TOOL KIT

Knowing What to Sell, When, and to Whom

by V. Kumar, Rajkumar Venkatesan, and Werner Reinartz

Predicting customer behavior is so difficult that companies spend millions inundating—and alienating—customers. Here’s a way to crunch the data that makes it possible to offer customers what they want, when they want it.
Knowing What to Sell, When, and to Whom

by V. Kumar, Rajkumar Venkatesan, and Werner Reinartz

You are a chief marketing officer contemplating your company’s quarterly mail-shot to customers. You know that if you can get some of your customers to buy from you, then you’ll have increased the chances that they’ll come back again in the future; second-time customers are more likely to become third-time customers than first-time customers are to become second-time ones, and so on. But the mailing is an expensive proposition, and you know that in the past only about 3% of customers have actually responded to mailings by making a purchase. Shareholders and financial analysts are keeping a close eye on your company’s marketing ROI, so you need to make each contact with your customers count.

You turn to your company’s newly implemented customer relationship management (CRM) system, which tracks what each customer purchases and when. Using these data, you should be able to determine the probability that a given existing customer with a certain buying history will purchase a given product at a given time. This information should enable you not only to target the customers who are most likely to purchase something but also to tailor your offering to what is most likely to appeal to them. And it should prevent you from spending money on customers who won’t follow through (and who might actually be put off by the avalanche of unsolicited offers coming from your organization). All of this should significantly improve your ROI—the benefits from more precise targeting and the reduction in the number of mailings more than outweighing the costs of customization in this digital age.

That, at least, is the theory. Unfortunately, despite the abundance of data that many companies collect, most do a poor job predicting the behavior of their customers. In fact, our research into the purchase patterns of thousands of customers at two large firms suggests that their predictions about whether a particular customer will buy a particular product at a particular time are correct only around 60% of the time, a result that hardly justifies the costs of having a CRM system in
the first place. After all, you would accurately predict the outcome of a coin toss 50% of the time. Most companies take studies like this as evidence that it’s impossible to use the past to predict the future, and they revert to the timeworn marketing practice of inundating their customers with offers.

But as we will demonstrate, the poor predictions are not the result of any basic problem with CRM systems or any failure of the predictive power of past behavior. Rather, the problem lies in the limitations of the mathematical methods most companies use to interpret the data. We have developed a new way of predicting customer behavior, based on the work of the Nobel Prize–winning economist Daniel McFadden, that delivers vastly improved results. Indeed, the new methodology ups the odds of successfully predicting a specific purchase by a specific customer at a specific time to about 80%, a number that will have a major impact on any company’s marketing ROI. Using our methodology, managers can actually increase revenues while reducing the frequency of customer contact, evidence that overcommunication does indeed damage a company’s sales.

A Problem of Probabilities
To understand why companies do such a poor job of predicting customer behavior, we must first take a closer look at the methods they use. The most common method involves two separate steps. This first step is to estimate the probability that a customer will choose to purchase a particular product. The second is to estimate the probability that a customer will make a purchase at a particular time. Most firms stop at the first step, which limits their ability to make accurate predictions about the timing of purchases, but even those companies that follow the process through end up with bad data, as we shall see.

The probability that a customer will choose to buy a particular product is assumed to be a function of a range of variables. Some of the variables will be a customer’s demographic data, some will reflect the person’s past purchasing behavior, and still other factors will have to do with the company’s actions, such as the customer’s familiarity with the brand or the nature of the company’s contact with him or her. Marketers using the traditional method determine the relative importance of the variables by looking at a sample of customers, usually those on which the firm has the richest data. Some form of regression analysis is then applied to these data to derive an equation for the desired probability, in which each variable has a coefficient, or weighting, that determines its relative importance. The equation is then used to estimate the product choice probabilities for all the customers on which the firm has enough data, the assumption being that the coefficients of the sample will remain valid for all customers in the future. What marketers get at the end of this exercise is a series of probabilities that tells them (theoretically) which customers are most likely to buy a particular product and which products a particular customer is most likely to buy. It does not, however, tell them anything about when a customer will buy.

