NEXT-PRODUCT-TO-BUY MODELS FOR CROSS-SELLING APPLICATIONS

Aaron Knott
Andrew Hayes
Scott A. Neslin

AARON KNOTT is the Director of Special Projects, Harte-Hanks Data Technologies, Billerica, Massachusetts.

ANDREW HAYES is the Research Statistician, Tuck School of Business, Dartmouth College, Hanover, New Hampshire.

SCOTT A. NESLIN is the Albert Wesley Fry Professor of Marketing, Tuck School of Business, Dartmouth College, Hanover, New Hampshire.

ABSTRACT

We present and evaluate next-product-to-buy (NPTB) models for improving the effectiveness of cross-selling. The NPTB model reduces the waste of poorly targeted cross-selling activities by predicting the product each customer would be most likely to buy next. We describe the model-building process and discuss theoretical and practical issues in developing a NPTB model. We then illustrate the effectiveness of the NPTB approach with a field test. The field test shows that the NPTB model increases profits compared to a heuristic approach, and that profits are incremental over and above sales that would have occurred through other channels. We then conduct an empirical test of methodological issues. We find that incorporating current product ownership as a predictor enhances predictive accuracy the most, followed by customer monetary value to the company, and demographics. We

JOURNAL OF INTERACTIVE MARKETING
VOLUME 16 / NUMBER 3 / SUMMER 2002
Published online in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/dir.10038

find that statistical method makes little difference in predictive accuracy, with neural nets having a slight edge. A simple random sample to create the calibration database increases predictive accuracy more than a stratified random sample, although the stratified sample may be preferred to avoid underpredicting unpopular products. We explore the potential for incorporating purchase incidence models in the NPTB approach, and find that this potentially enhances the effectiveness of the NPTB model. We close with recommendations for practitioners and for future academic research.

INTRODUCTION

A retail bank has more than a million customers, but most utilize only one or two of the bank's services. The bank wants to increase usage of its full product line. A software company knows which of its customers have bought which of its products. It wants to know which additional selections it should target to which customers. An online book company wants to know which additional titles it should target to its customers. Managers of an in-store self-serve photo kiosk wish to know which additional features it should advertise to its customers, as the customer is using the kiosk.

The above are all examples of cross-selling—encouraging a company's customers who have already bought its Product A to also buy its Product B (Deighton et al., 1994, p. 61; Nash, 1993, p. 91). The challenge of cross-selling is to know which product to target to which customer. Typically a company has several candidate products. Unfortunately, it is unable to target all of these products to each customer. This may be too expensive (due to contact costs), too time-consuming (in a telemarketing context), or ineffective (due to information overload on the customer). Or, the company simply may not want to turn off the customer by flooding him or her with too many offers.

A next-product-to-buy (NPTB) model promises to enhance the effectiveness of cross-selling by specifying which product to target to which customer. The NPTB model predicts which product (or products) a customer would be most likely to buy next, given what we know so far about the customer. These products can then be targeted to that customer.

There are several important questions regarding the implementation of NPTB models. First and perhaps foremost is do they actually work, i.e., do they enhance the profitability of cross-selling campaigns? Second are a host of methodological issues: What data do we need in order to make these predictions? What statistical technique should we use to develop the model? How should we select customers for estimating the model? Is it enough to predict just what product will be bought next—do we also need to predict when?

Accordingly, the purpose of this article is to demonstrate the value of NPTB models and provide methodological guidelines for implementing them successfully. We organize the paper into four sections. First we discuss the steps involved in building a NPTB model. Next we present a field test in which we evaluate the performance of an NPTB model. Then we investigate three methodological issues: the appropriate statistical model (logistic regression, multinomial logit, discriminant analysis, or neural nets), the most valuable data (demographics, monetary value, or purchase history), and how to draw a sample for estimating the model (simple random versus proportional to current product ownership). We also investigate the value of modeling not only *what* the next product is likely to be, but *when* it will be bought. We conclude with recommendations for model builders, opportunities for additional applications, and suggestions for future research.

THE MODEL-BUILDING PROCESS

An NPTB model is a forecasting equation that can be used to predict which product a particular customer is most likely to buy. There are four steps in developing an NPTB model: (1) compiling data, (2) selecting a statistical model, (3) estimating and evaluating the model, and (4) scoring and targeting customers.

Compiling Data

The key data requirement is customer-specific data up through time period t, plus an observation of what the customer bought in period $t + 1$. For example, a book e-tailer may know how much money each customer spent and what books he or she bought up through March 2000, and then observe the books each customer buys during April 2000. A company may observe what products the customer uses in the first 10 minutes of using a kiosk, and then observe what product the customer uses next. A retail bank may have a great deal of demographic information on its customers, plus which of its 13 financial services these customers bought up until November 1999, and then what the customers buy during December.

The data through period t provide the independent, or predictor, variables for the NPTB model, and the product(s) bought in period $t + 1$ provide the dependent variable(s).[1] The predictor variables may include product ownership as of period t, recency, frequency, or monetary value (RFM) measures for each customer through period t, demographic descriptors (age, income, etc.), and marketing efforts the company targeted at each customer through period t. Since it typically costs both time and money to assemble the database, the decision of which predictor variables to collect is nontrivial. To provide guidance for this decision, we will later compare the contributions to predictive accuracy made by various predictor variables.

Selecting a Statistical Model

Statistical models for generating NPTB predictions include: logistic regression, multinomial regression, discriminant analysis, and neural nets.[2] We assume the reader is generally familiar with these methods. Excellent references are Hosmer and Lemeshow (2000) or Pindyck and Rubinfeld (1981) (logistic regression), Ben-Akiva and Lerman (1991) (multinomial logit), Lehmann, Gupta, and Steckel (1998) or Sharma (1996) (discriminant analysis), and Zahavi and Levin (1997) or David Shepard Associates (1995) (neural nets).

