MARKETING SCIENCE INSTITUTE

MSI Reports

Drivers of Technological Novelty and Superior Customer-Need Fulfillment in New Product Development (04-117)
Stefan Wuyts and Shantanu Dutta

Consumer Preferences for Mass Customization (04-118)
Benedict G. C. Dellaert and Stefan Stremersch

The Effect of Service Experiences over Time on a Supplier's Retention of Business Customers (04-119)
Ruth N. Bolton, Katherine N. Lemon, and Matthew D. Bramlett

A Spatial-Choice Model for Product Recommendations (04-120)

Sangkil Moon and Gary J. Russell

Firm-Sponsored Satisfaction Surveys: Positivity Effects on Customer Purchase Behavior? (04-121)
Utpal M. Dholakia, Vicki G. Morwitz, and Robert A.Westbrook

Analogies and Imaginary Consumers: A Case Study of New Product Development (04-122)
José Antonio Rosa, Steve Hoeffler, William Qualls, and Jonathan Bohlmann

2004

WORKING PAPER SERIES

ISSUE FOUR

NO. 04-004

MARKETING SCIENCE INSTITUTE

MSI Reports

2004

WORKING PAPER SERIES

ISSUE FOUR

NO. 04-004

The Marketing Science Institute supports academic research for the development—and practical translation—of leading-edge marketing knowledge on issues of importance to business performance. Topics are identified by the Board of Trustees, which represents MSI member corporations and the academic community. MSI supports academic studies on these issues and disseminates findings through conferences and workshops, as well as through its publications series.

Marketing Science Institute
1000 Massachusetts Avenue
Cambridge, MA
02138-5396

Phone: 617.491.2060
Fax: 617.491.2065
www.msi.org

MSI Reports (ISSN 1545-5041) is published quarterly by the Marketing Science Institute.

Working Paper Series
The articles that appear in *MSI Reports* have not undergone a formal academic review. They are released as part of the MSI Working Paper Series, and are distributed for the benefit of MSI corporate and academic members and the general public.

Subscriptions
Annual subscriptions to *MSI Reports* can be placed online at www.msi.org. Questions regarding subscriptions may be directed to pubs@msi.org.

Single reports
Articles in *MSI Reports* are available in downloadable (PDF) format at www.msi.org.

Past reports
MSI working papers published before 2003 are available as individual hard-copy reports; many are also available in downloadable (PDF) format. To order, go to www.msi.org.

Corporate members
MSI member company personnel receive all MSI reports (PDF and print versions) free of charge.

Academic members
Academics may qualify for free access to PDF (downloadable) versions of MSI reports and for special rates on other MSI print publications. For more information and to apply, go to "Qualify for academic membership" on www.msi.org.

Classroom use
Upon written request, MSI working papers may be copied for one-time classroom use free of charge. Please contact MSI to obtain permission.

Search for publications
See the searchable publications database at www.msi.org.

Submissions
MSI will consider a paper for inclusion in *MSI Reports*, even if the research was not originally supported by MSI, if the paper deals with a priority subject, represents a significant advance over existing literature, and has not been widely disseminated elsewhere. Only submissions from faculty members or doctoral students working with faculty advisors will be considered. "MSI Working Paper Guidelines" and "MSI 2004-2006 Research Priorities" are available in the Research section of www.msi.org.

Publication announcements
To sign up to receive MSI's electronic newsletter, go to www.msi.org.

Change of address
Send old and new address to pubs@msi.org.

A Spatial-Choice Model for Product Recommendations

Sangkil Moon and Gary J. Russell

Models can help managers match customers to products, but their predictive power depends on the firm's customer database. By "mapping" customers according to product preference similarities, this spatial-choice model captures the effects of variables typically absent from customer records that may drive choice behavior.

Sangkil Moon is Assistant Professor of Marketing, College of Management, North Carolina State University. **Gary J. Russell** is Professor of Marketing and Henry B. Tippie Research Fellow, Tippie School of Management, University of Iowa.

Report Summary
Product recommendation models are key tools in customer relationship management. Using information in a customer database, such models forecast the probability that a customer will buy a product.

In this study, Moon and Russell develop a product recommendation model based on the principle that customer preference similarity, as determined by prior purchase behavior, is a key element in predicting current product purchase.

Using a joint-space map based on past purchase behavior, they develop a predictive model in which the probability that a customer will buy a product depends upon the customer's relative distance on the map from customers who have already purchased the product. The space used in spatial modeling is thus a replacement for variables that may drive consumer choice, but which are usually absent from a firm's customer database.

By creating a map in which locational proximity is related to product preference, the spatial methodology allows the analyst to construct a model that implicitly contains more information about customer behavior than is apparent from the available data. In a certain sense, the spatial-choice methodology can be interpreted as a representation of unobserved heterogeneity in customer preferences.

Using data relating to purchases of insurance policies, Moon and Russell show that the proposed approach provides excellent forecasts relative to benchmark models.