The next step in the traditional method is to estimate the probability that a customer will make a purchase at a given time. This probability is a function of the average interval between purchases for all customers in the original sample, adjusted for a number of customer-specific variables, such as the time intervals between the most recent purchases, and how often marketing materials have been sent to each person. Historical data on these variables for a sample of customers are, once again, plugged into a form of regression analysis that produces an equation in which the relative importance of the determining variables is fixed. By feeding fresh data on these variables into the equation, marketers can derive the probability that each customer will buy a product at a particular time. This allows the marketer to determine at which times (for example, in which months) each customer is most likely to buy any of the company’s products.

The joint probability for each customer’s future purchase behavior is calculated by simply multiplying the two probabilities—which products the individual will buy and when. What marketers get from this is a probability cube whose three axes are customers, product groups, and time periods, as illustrated in the exhibit “The Customer Probability Cube.” Marketers can use the cube in various ways. They can identify what products each customer will buy over a period and when his or
her purchases are most likely to take place. Or they can identify the customers who are most likely to buy each product and the times when the product will be most actively in demand. Using those predictions, they can determine what products to offer to which customers at which times.

All this sounds very reasonable. The relationship between a customer’s decision to purchase and the choice of product seems to be captured by the fact that product choice probabilities are factored into timing probabilities. And in many industries, sample sizes are large and customer data rich. Why then are the numbers so unreliable?

Part of the problem is that the timing of a customer’s purchase is influenced by the type of product purchased. Suppose a customer purchases product A every three months and product B every four months. Let’s further suppose that two months have elapsed since the customer’s most recent purchase, and she bought product B at that time. Clearly, this customer is right now more likely to purchase product A than product B. But the approach of multiplying the product-choice and purchase-timing probabilities from two independent regression equations completely ignores any kind of interdependence between the two probabilities. The result is poor predictions of both when a customer will make a purchase and what product the customer will buy at that time. There are statistical corrections that can deal with this common regression-analysis phenomenon, however, so it is not the main problem with the method.

There’s another source of error in the traditional method that cannot be corrected: the fact that the two probability equations are based on data from a single sample. This gives rise to sampling error, the inaccuracy of results that occurs when a population sample is used to explain the behavior of the total population. To understand how this works, consider the following simple example. Suppose you have 20 million customers, and you want to know how highly they rate your product. You probably would not ask all 20 million of them what they thought and then average all the ratings. Rather, you’d be more likely to put the question to a random sample of 1,000 customers and average their responses. You’d then use their average rating of, say, four stars as a proxy for the entire population’s average rating. The problem is, if you were to take another random sample of 1,000, you might get a three-star average rating. If you took 100 such samples, you would find that the ratings from these samples followed a normal bell-shaped distribution pattern around a mean (say, 4.1 stars) that was closest to the entire population’s true average rating. The chance that the sample you started with actually had the same mean as the population as a whole is infinitesimal. To get close to the true population mean, you would have to repeat the test 100 times or more with different samples or use a much bigger sample.

The traditional approach to estimating probabilities is vulnerable to sampling error precisely because of the implicit assumption in all regression analyses that the weightings, or coefficients, of the independent variables of the sample group are representative of
population as a whole (the population here being all your existing customers and the sample being those existing customers you’re using to determine the relationship). But since that’s highly unlikely to be the case, it follows that the relationships between customer purchase decisions and the determining variables estimated through regression analysis are bound to be inaccurate. If the sampling error is severe enough, the company using this methodology can end up choosing the wrong product to push at the wrong time to the wrong customer—and even using the wrong channels (which channels a company uses are often a big determinant of both product choice and purchase timing).

Unfortunately, most companies have no option but to rely on often relatively small samples to perform the calculations. They frequently lack enough data on all their customers to estimate meaningful relationships between the various drivers of purchasing behavior. On top of that, the populations may simply be too big for their computers to handle—imagine trying to work with data from 1 million customers choosing every month from 1,000 products. Yet it’s precisely for companies with large customer populations that an accurate probability cube would create the most value.

**Eliminating Sampling Error**

So how can companies derive probabilities free of sampling error? The answer lies in a branch of statistical mathematics called Bayesian estimation. The methodology has been around for decades but is only recently entering the marketing mainstream.