There are a few practical considerations that might determine the choice of model a priori. Logistic regression is particularly easy to implement and widely available. Multinomial logit is particularly applicable when only one product will be purchased next; it is designed for the situation that the consumer is making a choice of which product to buy next, among all the products offered by the company. There are instances, however, when a consumer might purchase two products at the same time. In addition, there are instances when the consumer is choosing one product but only considering a subset of the products offered by the company. In this case, one might use a "consideration set" model (e.g., Siddarth, Bucklin, & Morrison, 1995). Discriminant analysis is also widely available and under assumptions of multivariate normality and, equal variance-covariance matrices between groups who buy versus do not buy the product, minimizes prediction errors (Dillon & Goldstein, 1984). How it performs under the common circumstances when these assumptions are violated can only be evaluated through prediction tests. In fact, Stevens (1996, p. 287) suggests that logistic regression may be a better procedure when the normality assumption is violated.

Neural nets are adept at capturing nonlinearities and interactions. They predict NPTB based on a nonlinear combination of nonlinear functions, typically logistic, so the model is highly nonlinear. The ability to capture interactions is also important, since need for new products might be created not merely by the main effects of owning various products, but the combinations of products owned.[3] This would create interactions that are uncovered implicitly through the process of estimating a neural net. To capture these interactions using the other

[1] Note that not each customer will purchase in period $t + 1$. Our practice has been to use observations only for customers who do buy in period $t + 1$. When including purchase incidence in the analysis, as we do later in this article, one includes all customers, whether or not they bought in period $t + 1$.

[2] See Kamakura, Ramaswami, and Srivastava (1991) for a different approach using latent trait modeling.

[3] We are grateful to one of the reviewers for this point.

techniques would be tedious since it would require specification of several interactions and then winnowing out the insignificant effects, etc.

Although practical issues can be important, the final arbiter of which statistical model to select is often predictive accuracy. We therefore later provide an empirical comparison of the predictive accuracy of the various statistical models.

Estimating and Evaluating the Model

The process for estimating an NPTB model is similar no matter what the technique. An observation consists of a given customer's set of predictor variables, together with which product the customer chose next. These observations provide the inputs for each statistical model, and the statistical software estimates the coefficients needed to make predictions. For example, in the logistic regression model, predicted probabilities are generated by:

$$\text{Prob(Customer } i \text{ next purchases Product } j) = \frac{1}{1 + e^{-V_{ij}}} \quad (1)$$

$$V_{ij} = \sum_{k=1}^{K} \beta_{jk} X_{ijk} \quad (2)$$

where the X_{ijk} are K predictor variables measured for customer i that predict whether the customer will next purchase Product j; β_{jk} are the coefficients estimated from the data. The task is to estimate these coefficients, which is accomplished by maximum likelihood. If we have J products we have J separate logistic regression models, one for each product.[4] The dependent variable for each model is 0-1, did the customer buy or not buy product j as their next product. The independent variables are the Xs.

As in all response modeling, NPTB models should be tested on a holdout sample. The large number of parameters to be estimated, especially with neural nets, raises the danger of over-fitting the model to the calibration data, so that the model has no predictive validity outside that sample. We typically use a holdout sample roughly equal in size to the calibration sample (e.g., roughly 10,000).

There are two ways to evaluate predictive accuracy. The first is to compare one model versus another, in a "horse race" (Pfeifer, 1996). The horse race can be in terms of a lift chart that depicts the relationship between a predicted ordering of customers in terms of their likelihood to buy and their actual likelihood. Or it can be in terms of an overall measure such as probability that the next product bought was predicted correctly. A second way to evaluate predictive accuracy is to compare to a benchmark. For example, if there were P products available and we randomly guessed which product would be purchased next, our likelihood of guessing correctly would be $1/P$. If we gave ourselves G guesses, our likelihood of guessing correctly would be G/P. These calculations assume we do not take into account the *a priori* percentages of choices that customers make. This can be taken into account with relatively straightforward probability calculations (see Morrison, 1969).

Although often overlooked, another way to evaluate models is in terms of the insights they generate. We will have more confidence in extrapolating a model to some future time period if we believe the underlying parameters make sense. One interesting insight generated by NPTB models is product affinity, the incremental probability of next buying Product j given one already owns Product k.

Table 1 displays affinity relationships measured from a logistic NPTB model, expressed as odds ratios[5] for a subset of products in a retail banking application. The table reveals that base

[4] The basic application of logistic regression to NPTB does not take into account potential correlations between the equations for each product. This can be done using seemingly-unrelated-regression (SUR) techniques (Pindyck and Rubinfeld 1981).

[5] The odds ratio is defined as the odds of next buying Product j given that the customer already owns Product k, divided by the odds if the customer does not own Product k.

TABLE 1
An Example of An Odds Ratio Table of Product Affinities

	Product j				
Product k	*Base Checking*	*No-Fee Checking*	*Base Savings*	*No-Fee Savings*	*CDs*
Base Checking	2.16*†	.66	2.29*	.73	.36*
No-Fee Checking	.68	2.66*	1.55*	1.48	.69*
Base Savings	1.67*	1.09	.83*	.96	1.36*
No-Fee Savings	1.47‡	.12*	.30*	2.54*	1.66*
CDs	.63*§	.45*	.44*	.51*	4.94*

* Statistically significant at $p < .05$.
† To be read as, owning Product A increases the odds of buying Product A next by 116%.
‡ Statistically significant at $p < .10$.
§ To be read as, owning Product A decreases the odds of buying Product B next by 37% ([1 − .63] ∗ 100%).

checking, no-fee checking, no-fee savings, and CDs are all "self-generating" products—owning one of these accounts increases the chances the customer will purchase another account next. The effect is sensibly and particularly strong for CDs—owning a CD significantly enhances the odds that one will purchase another CD next (odds ratio = 4.94). Also sensible is that owning savings accounts enhances the odds of next purchasing a CD (odds ratios = 1.36 and 1.66 for base savings and no-fee savings, respectively), whereas owning checking accounts significantly decreases those odds (odds ratios = .36 and .69 for base checking and no-fee checking, respectively). This is because savings account owners are probably in a savings mode, and purchase a CD when they have accumulated enough of a balance. Checking account owners, on the other hand, are spenders, and less likely to purchase CDs. The face validity of these relationships enhances our confidence in the predictive validity of the model.

