Journal of Marketing Research*, Vol. XI, 286-94.

Dagger, Tracey S., and Peter J. Danaher (2014), "Comparing the Effect of Store Remodeling on New and Existing Customers," *Journal of Marketing*, 78, May, 62-80.

Danaher, Peter J. (2002), "Optimal Pricing of New Subscription Services: Analysis of a Market Experiment," *Marketing Science*, 21(2), 119-138.

Dhar, Sanjay, and Stephen J. Hoch (1996), "Price Discrimination Using In-Store Merchandising," *Journal of Marketing*, January, 17-30.

Dholakia, Utpal M. (2006), "How Customer Self-Determination Influences Relational Marketing Outcomes: Evidence from Longitudinal Field Studies," *Journal of Marketing Research*, Vol. XLIII, February, 109-120.

Dholakia, Utpal M., and Itamar Simonson (2005), "The Effect of Explicit Reference Points on Consumer Choice and Online Bidding Behavior," *Marketing Science*, 24(2), 206-217.

Dubé, Jean-Pierre, Xueming Luo, and Zheng Fang (2016), "Self-Signaling and Prosocial Behavior: a Cause Marketing Experiment," working paper, University of Chicago.

Eskin, Gerald J. and Penny H. Baron (1977), "Effects of Price and Advertising in Test-Market Experiments," *Journal of Marketing Research*, Vol. XIV, 499-508.

Gaurav, Sarthak, Shawn Cole, and Jeremy Tobacman (2011), "Marketing Complex Financial Products in Emerging Markets: Evidence from Rainfall Insurance in India," *Journal of Marketing Research*, Vol. XLVIII, Special Issue 2011, S150-S162.

Gneezy, Ayelet, Uri Gneezy, and Dominique Olié Lauga (2014), "A Reference-Dependent Model of the Price-Quality Heuristic," *Journal of Marketing Research*, Vol. LI, April, 153-164.

Godes, David and Dina Mayzlin (2009), "Firm-Created-Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, 28(4), 721-739.

Guadagni, Peter M. and John D. C. Little (2008), "A Logit Model of brand Choice Calibrated on Scanner Data," *Marketing Science*, 15(1), 1-20.

Haisley, Emily and George Loewenstein (2011), "It's Not What You Get but When You Get It: The Effect of Gift Sequence on Deposit Balances and Customer Sentiment in a Commercial Bank," *Journal of Marketing Research*, Vol. XLVIII, February, 103-115.

Haruvy, Ernan, and Peter T. Popkowski Leszczyc (2010), "Search and Choice in Online Consumer Auctions," *Marketing Science*, 29(6), 1152-1164.

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

Ju-Young Kim, Martin Natter and Martni Spann (2009), "Pay What You want: A New Participative Pricing Mechanism," *Journal of Marketing*, 73, January, 44-58.

Krider, Robert E., Priya Raghubir, Aradhna Krishna (2001), "Pizzas: π or Square? Psychological Biases in Area Comparisons," *Marketing Science*, 20(4), 405-425.

Kronrod, Ann, Amir Grinstein, and Luc Wathieu (2012), "Go Green! Should Environmental Messages Be So Assertive?" *Journal of Marketing*, 76, 95-102.

Kumar, V., J. Andrew Peterson, and Robert P. Leone (2010), "Driving Profitability by Encouraging Customer Referrals: Who, When, and How," *Journal of Marketing*, 74, September, 1-17.

Kumar, V., Rajkumar Venkatesan, and Werner Reinartz (2008), "Performance Implications of Adopting a Customer-Focused Sales Campaign," *Journal of Marketing*, 72, September, 50-68.

Lambrecht, Anja and Catherine Tucker (2012), "Paying with Money or Effort: Pricing When Customers Anticipate Hassle," *Journal of Marketing Research*, Vol. XLIX, 66-82.

Lambrecht, Anja and Catherine Tucker (2013), "When Does Retargeting Work? Information Specificity in Online Advertising," *Journal of Marketing Research*, Vol. L, October, 561-576.

Laran, Juliano, and Michael Tsiros (2013), "An Investigation of the Effectiveness of Uncertainty in Marketing Promotions Involving Free Gifts," *Journal of Marketing*, 77, 112-123.

Leszczyc, Peter T. L. Popkowski, and Gerald Haübl (2010), "To Bundle or Not to Bundle: Determinants of the Profitability of Multi-Item Auctions," *Journal of Marketing*, Vol. 74, 110-124.

Lewis, Randall A., and David H. Reiley (2014), "Online Ads and Offline Sales: Measuring the Effects of Retail Advertising Via a Controlled Experiment," *Quantitative Marketing and Economics,* 12, 235-266.