Bayesian estimation gets around the problem in the following way. Rather than estimating a single weighting for each variable (as regression analysis does), the formula at the heart of this technique first specifies the range of weightings that could have produced the observed data of the sample being analyzed. Then, through an iterative chain of calculations, it allows the analyst to determine the most probable weightings for the variables involved, those that would most likely have produced the observed data. You can think of a Bayesian estimation as reproducing the dots on a scatter diagram rather than finding the best-fit line, which is what regression analysis does. This kind of calculation has greater predictive power because it reproduces the actual behavior of a sample rather than estimating a set of weightings from one sample and then assuming that those weightings are valid for the whole population.

We have built on the pioneering work of Daniel McFadden to develop a multivariate formula called a likelihood function, which can accurately compute purchase and timing probabilities for a customer population choosing from more than two products. That’s obviously important because most companies offer more than two products and many of their customers—especially those they are likely to use in a sample—will have purchased more than two different ones. While a full discussion of the mathematics of the model is beyond the scope of this article, we provide a summary description of the formula in the exhibit “Estimating the Likelihood of Purchase.” (We refer those interested in a com-

---

**Estimating the Likelihood of Purchase**

At the heart of our method for predicting customer behavior is what we call the likelihood function. The function estimates the likelihood \( L_i \) that a customer or household \( (i) \) will purchase a given product at a given time:

\[
L_i = \prod_{r_i=1}^{R_i} \left[ \left( \prod_{j=1}^{J} \left( f_j(t,j) \right) \delta_{ijt} \right)^{C_{rij}} S_j(t)^{1-C_{rij}} \right]
\]

where:

- \( R_i \) is the number of interpurchase times for customer or household \( i \)
- \( C_{rij} = \begin{cases} 0 & \text{if the } r_i^{th} \text{ interpurchase time extends beyond the observation window} \\ 1 & \text{otherwise} \end{cases} \)
- \( \delta_{ijt} = \begin{cases} 1 & \text{if product } j \text{ is bought by customer or household } i \text{ at time } t; \\ \text{the probability that } \delta_{ijt} = 1 \text{ is } P_{ij}(t) \\ 0 & \text{otherwise; the probability that } \delta_{ijt} = 0 \text{ is } (1-P_{ij}(t)) \end{cases} \)

and \( f_j(\cdot) \) and \( S_j(\cdot) \) denote the density and survivor functions, respectively. The term involving \( S_j(t) \) accounts for right censoring of the data, because the end of the data collection period usually does not coincide with a purchase for all households. Consequently, this term does not depend on \( P_{ij}(t) \).

Standard maximum likelihood methods can be used to estimate the model parameters. We have implemented the model-estimation procedure in a Gauss program.
Knowing What to Sell, When, and to Whom • TOOL KIT

The charts in this exhibit compare the probabilities of purchase for a single customer over time and across product types. The first chart shows the numbers according to our Bayesian estimation model; the second reflects the traditional method. As the numbers show, the two models predict very different purchasing behaviors for the same customer. The results in the first table indicate that a given customer is expected to buy product 1 (say, a router) in Q1, products 1 and 2 (a router and an antivirus software program) in Q2, and product 2 (an antivirus program) in Q3. The results in the bottom table indicate that the same customer is expected to purchase a router in Q2, a router and an antivirus program in Q3, and an antivirus program in Q4.

**B2B FIRM: PROBABILITIES OF PURCHASE USING**

<table>
<thead>
<tr>
<th>OUR MODEL</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product 1</td>
<td>0.75</td>
<td>0.1</td>
<td>0.17</td>
<td>0.36</td>
</tr>
<tr>
<td>Product 2</td>
<td>0.1</td>
<td>0.15</td>
<td>0.66</td>
<td>0.35</td>
</tr>
<tr>
<td>Both Products 1 &amp; 2</td>
<td>0.1</td>
<td>0.7</td>
<td>0.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TRADITIONAL MODEL</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product 1</td>
<td>0.5</td>
<td>0.7</td>
<td>0.05</td>
<td>0.2</td>
</tr>
<tr>
<td>Product 2</td>
<td>0.4</td>
<td>0.1</td>
<td>0.25</td>
<td>0.6</td>
</tr>
<tr>
<td>Both Products 1 &amp; 2</td>
<td>0.05</td>
<td>0.15</td>
<td>0.65</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Of course, for this new approach to be an improvement over traditional methods, it needs to generate more accurate results. To see if this was the case, we first applied it to a large multinational B2B company that sells high-tech products and services to professional and Fortune 500 clients. We examined three years of data (2000 through 2002) on a sample of 20,000 customers to determine coefficients for the customer variables and then applied the resulting equations to all the customers in the database to derive a probability cube covering the four quarters that started in January 2003. We looked at a range of factors related to purchase behavior (such as number of products purchased from different product categories and number of products bought within the same category) and timing (interpurchase times, for example, and frequency of marketing contact).