Scoring and Targeting Customers

All four techniques—logistic regression, multinomial logit, discriminant analysis, and neural nets—generate probabilities that serve as scores for each customer for each product. A perennial issue in database marketing is how far down the list to target. The answer depends on whether the probabilities generated by the statistical models can be interpreted in absolute terms. The probabilities generated by the NPTB model may be systematically higher or lower depending on market conditions. The ordering of customers will stay the same, but the absolute levels may change. Even if the numbers can be taken literally, there is still a Type I/Type II error issue of whether it is worse not to target a potential responder or to target a customer who doesn't respond. Because of these difficulties, another approach is simply to adopt a budget and target only say the top 20% of customers, etc.

Theoretical Issues and the Model Building Process

The above discussion provides a blueprint for developing an NPTB model. However, there are important issues with respect to competition and cross-channel cannibalization that merit consideration in this process. To motivate these issues, we present a brief theoretical discussion based on the logistic regression model.[6]

[6] There are different approaches to developing the theory of the logistic regression model. We find the threshold approach as articulated by Pyndyck and Rubinfeld (1981, pp. 280–301) to be most relevant to the NPTB approach, although it is equivalent to

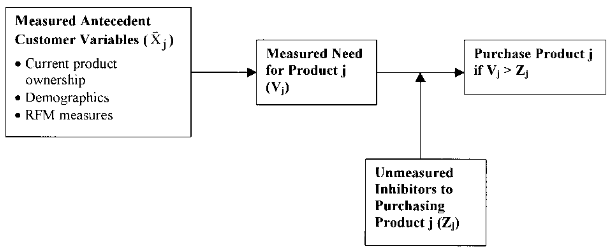

FIGURE 1
Behavioral framework underlying next-product-to-buy models

Figure 1 presents schematically the underlying theory of the logistic regression model as applied to the NPTB context. The key quantities are[7]:

$\vec{X}_j$ = A vector of variables the model-builder can measure that generate the customer's need for product *j* (see Engel, Blackwell, & Miniard, 1995, pp. 175–182). These predictor variables include current product ownership, demographics, and RFM measures (see equation 2).

V_j = The customer's need for product *j* based on the predictor variables.

Z_j = Net impact of factors the model builder cannot measure that inhibit the translation of need into purchase. Potential inhibitors include lack of problem recognition (the customer does not recognize his or her need) and marketing efforts by competitors.

The customer will buy product *j* if his or her need for the product exceeds the inhibitors. This can be depicted as a threshold model, i.e.,

$$\text{Buy Product } j \text{ if } V_j > Z_j \tag{3}$$

Logistic regression assumes that $V_j = f(\vec{X}_j)$, i.e., that the predictor variables cause need for the product. The function $f(.)$ is usually taken to be linear in the *X*s (see equation 2). Second, the logistic model assumes that Z_j is distributed across consumers according to a logistic distribution. This acknowledges that consumers vary in their levels of inhibitions. The logistic distribution is used because of analytical tractability in deriving a closed form (equation 2) for the probability a customer will next purchase product *j*, namely:

$$\text{Prob(Purchase Product } j) = \text{Prob}(Z_j < V_j) = \frac{1}{1 + e^{-V_j}} \tag{4}$$

An assumption in deriving equation 4 is that $V_j = f(\vec{X}_j)$ is independent of Z_j, i.e., measured need is independent of unmeasured inhibitors. This is akin to the standard assumption in regression that the error term (in this case *Z*) is independent of the variables included in the model (in this case the *X*s). This assumption might be erroneous if competitors are targeting the firm's customers based on the *X* variables. Competitors may have acquired a list of the firm's customers from a list broker and target based on income (one of the *X* variables). This would create a correlation between *V* and *Z*. The result would be biased estimates of the *f(X)* function and less accurate predictions. Another way this assumption could be violated would be if customers with high need were more "involved" in the category (Engel et al., 1995, pp. 161–162; 237) and hence more likely to recognize that need. In that case, high need customers may not require a targeted marketing effort since they will purchase the product anyway through the firm's existing marketing channels. In summary, violation of the independence assumption could decrease the effectiveness of targeting via NPTB models because either (1) competitive efforts will distort estimation of the model, or (2) we may end up targeting consumers who would have bought through a different channel anyway. The way to approach the competitive issue is first to ask if competitors are targeting our customers based on variables that we are using to predict NPTB. If the answer is yes, the ideal solution is to measure competitive activity. Since this is often difficult, the most practical approach is see if predictive accuracy is

the econometric approach presented in Ben-Akiva and Lerman (1991, pp. 59–72).

[7] We suppress the household subscript for simplicity.

high enough for the NPTB model to be useful. The cannibalization issue is very difficult to measure since it requires data on previous marketing efforts and formal modeling of "cross-channel elasticities." Often, the best way to investigate cross-channel cannibalization is to create experimental control groups. This is exactly what we do in the field test to be described next.

A FIELD TEST OF AN NPTB MODEL

Problem Setting

The setting was a retail bank interested in increasing sales of a particular loan product.[8] Normally a direct mailing for this loan would be targeted at customers selected by a heuristic based on wealth-related variables such as home value, income, etc. These demographics were the best that could be linked intuitively to purchasing the loan, but response rates were still typically quite low. In addition, there was the cannibalization issue, as raised above by our theoretical analysis, of whether direct mail offers generate incremental response over what the customer would eventually have done anyway albeit through a different channel (e.g., buying from the bank directly). Finally, there was also a question as to whether the bank would be better off targeting their marketing efforts to prospects purchased from commercial list brokers. In this context, the bank's officers decided to test an NPTB model.

The Field Test

The field test was designed to test (1) Would the NPTB model improve response rate and revenue per customer over the demographics-based heuristic? (2) Does direct mail solicitation, based either on the model or the heuristic, improve over what the customer would have done anyway, and (3) How do either the NPTB or heuristic methods of targeting current customers compare with targeting prospects? Accordingly, six groups were set up:

Heuristic Mail Group: Customers selected using the demographics heuristic, in this case whether home value was ≥ $100,000 or unknown, and then mailed an offer for the loan (n = 23,639).

NPTB Mail Group: Customers selected using the NPTB model, and then mailed an offer for the loan (n = 23,877). Customers were selected if the NPTB model predicted the loan would be their first- or second-most likely NPTB.

Prospect Mail Group: Prospective customers not currently customers of the bank, who were mailed an offer for the loan. Names were purchased from a list broker (n = 49,905).