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

Li, Jimmy Q., Paat Rusmevichientong, Duncan I. Simester, John N. Tsitsiklis and Spyros I. Zoumpoulis (2015), "The Value of Field Experiments," *Management Science*, forthcoming.

Lim, Noah, Michael J. Ahearne, and Sung H. Ham (2009), "Designing Sales Contests: Does the Prize Structure Matter?," *Journal of Marketing Research*, Vol. XLVI, 356-371.

Mantrala, Murali K., P. B. Seetharaman, Rajeeve Kaul, and Srinath Gopalakrishna, and Antonie Stam (2006), "Optimal Pricing Strategies for an Automotive Aftermarket Retailer," *Journal of Marketing Research*, Vol. XLIII, November, 588-604.

Neslin, Scott A., Thomas P. Novak, Kenneth R. Baker, and Donna L. Hoffman (2009), "An Optimal Contact Model for Maximizing Online Panel Response Rates," *Management Science*, 55(5), 727-737.

Raju, Sekar, Priyali Rajagopal, and Timothy J. Gilbride (2010), "Marketing Healthful Eating to Children: The Effectiveness of Incentives, Pledges and Competitions," *Journal of Marketing*, Vol. 74, May, 93-106.

Ramanathan, Suresh and Sanjay K. Dhar (2010), "The Effect of Sales Promotions on the Size and Composition of the Shopping Basket: Regulatory Compatibility from Framing and Temporal Restrictions," *Journal of Marketing Research*, Vol. XLVII, June, 542-552.

Schumann, Jan H., Florian von Wangenheim, and Nicole Groene (2014), "Targeted Online Advertising: Using Reciprocity Appeals to Increase Acceptance Among Users of Free Web Services," *Journal of Marketing*, 78, January, 59-75.

Shu, Suzanne B. and Ayelet Gneezy (2010), "Procrastination of Enjoyable Experiences," *Journal of Marketing Research*, Vol. XLVII, October, 933-944.

Simester, Duncan I., Peng Sun, and John N. Tsitsiklis (2006), "Dynamic Catalog Mailing Policies," *Management Science*, 52(5), 683-696.

Skiera, Bernd, and Nadia Abou Nabout (2013), "PROSAD: A Bidding Decision Support System for Profit Optimizing Search Engine Advertising," *Marketing Science*, 32(2), 213-220.

Soman, Dilip, and Amar Cheema (2011), "Earmarking and Partitioning: Increasing Saving by Low-Income Households," *Journal of Marketing Research*, Vol. XLVIII, Special Issue 2011, S14-S22.

Spangenberg, Eric R., David E. Sprott, Bianca Grohmann, and Ronn J. Smith (2003), "Mass-Communicated Prediction Requests: Practical Application and a Cognitive Dissonance Explanation for Self-Prophecy," *Journal of Marketing*, Vol. 67, July, 47-62.

Toubia, Olivier, and Andrew T. Stephen (2013), "Intrinsic vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?," *Marketing Science*, 32(3), 368-392.

Tsiros, Michael, and David M. Hardesty (2010), "Ending a Price Promotion: Retracting It in One Step or Phasing It Out Gradually," *Journal of Marketing*, 74, January, 49-64.

Tucker, Catherine E. (2014), "Social Networks, Personalized Advertising, and Privacy Controls," *Journal of Marketing Research*, Vol. LI, October, 546-562.

Tucker, Catherine E., and Juanjuan Zhang (2010), "Growing Two-Sided Networks by Advertising the User Base: A Field Experiment," *Marketing Science*, 29(5), 805-814.

Tucker, Catherine E., and Juanjuan Zhang (2011), "How Does Popularity Information Affect Choices? A Field Experiment," *Management Science*, 57(5), 828-842.

Urban, Glen L., John R. Hauser, William J. Qualls, Bruce D. Weinberg. Jonathan D. Bohlmann, and Roberta A. Chicos (1997), "Information Acceleration: Validations and Lessons from the Field," *Journal of Marketing Research*, Vol. XXXIV, February, 143-153.

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

Villas-Boas, Miguel J., and Russell S. Winer (1999), "Endogeneity in Brand Choice Models," *Management Science*, 45(10), 1324-1338.

Wansink, Brian, Robert J. Kent, and Stephen J. Hoch (1998), "An Anchoring and Adjustment Model of Purchase Quantity Decisions," *Journal of Marketing Research*, Vol. XXXV, February, 71-81.

White, Katherine, and Bonnie Simpson (2013), "When Do (and Don't) Normative Appeals Influence Sustainable Consumer Behaviors?" *Journal of Marketing*, 77, March, 78-95.