We then applied the traditional methodology (using the same set of customer variables) to derive a second probability cube. The probabilities obtained by the two methods produced very different numbers, as can be seen in the exhibit “How Different Are Our Numbers?” The exhibit compares the probabilities of a single, very frequent customer choosing to buy one or both of two products over a period of four quarters.

We compared the probabilities we derived by both methods with the actual observed behavior of a number of customers of our B2B firm during 2003 and 2004, and a sample of our findings is given in the exhibit “How Accurate Were We?” Our method was much better at predicting what customers would actually do than the traditional method. When our methodology, for example, predicted that a particular customer had a high probability (defined as more than 50%) of purchasing product 1 in a given quarter, in 85% of cases, the customer did indeed purchase that product. (The hit rates for buying product 2 individually and buying both products 1 and 2 together were 74% and 80% respectively.) But when the traditional methodology indicated that a customer had a high
probability of purchasing product 1, the prediction was correct in only 55% of cases. Thus, our new methodology improved the B2B company's ability to accurately predict customer behavior by about 54%. The main flaw in the traditional method is that even though it accurately predicts which products the customer will buy, it performs poorly in predicting the purchase time. The greater accuracy of our method was also reflected in the reduction of the standard deviation of our predictions. Our predictions typically varied from the outcome by 3.4 months rather than the 4.4 months of the traditional approach.

We performed the same experiment for a large corporation selling financial investment, banking, and insurance products directly to consumers. This time, we used four years of data on 10,000 customers to estimate probabilities over the fifth year. The customer variables we input were the same as those we used for the B2B company, with some adjustments reflecting the different nature of the businesses (customers can't, for instance, return financial services). The results we got were strikingly similar to the B2B case. Our model predicted actual purchases correctly 71% to 89% of the time, which compared favorably with a hit rate of between 58% and 65% for the traditional model. On average, the improvement in performance for the proposed model is about 33% compared with the traditional model. The average deviation in purchase timing was about 3.1 months for our method compared with 4.2 months for the traditional method.

More Bang for the Marketing Buck

Our experiments highlighted the importance of interdependencies between the variables. Of particular interest was our finding that purchase acceleration was linked to marketing communication in a highly nonlinear fashion. Below a certain threshold frequency of marketing contact, customers were held back from purchasing; but above a certain threshold, customers were put off. In other words, communicating too much can harm you as much as communicating too little. Clearly, many companies may be actively damaging their customer revenues in attempts to make sure that no opportunity for a sale is missed. This finding reinforces anecdotal evidence: How do you, as a customer, feel about the space taken up in your mailbox by special offers from credit card companies?

The corollary is that a careful reduction in communication by these same companies to the right levels would lead not only to lower costs but to an increase in revenues per customer. To test the effectiveness of our methodology in helping companies find those right levels, we conducted a field study to see what impact applying strategies suggested by the model would actually have on the profits and revenues at our two companies, both of which we suspected were guilty of overcommunicating.

We split each of our samples (20,000 customers at the B2B firm and 10,000 at the financial services firm) into a test group and a control group. The communication strategy for the customers in the test groups was determined by the variable relationships and the probability predictions generated by our model. The contact strategy for customers in the control groups was determined by their company's traditional approach. Over the course of a year, we collected per-customer data on revenues, costs of sales and communication, number of contacts before a purchase is induced, profit, and return on investment for the sample customers.

The exhibit "What Was the Impact on the Bottom Line?" gives a breakdown of the differences between the two groups at each company for each measure tracked. The communications plans determined by our model
resulted in sharp improvements in profitability. At the B2B firm, the new methodology increased profits by an average of $1,600 per customer, representing an improvement in ROI of 160%. Given the sample size of over 20,000 customers, the increase in profits amounted to about $32 million for the sample group alone. Since the company’s entire customer base numbered 200,000, the potential profit improvement would total $320 million. For the financial services firm, the average profitability improvement per customer was about $400, representing an ROI improvement of 200%. Given the sample size of more than 10,000 customers, the increase in profits amounted to over $4 million. Extended to the firm’s total customer population, the profit improvement would amount to $200 million.