Heuristic Control Group: Customers selected using the demographics heuristic, in this case whether home value was ≥ $100,000, but not mailed an offer (n = 1,186).

NPTB Control Group: Customers selected using the NPTB model, using the same procedure as for the NPTB mail group, but not mailed an offer (n = 1,209).

Prospect Control Group: Prospective customers not currently customers of the bank, and not mailed the offer (n = 2,500).

Comparing the NPTB and Heuristic mail groups allowed the bank to determine if the NPTB model outperformed a commonsense but not statistically rigorous approach. Comparing the NPTB and Heuristic mail groups to their respective control groups allowed the bank to determine whether responses were truly incremental and did not cannibalize the bank's other channels. Comparing with the Prospect mail group would allow the bank to access the relative benefits of retention versus acquisition efforts.

Model Development

A neural net NPTB model was used to predict purchase likelihood for nine products mar-

[8] There were actually two versions of the product; we aggregate the results across these two varieties for simplicity.

TABLE 2
Results of Field Test

Group	Number of Customers	Number Respond	Response Rate	Revenues	Revenues/Responder	Revenues/Customer
Heuristic mail	23,639	104	0.44%	$700,449	$6,735	$29.63
NPTB mail	23,877	270	1.13%	$2,227,146	$8,249	$93.28
Prospect mail	49,905	49	0.10%	$365,204	$7,453	$7.32
Heuristic control	1,186	5	0.42%	$26,346	$5,269	$22.21
NPTB control	1,209	6	0.50%	$44,850	$7,475	$37.10
Prospect control	2,500	0	0	NA	NA	NA

Note: Dollar numbers are disguised but preserve the true ratios for comparisons.
NPTB, next-product-to-buy model.

keted by the bank, including the loan product that was the target in this case, plus various other offerings such as IRAs, CDs, and savings accounts. The model was estimated on 7,200 households, 800 for each product,[9] and tested on 19,082 households who opened accounts in a period subsequent to which the model was estimated. Predictor variables included current ownership of 12 of the bank's products, total deposit and loan dollars at the bank, age, length in current residence, income, and home ownership. The actual product purchased corresponded to the product predicted most likely to buy 45.8% of the time, compared with 1/9 = 11% based on random prediction, and to the product predicted most- or second-most likely to buy 63.5% of the time, compared with 22% random prediction. These results are impressive, but the best test is in the field, where we tabulate actual revenues generated and ROI, and examine the level of cannibalization with the regular marketing channel—buying directly from the bank.

Results

The results are shown in Table 2[10] and suggest three important conclusions. First, the NPTB model outperforms the heuristic approach. Response rate, revenues per responder, and revenues per customer are all higher. Revenues per customer are more than three times higher for the NPTB model.

Second, the NPTB model outperforms its control group, especially in terms of response rate.[11] This shows that the NPTB model does not merely identify customers who will buy the product anyway—it identifies customers who will respond to a direct marketing effort. Note the same cannot be said for the heuristic approach. The response rate is only marginally higher using a heuristic-based mailing compared with what these customers would have done anyway.

Third, both the NPTB model and the heuristic approach outperform the Prospect group in terms of response rate and revenues per customer, although not in terms of revenues per responder.

To shed further light on the financial performance of the three approaches (NPTB, heuristic, and prospecting), we compute an ROI for each method and display the calculation in Table 3.[12] The first step was to calculate "lift" from

[9] Neural net models achieve better results when there are roughly even numbers of observations for each product.

[10] Note that the absolute dollar levels are disguised, but the response rates and relative comparisons between groups are accurate.

[11] Note "response rate" is somewhat of a misnomer in referring to the control groups. They did not respond to a direct mailing. They purchased the loan on their own and based on the bank's other marketing efforts that are common to all customers.

[12] Again the absolute dollar levels are disguised but ratios are preserved, including the ROI percentages.

TABLE 3
Field Test Profits

Method	*Lift*	*Gross Profit Contribution*	*Acquisition Cost/Mailee*	*Total Acq. Cost*	*Total Profit*	*ROI*
NPTB	$1,341,362*	$36,485†	$.2425	$5,790‡	$30,695§	530.1%‖
Heuristic	$175,342	$4,769	$.2425	$5,732	($963)	−16.8%
Prospects	$365,204	$9,934	$.2850	$14,223	($4,289)	−30.2%

Note: See Table 1.
* $2,227,146 − 23,877 ∗ .50% ∗ $7,475.
† $1,341,362 ∗ .0272 (272 basis points profit contribution).
‡ 23,877 ∗ $.2425.
§ $36,485 − $5,790.
‖ $36,485/$5,790.

the mailing. This was done by starting with the total revenue generated by each method and subtracting the revenues that would have been expected had the mailed group not been mailed. This expectation was calculated by extrapolating revenues from the appropriate control group. Then we assumed since this is a loan product that profit margin could be expressed in terms of interest basis points representing the difference between the bank's investment return and the rate it can charge for the loan. This generates gross profit contribution. We then calculate total mail costs based on an assumed cost per mailed customer, and total profit is gross profit contribution minus these mailing costs. ROI is simply the ratio of total profit to total mailing costs.

Table 3 shows that only the NPTB model was profitable and strongly so, generating a return of 530%. The other two methods generate negative ROI. A few caveats need to be stated. First, the NPTB model could arguably be allocated additional fixed costs due to the cost of running the model and scoring the database. This would lower its ROI. But these costs would probably not be significant since this was one of several applications of the NPTB model. In addition, the raw data needed from the model had already been assembled by the bank and were used for the heuristic approach as well. Second, the prospecting efforts show up as a loss, but arguably this group could be the source of long-term gains that could make the prospecting efforts profitable. This would especially bode poorly for the heuristic approach, which also lost money from an ROI standpoint. However, one could argue that both the NPTB and heuristic approaches also help to retain the customer in the bank's franchise, and thus have a long-term effect as well. A full evaluation of this could be done within a migration model framework (Nasr & Berger, 1998) and would be a very interesting avenue for future research.

Overall, the results suggest strongly that the NPTB model helped to improve the effectiveness of the bank's cross-selling efforts. The improvements are relative both to current heuristics for direct mail targeting, and to what the selected customers would have done anyway had they not been mailed the offer. In addition, the NPTB approach appeared to outperform prospecting for new customers, whereas the heuristic approach quite possibly did not.