Yang, Sha, and Anindya Ghose (2010), "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative or Zero Independence?" *Marketing Science*, 29(4), 602-623.

Yao, Song, Carl F. Mela, Jeongwen Chiang, and Yuxin Chen (2012), "Determining Consumers' Discount Rates with Field Studies," *Journal of Marketing Research*, Vol. XLIX, December, 822-841.

Zhu, Rui (Juliet), Utpal M. Dholakia, Xinlei (Jack) Chen, and René Algesheimer (2012), "Does Online Community Participation Foster Risky Financial Behavior?" *Journal of Marketing Research*, Vol. XLIX, June, 394-407.

Appendix: Summary of the Papers

Pricing Issues

Who to target with discounts?

Avoid sending discounts to customers who previously paid full price.	Dholakia (2006)
Avoid offering discounts to customers who are more sensitive to quality than price.	Anderson and Simester (2001a)
Deep discounts have a more favorable long-term impact for new customers than for existing customers.	Anderson and Simester (2004)

What types of discounts are most effective?

More than 50% of the promotions are unprofitable because the loss of margin is not fully offset by the incremental sales.	Ailawada, Harlam, César and Trounce (2007)
Coupons are more effective for increasing sales and profits than discounts off the regular price.	Dhar and Hoch (1996)
Multiple unit promotions (e.g. 2 for $1.50) are more effective than single unit promotions (75-cents each). Larger quantity limits increase the number of units that customers buy.	Wansink, Kent and Hoch (1998)
Framing a discount as "50% more free" is more effective than "33% off the regular price".	Chen, Marmorstein, Tsiros and Rao (2012)
The optimal design of a coupon depends upon the customer's psychological orientation.	Ramanathan and Dhar (2010)
Free samples can lead to positive long-term demand effects.	Bawa and Shoemaker (2004)
A sequence of increasing customer bonuses is more effective at driving usage and retention than a sequence of decreasing bonuses.	Haisley and Loewenstein (2011)
Longer deadlines on gift cards lead to lower redemption rates.	Shu and Gneezy (2010)
Uncertainty about a free gift can influence the effectiveness of promotion, depending upon whether customers are primed to make emotional or thoughtful decisions.	Laran and Tsiros (2013)
Returning promoted prices to regular levels gradually will generate more demand.	Tsiros and Hardesty (2010)
Promotions that extend store trips can lead to an increase in unplanned expenditure.	Hui, Inman, Huang and Suher (2013)
Training customers to appreciate the benefits of the product can be more effective than promotions.	Gaurav, Cole and Tobacman (2011)

Price as a Signal

9-digit price endings can signal a price is discounted, particularly for new products.	Anderson and Simester (2003)
"Sale" signs are a less credible signal when they are used too often.	Anderson and Simester (2001b)
Higher prices set higher expectations, and so customers evaluate a low quality product more negatively when it has a higher price.	Gneezy, Gneezy and Lauga (2014)
Promotions can lower demand by signaling that a product is suitable for customers who are more sensitive to quality than price.	Anderson and Simester (2001a)

Using discounts to attract first-time customers can make these customers more loyal by signaling that the firm offers good value.

Anderson and Simester (2004)

Multi-Part Tariffs and Other Schemes

Customers evaluate outcomes within a period, rather than across the entire contract.

Lambrecht and Tucker (2012)

A multi-part pricing system can lower demand if customers interpret it as unfair.

Anderson and Simester (2008)

When designing a revenue maximizing 2-part tariff in the telecommunications market it is important to account for both usage and customer retention, as solely focusing on usage will substantially under-estimate the sensitivity of revenue to price.

Danaher (2002)

Consumers are surprisingly short-term focused when choosing between linear price schemes and 3-part tariffs.

Yao, Mela, Chiang and Chen (2012)

How do customers respond to a pay-what-you-want pricing scheme?

Kim, Natter and Spann (2009)

Advertising Issues

Is there an impact on purchasing?

It takes a lot of data to measure the impact of Internet advertising on consumer spending.

Lewis and Riley (2014)

Retargeting is generally ineffective, unless consumers are close to purchasing.

Lambrecht and Tucker (2013)

DVR's do not appear to affect household spending.

Bronnenberg, Dube and Mela (2010)

Competitors' advertising can increase your demand.

Anderson and Simester (2013)

Which messages are most effective?

Framing an appeal as a self-prediction increases its effectiveness.

Spangenberg, Sprott, Grohman and Smith (2003)

Matching the type of appeal with a focus on either individual versus social benefits can improve advertising effectiveness.

White and Simpson (2013)

Highlighting the benefits of targeting is less effective than emphasizing reciprocity for free access.