A great proportion of this improved profitability, of course, can be attributed to the costs saved by reducing the level of communication (31% at the B2B firm and 26% at the financial services firm). But note that the revenues generated for all product groups also went up. The $365 average per-customer difference in revenues at the B2B firm, for example, implies that sales could be as much as $73 million higher if the new methodology were rolled out to all 200,000 customers. It appears, therefore, that our model does indeed do more than just allow companies to stop spending money on unreceptive customers—it actually helps companies recover sales that their traditional marketing strategies may currently be losing.

The secret to achieving a good marketing ROI is simple: Give customers more of what they truly want and less of what they don’t. It’s always been hard to work out what customers do and don’t want, let alone when they do or don’t want it, so marketers have resorted to offering them everything all the time. Our new technique makes it perfectly feasible for companies to avoid this trap. And thanks to the

<table>
<thead>
<tr>
<th>IMPROVEMENTS IN PERFORMANCE (PER CUSTOMER)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>HIGH-TECHNOLOGY COMPANY (B2B)</td>
</tr>
<tr>
<td>Revenue ($)</td>
</tr>
<tr>
<td>Product 1</td>
</tr>
<tr>
<td>Product 2</td>
</tr>
<tr>
<td>Products 1 &amp; 2</td>
</tr>
<tr>
<td>Product 3</td>
</tr>
<tr>
<td>Products 1 &amp; 2</td>
</tr>
<tr>
<td>Products 1 &amp; 3</td>
</tr>
<tr>
<td>Products 2 &amp; 3</td>
</tr>
<tr>
<td>Products 1, 2, &amp; 3</td>
</tr>
<tr>
<td>FINANCIAL SERVICES COMPANY (B2C)</td>
</tr>
<tr>
<td>Revenue ($)</td>
</tr>
<tr>
<td>Product 1</td>
</tr>
<tr>
<td>Product 2</td>
</tr>
<tr>
<td>Product 3</td>
</tr>
<tr>
<td>Products 1 &amp; 2</td>
</tr>
<tr>
<td>Products 1 &amp; 3</td>
</tr>
<tr>
<td>Products 2 &amp; 3</td>
</tr>
<tr>
<td>Products 1, 2, &amp; 3</td>
</tr>
</tbody>
</table>
widespread availability of rich databases, computing power, methodological advancements, and quantitative empirical thinking, the list of companies that can benefit from this approach is large and growing larger. Companies that take advantage of the new technology in the right way will doubly benefit—the overall reduced level of marketing will stop them from alienating customers while making more dollars available for tailored pitches to existing customers and for outreach initiatives to new ones. When companies offer customers what they want, when they want it, sales will rise.

Reprint R0603J
To order, see the next page or call 800-988-0886 or 617-783-7500 or go to www.hbr.org
Harvard Business Review OnPoint articles enhance the full-text article with a summary of its key points and a selection of its company examples to help you quickly absorb and apply the concepts. Harvard Business Review OnPoint collections include three OnPoint articles and an overview comparing the various perspectives on a specific topic.

Further Reading

The Harvard Business Review Paperback Series

Here are the landmark ideas—both contemporary and classic—that have established Harvard Business Review as required reading for businesspeople around the globe. Each paperback includes eight of the leading articles on a particular business topic. The series includes over thirty titles, including the following best-sellers:

- **Harvard Business Review on Brand Management**  
  Product no. 1445

- **Harvard Business Review on Change**  
  Product no. 8842

- **Harvard Business Review on Leadership**  
  Product no. 8834

- **Harvard Business Review on Managing People**  
  Product no. 9075

- **Harvard Business Review on Measuring Corporate Performance**  
  Product no. 8826


To Order

For reprints, Harvard Business Review OnPoint orders, and subscriptions to Harvard Business Review:
Call 800-988-0886 or 617-783-7500.
Go to www.hbr.org

For customized and quantity orders of reprints and Harvard Business Review OnPoint products:
Call Rich Gravelin at 617-783-7626,
or e-mail him at rgravelin@hbsp.harvard.edu

This document is authorized for use only in Marketing Management - I, (BM-2012-14), Term-I by Debasis Pradhan from August 2012 to February 2013.