AN EMPIRICAL INVESTIGATION OF METHODOLOGICAL ISSUES

Data Requirements, Statistical Model, and Sample Selection

Problem setting. The situation involves cross-selling opportunities for a retail bank. The bank had a product line of 13 products, and the average customer currently owned only a small number of these products. We will investigate the predictive accuracy of NPTB models based on statistical technique, data availability, and type of sample used for calibration. The four statistical models tested are logistic regression, multinomial logit, discriminant analysis, and neural networks. The data available include demographics (home value, income, length of residence, marital status, presence of children, and state of residence), current product ownership, and account value. Customers could currently own one or more of a particular product, e.g., 1, 2, 3, or more certificates of deposit. We did a subtest on whether to define the ownership variables in terms of number of accounts or 0–1 indicating whether the customer had at least one account (1) or not (0).

Sample selection. There are two ways to draw the sample. One, the simple random approach, is to select say 10,000 customers at random from the database of all customers who purchased a product during a particular time period. The other approach, stratified random, is to select a sample so that there is at least a minimum number of product purchases for each product. Neural net programs, for example, often recommend a minimum number of "switched-on" outputs for accurate model calibration (Garson, 1998, pp. 88–89).

We selected three data sets: First we drew a simple random sample of 9230 customers for whom predictor data were available as of October 1997, and who had purchased a product during November or December 1997.[13] Second, we drew a stratified sample from the November–December purchase pool. Specifically, we divided the pool into 13 strata depending on which product they bought. If there were more than 1,000 purchasers of a particular product, we randomly sampled 1,000. This yielded a maximum of 1,000 purchases of each product; the minimum was 328. The total sample size was 10,000. Third, we drew a holdout sample of 11,287 customers who purchased during January 1998.[14]

There are three issues to investigate in our test: statistical technique, independent variables, and sampling technique. There are four statistical techniques (logistic regression, multinomial logit, discriminant analysis, and neural networks), three sets of independent variables (product ownership, demographics, and monetary value), and two samples (simple and stratified). In addition, product ownership could be coded 0-1 or as the number of accounts. Certain combinations of these variables were not possible (e.g., we had to use at least one independent variable set and at least one statistical technique). We therefore tested a total of 88 combinations spanning the reasonable possibilities.

Results. Table 4 displays the predictive accuracy results for 88 models. Predictive accuracy based on first- or second-highest product ranges from about 40% to 55%, clearly exceeding the two-guess benchmark of $2/13 = 15.4\%$.

An instructive way to evaluate these results is to code the methods as indicator variables and use them as independent variables in a regression where predictive accuracy is the dependent variable.[15] The βs for this regression show the effect of the various methodological choices on predictive accuracy. Specifically, for each model, or row in Table 4, define:

ACCURACY = Predictive accuracy of the model.
PRODUCTS = 1 if current products owned is included in the model.
0 otherwise.
0-1 = 1 if current products owned coded as a 0-1 variable.
0 otherwise.
DEMOS = 1 if demographics included.
0 otherwise.

[13] The total pool was 18,690 customers who had purchased a product during November or December 1997. We randomly drew a 50% sample.

[14] Note the selection of a holdout purchase period $t + 1$ different than the calibration purchase period $t + 1$ (January vs. November–December) provides a test of whether seasonal factors diminish predictive accuracy.

[15] This analysis is in the spirit of a meta-analysis (Lehmann et al., 1998). See Lodish et al. (1995) and Abramson, Andrews, Currim, and Jones (2000) for a similar approach.

TABLE 4
Predictive Accuracy of Various Models

Percentage Correct First or Second Prediction

Predictor Variables				*Statistical Technique*			
Current Prod. Ownership	*Demographics*	*Activity Level*	*Sampling for Calibration Sample*	*Discrim Analysis*	*Multinomial Logit*	*Logistic*	*Neural Net*
Number	No	No	Stratified	47.0	47.9	48.0	50.9
0–1	No	No	Stratified	48.1	49.1	48.6	47.5
Number	Yes	Yes	Stratified	47.7	48.6	48.6	49.8
0–1	Yes	Yes	Stratified	48.7	49.6	49.3	50.1
No	Yes	Yes	Stratified	45.5	45.9	46.4	47.9
No	Yes	No	Stratified	38.6	38.8	38.3	39.0
Number	Yes	No	Stratified	46.9	47.6	47.7	48.3
0–1	Yes	No	Stratified	48.1	48.5	48.6	49.8
No	No	Yes	Stratified	40.7	42.3	42.5	40.9
Number	No	Yes	Stratified	47.1	48.6	48.8	50.4
0–1	No	Yes	Stratified	48.0	49.6	49.0	51.2
Number	No	No	Simple	52.9	53.4	52.8	51.8
0–1	No	No	Simple	53.6	53.3	52.9	53.9
Number	Yes	Yes	Simple	54.9	55.1	55.0	55.5
0–1	Yes	Yes	Simple	55.0	54.9	55.1	55.6
No	Yes	Yes	Simple	51.8	52.4	52.1	52.9
No	Yes	No	Simple	47.4	47.6	47.8	47.4
Number	Yes	No	Simple	53.9	53.6	54.0	54.6
0–1	Yes	No	Simple	54.3	54.5	54.1	54.6
No	No	Yes	Simple	48.1	48.3	48.1	50.0
Number	No	Yes	Simple	54.3	54.3	54.1	53.4
0–1	No	Yes	Simple	54.5	54.4	54.2	55.2

VALUE = 1 if customer monetary value included in the model.[16]
0 otherwise.
SAMPLE = 1 if random sample used.
0 if nonrandom sample used.
DISCRIM = 1 if discriminant analysis used.
0 otherwise.
LOGIST = 1 if logistic regression used.
0 otherwise.

[16] Monetary value was operationalized using two variables: the logarithm of the total value of all deposit-type accounts, and the logarithm of the total value of all loan-type accounts. Both these variables were included when VALUE = 1, and neither was includes when VALUE = 0.

MNLOGIT = 1 if multinomial logit used.
0 otherwise

We ran a regression with ACCURACY as the dependent variable and the other variables listed above as independents, using the data in Table 4. The results are in Table 5.