Schumann, von Wangenheim and Groene (2014)

Control over personalization can make personalized advertisements more effective.

Tucker (2014)

Markets with more competitors can appear more attractive to entrants, as there must be sufficient demand to attract so many competitors .

Tucker and Zhang (2010)

Advertising popularity information can benefit niche brands, because customers recognize that these brands must offer high quality to overcome their narrow reach.

Tucker and Zhang (2011)

Appeals to explicitly compare an offer with competing offers produces more cautious bidding behavior in auctions.

Dholakia and Simonson (2005)

The effect of advertising on word-of-mouth is stronger for those who do not already associate the brand with the message.

Berger and Schwartz (2011)

Assertive messaging is more effective if customers believe in the cause, but less effective

Kronrod, Grinstein and Wathieu

if they are not yet convinced.   (2012)

Optimizing paid search advertising

Clicks and revenues for organic search are higher in the presence of paid search.   Yang and Ghose (2010)

Clicks decrease with search position but conversion may increases with search position.   Agarwal, Hosanagar and Smith (2011)

Market Research and Product Development Issues

Market Research

Mere measurement of purchase intentions may increase the correlation between stated purchase intentions and actual purchase probabilities.   Chandon, Morwitz and Reinartz (2005)

Participation in customer satisfaction surveys increases subsequent purchase likelihoods.   Borle, Dholakia, Singh and Westbrook (2007)

Multimedia computer interactions can realistically portray customer's information sources, and provide forecasts that are not significantly different from traditional methods.   Urban, Hauser, Qualls, Weinberg, Bohlmann and Chicos (1997)

An optimal contact model for increasing response rates to online marketing research panels.   Neslin, Novak, Baker and Hoffman (2009)

Product Sizes and Bundling

If a customer needs a fixed quantity, they will purchase more units when the product is in a round package compared to a square package of equal volume.   Krider, Raghubir and Krishna (2001)

Bundling substitutes or unrelated items in an auction is less profitable than selling them separately, but bundling is more profitable for complementary products.   Leszczyc and Haubl (2010)

Multiple unit promotions (e.g. 2 for $1.50) are more effective than single unit promotions (75-cents each).   Wansink, Kent and Hoch (1998)

Validating Optimization Models

Pricing Models

Category pricing of automobile parts.   Mantrala, Seetharaman, Kaul, Gopalakrishna and Stam (2006)

Scholarships and university admission decisions   Belloni, Lovett, Boulding and Staelin (2012)

Advertising Models

Optimizing search engine advertising bids.   Skiera and Nabout (2013)

An advertising contact model for catalog mailing.   Simester, Sun and Tsitsiklis (2006)

Estimating the incremental impact of marketing activities in different channels.   Li and Kannan (2014)

Banner advertising optimization.	Urban, Liberali, MacDonald, Bordley and Hauser (2014)

Market Research and Product Development Models

Increasing response rates to online marketing research panels.	Neslin, Novak, Baker and Hoffman (2009)
Forecasting customer demand for really new products.	Urban, Hauser, Qualls, Weinberg, Bohlmann and Chicos (1997)

Other Topics

Salesforce Optimization

Coordinating sales calls to match forecasts of customers' purchasing decisions can increase profits.	Kumar, Venkatsean and Reinartz (2008)
Sales force sales contest should include multiple prize winners, and rank ordering prices in contests with multiple winners does not increase sales or revenues.	Lim, Ahearne and Ham (2009)

Word-of-Mouth and Referrals

Word-of-mouth is more effective at increasing sales if it is comes from relatively less loyal customers, and from acquaintances instead of friends.	Godes and Mayzlin (2009)
Which customers should a firm target when designing customer referral programs?	Kumar, Peterson and Leone (2010)

Participation in an Online Community

Participation in an online community leads to less spending and fewer listings.	Algesheimer, Borle, Dholakia and Singh (2010)
Participation in an online community makes customers more willing to engage in risky bidding behavior, particularly when ties to other community members are strong.	Zhu, Dholakia, Chen and Algesheimer (2012)
Users contribute to Twitter because they care about how people perceive them, rather than because they derive intrinsic utility from the activity.	Toubia and Stephen (2013)

Encouraging Positive Behaviors

Competitions, pledges and incentives can all improve children's food choices.	Raju, Rajagopal and Gilbride (2010)
Earmarking earnings can increase savings rates.	Soman and Cheema (2011)

Other

Remodeling a retail space increases sales more for new customers than established customers.	Dagger and Danaher (2014)
Auction outcomes can be influenced by shipping costs, secret reserve prices, and other auction features.	Haruvy and Leszczyc (2011)