The overall F-tests shown in Table 5 indicate that the type of customer data and the sampling technique have statistically significant impacts on predictive accuracy. The effect of statistical method is marginal (p value = .104). The specific findings follow.

First, simple random sampling increases predictive accuracy significantly. This is because with

TABLE 5
Predictive Accuracy As a Function of Method

Predictive Accuracy As a Function of Method

Category	*F-Statistic*	*Variable*	*Regression Coefficient*	*t-Statistic*	*Significance Level*
		Constant	41.66	80.10	.000
Customer data	77.7 ($p < .001$)	PRODUCTS	5.76	14.75	.000
		0–1	.51	1.45	.152
		DEMOS	.83	2.69	.009
		ACTIVITY	1.83	5.93	.000
Sampling	392.0 ($p < .001$)	SAMPLE	5.98	19.80	.000
Method	2.13 ($p = .104$)	DISCRIM	−1.07	−2.50	.014
		LOGIST	−.66	−1.55	.124
		MNLOGIT	−.56	−1.31	.194

$R^2 = .988$, $n = 88$ models.

simple random sampling, we obtain a more representative baseline sales level for each product.

Second, while all three types of data improve predictive accuracy, current product ownership is clearly most important, followed by monetary value and demographics.[17] Whether product ownership is defined as 0-1 or the number of accounts owned for each product does not make a big difference, with a slight edge toward defining as 0-1. Exactly why current ownership is such a strong predictor is an interesting question. It could be due to customer learning that occurs when customers own a particular product (e.g., see Gedenk & Neslin, 1999). In addition, the result is consistent with Rossi, McCulloch, and Allenby (1996), who similarly found previous behavior to be the best prediction of future behavior in a coupon targeting application.

Third, while the overall F-test is only marginally significant, there is some indication of the power of neural networks. Neural networks are marginally better than logistic and multinomial logit, although the improvement is not statistically significant. The advantage over discriminant analysis is statistically significant and amounts to one percentage point in additional accuracy. Overall, the advantage of neural nets is not strong statistically and the magnitude of the advantage is smaller than the impact of data and sampling.

It is interesting to note we found that sometimes a model predicted well by overpredicting the popular products and underpredicting the less popular products. Using product ownership variables and stratified samples significantly reduced this tendency. We believe the product ownership variables help because many products are self-generating (recall the discussion of product affinities). So when we include product ownership, the model has a basis for predicting Product A, even if this product is not popular. The stratified sample over-samples less popular brands, so we gain a better range of experience for understanding the factors that contribute to buying this product. These observations support the use of product ownership variables in NPTB models. They result in accurate aggregate predictions and predict a range of potential products. The simple random sample produces more accurate aggregate predictions but

[17] Please note one demographic, namely age, was not available for this test. Using this variable for a subtest with another sample indicated that age would add significantly to the value of demographics, although not more than having product ownership. Age adds predictive accuracy because it is key for predicting certain banking products such as retirement accounts.

may underpredict unpopular products. If the goal simply is to find the first- or second-most likely product for a customer, random sampling is best. However, if one wants to predict a more varied set of products, a stratified sample would be preferred.

Incorporating Purchase Incidence

Overview. The NPTB model analyzed so far attempts to predict the product the customer is most likely to buy next, and targets that product to the customer. But missing is a consideration of *when* the customer is likely to buy. It may be that the customer is likely to purchase a CD next, but what if "next" means 3 years from now? The high potential CD buyer may really be a poor target if the timing of the purchase is not likely for a long time. Targeting that customer now is premature.

There is a long history of models in the marketing literature that consider when the customer will buy. These include approaches that consider what and when jointly (Ben-Akiva & Lerman, 1991; Haldar & Rao, 1998; Chintagunta & Prasad, 1998), as well as several standalone purchase incidence models based for example on hazard models or regression (Neslin, Henderson, & Quelch, 1985; Gupta, 1988; Jain & Vilcassim, 1991; Helsen & Schmittlein, 1993).[18] In this section we explore the potential gain of incorporating a purchase incidence model in predicting NPTB.

Assume a firm is about to embark on a direct mail campaign for product j, and wants to take into account not only whether the customer is likely to next buy product j, but whether the customer is likely to buy over the next 3 months (when the campaign will be conducted). Let:

A = The event that the customer buys some banking product over the next 3 months.

B = The event that the customer next buys a CD.

[18] Note another ad hoc way to incorporate purchase incidence in neural nets or multinomial logit models is to define an additional "product," namely "did not buy in this month," and estimate the model on a monthly basis.

We are then interested in:

$$\text{Prob}(A \text{ and } B) = \text{Prob}(A) \times \text{Prob}(B|A) \quad (5)$$

Prob(B | A) is simply the NPTB probability we have been calculating all along. Prob(A) is the purchase incidence prediction. Under this extended model, it is possible that the customer might have a large Prob(B | A), meaning that the next product they buy is likely to be a CD. However Prob(A) may be small. As a result, the overall probability the customer would next buy a CD over the next 3 months will be low and the customer should not be targeted.

Application and data. The example will also be a retail bank. Data were available for 270,842 customers who opened 425,763 new accounts, from among 14 retail products, over the 5-year horizon of January 1, 1995, through December 31, 1999. The NPTB and purchase incidence models were estimated on these data, and tested on 20,000 randomly selected households over the period January 1, 2000, through March 31, 2000.

Model selection. As mentioned earlier, there are many different approaches to modeling equation (5). In selecting a model for this exploratory analysis, one is immediately struck by the sparseness of the data. The average customer made 425,763/270,842 = 1.57 purchases over a 5-year period. That is 0.31 purchases per year or one every 3.18 years. This is a very different scenario than in the packaged good arena, where many of the marketing models have been applied. We decided to use a simple model that could be applied across all 270,842 customers, rather than a more sophisticated model that would be prohibitive to run over that many customers. We therefore chose a proportional hazards model (Cox, 1972) for purchase incidence and a logistic regression for the next-product model (Prob(B | A)). The hazard model can be stated as:

$$H(t) = H_0(t) \times \exp(\beta X) \quad (6)$$

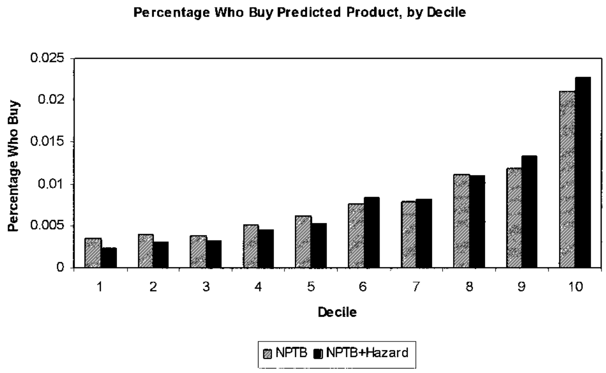

FIGURE 2
Comparative lift chart: NPTB model alone and NPTB + hazard model

H(t) is the hazard function or likelihood a purchase will take place at time *t* since the last purchase, $H_0(t)$ is the baseline hazard function that shows in general how the hazard function changes in the time since the last purchase. $Exp(X\beta)$ represents the impact of a set of predictor variables *X* such as current product ownership, RFM measures, and demographics, with β a set of parameters to be estimated. Once the baseline hazard function and the βs have been estimated, the model can generate "survivor" probabilities over any specified time period. This yields 1 − Prob(A), and therefore Prob(A). We then multiply Prob(A) times the NPTB probabilities, yielding the probability the customer is likely to buy a particular product over a 3-month span (equation 5). We can then rank order customers according to this score, and test whether these predictions hold in the test period.

Results. We included the following variables in the hazard model (as well as in the logistic regression NPTB model): current product ownership, years the customer has been with the bank, number of accounts opened in last 5 years, home equity and value, income, age, and a measure of market potential compiled by the bank. In all, this totaled 38 predictor variables. Many of these variables were statistically significant and the overall chi-square for the hazard model was highly significant.

We compared predictive accuracy for the test period for both the NPTB + Hazard and NPTB alone approaches. Figure 2 shows a "lift chart" comparing the two approaches. Customers have been ordered according to either Prob(B | A) (the NPTB approach) or Prob(A) × Prob(B | A) (the NPTB + Hazard approach) and then grouped into deciles. We then observe in the test period how many of them actually buy the product during the test period. Graphs such as Figure 2 are available for each product, but the summary in Figure 2 is across all products.

One sees in Figure 2 that the NPTB + Hazard approach improves over the NPTB approach. The NPTB + Hazard approach has more of the people it classifies in the top two deciles actually buy the product predicted, and fewer of the ones it classifies in the bottom deciles actually buy. It produces a better lift chart. The improvement shown in Figure 2 is not spectacular, but is suggestive of the potential of incorporating purchase incidence. In the banking industry, response rates are very low, so any improvement is important from a profitability standpoint.

To illustrate the profit implications of the model, we calculated the percentage increase in profits one could expect by targeting the top two deciles of the NPTB + Hazard approach versus the top two deciles of the NPTB approach. Assuming the same costs and profit margins, the difference is due to differences in purchase rates between the two groups. Figure 3 shows the results. The NPTB + Hazard approach yields up to 25% higher profits, e.g., for home equity loans, and performs well for credit cards, investment banking services, checking, and regular savings. It does not perform as well for IRAs and mortgages (MMA). The reason for this is not clear—there are no obvious differences in the nature of these products that would explain the lower performance.

Overall, the results suggest potential for incorporating purchase incidence into the NPTB framework. Our investigation is only exploratory, but suggests that even a relatively simple hazard model for incorporating incidence can improve results over the stand-alone NPTB model.

SUMMARY AND DISCUSSION

Our field test and methodological investigation suggest the following conclusions:

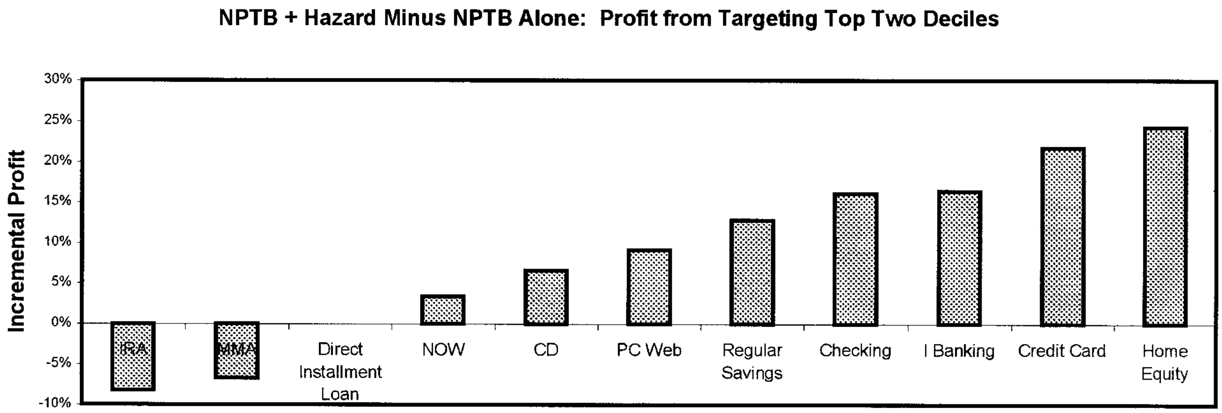

FIGURE 3
Comparative profits by product: NPTB model alone and NPTB + hazard model

- *NPTB models generate incremental cross-selling profits.* The gain is net of cross-channel cannibalization, and is greater than that using a commonsense but statistically nonrigorous heuristic for targeting customers. NPTB models also appear to be more effective than prospecting for new customers as a method of generating incremental sales. This suggests that NPTB models can help companies focus on customer retention rather than acquisition.
- *The most crucial predictor to include in an NPTB model is current product ownership.* This is consistent with the choice models literature (Rossi et al., 1996). Product ownership also generates diagnostics based on product affinities. Demographics and monetary value add predictive power, but product ownership is the single most valuable predictor. This may provide companies with a competitive advantage, in that competitors may be able to obtain demographic data on a company's customers, but they will have much more difficulty obtaining product ownership.
- *Predictive power is not much influenced by the statistical technique used for estimation.* Logistic regression, multinomial logit, discriminant analysis, and neural networks provide similar predictive accuracy, with an edge for neural nets (see Kumar, Rao, and Soni [1995] and Zhavi and Levin [1997] for further discussion of neural nets). Selection of the statistical technique can therefore be based largely on software availability and familiarity with the "ins and outs" of the technique.
- *The method for selecting customers for estimation can be either simple random or stratified random sampling.* If the goal is to find the most likely product(s) each customer will buy next, the simple random sample is preferred because of higher predictive accuracy. However, when there is large variation in product popularity, the model based on a random sample will skew its predictions, and hence its targeting recommendations, toward the most popular products because they are the most likely product for many customers. If this is a concern, the stratified random sample is better.
- *There is strong potential from incorporating purchase incidence models into the NPTB framework.* This can improve predictive accuracy and profits. However, the gains are not uniform, and more work is needed to hone in on the best approach for applying incidence models in cross-selling applications.

There are several additional issues to be researched. First, one could test additional predictors such as past marketing effort. Second would be a comparison between NPTB models and collaborative filtering (Mena, 1999; Oberndorf, 1999). Third, it would be valuable to develop optimal product targeting models. The question would be which products should be targeted to which customers to maximize profits, under the constraints that only a limited number can be targeted to each customer, and each product has a minimum sales target. Fourth would be to examine the effectiveness of NPTB models in managing the lifetime value of the customer (Nasr & Burger, 1998).

In short, NPTB models provide an already valuable tool for improving cross-selling efforts, yet there are many further research opportunities for making them even more valuable. We therefore view this article as merely a first step in a promising future research stream.

REFERENCES

Abramson, C., Andrews, R.L., Currim, I.S., & Jones, M. (2000). Parameter bias from unobserved effects in the multinomial logit model of consumer choice. Journal of Marketing Research, 37, 410–426.

Ben-Akiva, M., & Lerman, S.R. (1991). Discrete choice analysis: Theory and application to travel demand. Cambridge, MA: MIT Press.

Chintagunta, P.K., & Prasad, A.R. (1998). An empirical investigation of the "Dynamic McFadden" model of purchase timing and brand choice: Implications for market structure. Journal of Business and Economic Statistics, 16, 2–12.

Cox. D.R. (1972). Regression models and life tables. Journal of the Royal Statistical Society Series B. 34, 187–200.

David Shepard Associates. (1995). The new direct marketing (2nd ed.). New York: Irwin Professional Publishing.

Deighton, J., Peppers, D., & Rodgers, M. (1994). Consumer transaction databases: Present status and prospects. In R.C. Blattberg, R. Glazer, & J.D.C. Little (Eds.), The marketing information revolution. Boston, MA: Harvard Business School Press.

Dillon, W.R., & Goldstein, M. (1984). Multivariate analysis: Methods and applications. New York: Wiley.

Engel, J.F., Blackwell, R.D., & Miniard, P.W. (1995). Consumer behavior (8th ed.). Orlando, FL: The Dryden Press.

Garson, C.D. (1998). Neural networks: An introductory guide for social scientists. London: Sage Publications.

Gedenk, K., & Neslin, S.A. (1999). The role of retail promotion in determining future brand loyalty: Its effect on purchase event feedback. Journal of Retailing, 75, 433–459.

Gupta, S. (1998). Impact of sales promotions on when, what, and how much to buy. Journal of Marketing Research, 25, 342–355.

Haldar, S., & Rao, V.R. (1998). A micro-analytic threshold model for the timing of first purchases of durable goods. Applied Economics, 30, 959–974.

Helsen, K., & Schmittlein, D.C. (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. Marketing Science, 11, 395–414.

Hosmer, D.W., & Lemeshow, S. (2000). Applied logistic regression (2nd ed.). New York: Wiley.

Jain, D.C., Vilcassim, N.J., & Chintagunta, P.K. (1994). A random-coefficients logit brand-choice model applied to panel data. Journal of Business & Economic Statistics, 12, 317–328.

Kamakura, W.A., Ramaswami, S.N., & Srivastava, R.K. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. International Journal of Research in Marketing, 8, 329–349.

Kumar, A., Rao, V.R., & Soni, H. (1995). An empirical comparison of neural network and logistic regression models. Marketing Letters, 6, 251–264.

Lehmann, D.R., Gupta, S., & Steckel, J.H. (1998). Marketing research. Reading, MA: Addison-Wesley.

Lodish, L.M., Abraham, M., Kalmenson, S., Livelsberger, J., Lubetkin, B., Richardson, B., & Stevensk, M.E. (1995). How T.V. advertising works: A meta-analysis of 389 real world split cable T.V. Advertising experiments. Journal of Marketing Research, 32, 125–139.

Mena, J. (1999). Data mining your website. Boston: Digital Press.

Morrison, D.G. (1969). On the interpretation of discriminant analysis. Journal of Marketing Research, 6, 156–163.

Nasr, N.I., & Berger, P.D. (1998). Customer lifetime value: Marketing models and applications. Journal of Interactive Marketing, 12, 17–30.

Nash, E.L. (1993). Database marketing: The ultimate marketing tool. New York: McGraw-Hill.

Neslin, S.A., Henderson, C., & Quelch J. (1985). Consumer promotions and the acceleration of product purchases. Marketing Science, 2, 147–165.

Oberndorf, S. (1999, January). Predicting that next purchase. Catalog Age, 49.

Pfeifer, P.E. (1996). Capital One Financial Corporation: Response modeling. Case UVA-M-0507. Charlottesville, VA: University of Virginia Darden School Foundation.

Pindyck, R.S., & Rubinfeld, D.L. (1981). Econometric models and economic Forecasts (2nd ed.). New York: McGraw-Hill.

Rossi, P.E., McCulloch, R.E., & Allenby, G.M. (1996). The value of purchase history data in target marketing. Marketing Science, 15, 321–340.

Sharma, S. (1996). Applied multivariate techniques. New York: Wiley.

Siddarth, S., Bucklin, R.E., & Morrison, D.G. (1995). Making the cut: Modeling and analyzing choice set restriction in scanner panel data. 32, 255–267.

Stevens, J. (1996). Applied multivariate statistics for the social sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Zahavi, J., & Levin, N. (1997). Issues and problems in applying neural computing to target marketing. Journal of Direct Marketing, 11, 63–75.