For managers, the approach promises to be very effective in capturing the effects of variables that drive choice behavior but are not explicitly included in customer records: lifestyle, psychographics, financial resources, word-of-mouth, and product features, among others. In addition, the proposed model is flexible. It can integrate various sources of information on customer similarity and differentially weight this information to capture its effect on choice behavior. ■

Introduction

Product recommendation models are key tools in customer relationship management (CRM). An effective recommendation model contributes to the CRM goal of customer expansion by offering high-valued products to regular customers. When regular customers purchase more products from the same company, long-term customer retention is likely to be improved because of increased customer benefits and higher switching costs.

This research develops a product recommendation system based on the principle that customer preference similarity, as determined by prior response behavior, is a key element in predicting product purchase. Our study brings together two complementary research methodologies: joint-space mapping methodology (placing customers and products on the same psychometric map) and spatial-choice modeling (allowing observed choices to be correlated across customers). It differs from existing research in two respects. First, while marketing science models based on psychometric maps typically examine distances between products and customers in order to infer the attractiveness of each product to each customer (e.g., Moore and Winer 1987), we aim to infer product preferences from the relative similarity of customers to one another based on distances between them (Pollak 1976; Kapteyn et al. 1997; Yang and Allenby 2003). Because this approach relies only upon customer positions on a psychometric map, it is less susceptible to measurement error and does not require that the researcher accurately locate a new product on an existing product map.

Second, unlike previous work, this research uses a choice model adapted from the spatial-statistics literature (Cressie 1993; Russell and Petersen 2000; Bronnenberg and Mahajan 2001; Bronnenberg and Sismeiro 2002). Simply put, spatial models assume that entities (such as customers) can be located in a space. Responses by entities are assumed to be correlated in such a manner that entities near one another in the space generate similar outcomes. The methodology can integrate complex spatial correlations between entities into a model in a parsimonious and flexible manner. Typically, the space in a spatial model is physical geography. Our research differs from the spatial-statistics literature by analyzing a psychometric map rather than a physical map. Choice probabilities are predicted on the basis of an autologistic (AL) spatial model (Russell and Petersen 2000). Such a model, which takes a map as a key input, allows the researcher to construct a multivariate distribution of correlated binary (buy/not buy) response variables (Besag 1974; Cressie 1993). Because our AL model links these intercustomer correlations to distances on the psychometric map, we are able to infer product preferences based on a customer's map location.

Review of Product Recommendation Models

Product recommendation models match customers to products using information that allows the firm to infer product preferences. As shown in Table 1, models have been proposed in both the computer science and marketing science literatures. Below, we briefly review previous models, noting relationships to our proposed recommendation system.

Computer science literature

Many product recommendation models can be regarded as types of collaborative filtering (Herlocker et al. 1999; Ariely, Lynch, and Aparicio 2004). In collaborative filtering, a customer's choices are predicted using purchase information from other, similar customers. For example, nearest-neighbor algorithms may be created by computing the similarity between customers, based on their preference history, using cluster analysis or the Pearson correlation coefficient. Predictions of the degree to which a customer will like an unknown product are made by taking a weighted average of the opinions of a set of nearest neighbors for the target product.

Table 1
Comparison of Various Recommended Models

Model		Information Used (explanatory variables)	Information Forecasted (dependent variables)	Concept	Technique	Uses Statistical Theory and Testing
Computer science systems	collaborative filtering	overall preference ratings	overall preference ratings	customer similarity	cluster analysis	no
	content filtering	product attribute ratings	overall preference ratings	product similarity	regression and classification learner	no
	hybrid model	overall preference ratings and product attribute ratings	overall preference ratings	combining collaborative filtering and content filtering	regression, classification learner, and cluster analysis	no
Statistical-marketing models	Gershoff and West (1998)	product attribute ratings	overall preference ratings	community knowledge, similar to collaborative filtering	regression and residual analysis	yes
	Ansari, Essegaier, and Kohli (2000)	overall preference ratings, product attribute ratings, and customer characteristics	overall preference ratings	customer heterogeneity and product heterogeneity	hierarchical Bayesian model	yes
	Autologistic model (proposed in this research)	past purchases and customer characteristics for all customers; purchases of target products for calibration customers	binary choice	customer similarity	psychometric mapping (pick-any scaling) and spatial statistics (autologistic model)	yes

In contrast, content filtering (Mooney and Roy 2000; Ariely, Lynch, and Aparicio 2004) uses a given customer's product attribute ratings for other products to predict his or her response to a product of interest. The idea behind content filtering is similar to the theory underlying conjoint analysis: Preferences for a product can be predicted by appropriately weighting the values of product *attributes.* Whereas collaborative filtering requires overall preference ratings for different products, content filtering requires customers' ratings for product attributes over different products. The lighter data demands of collaborative filtering make it more popular than content filtering among Internet retailers such as amazon.com and barnesandnoble.com.

Despite its high popularity in commercial applications, collaborative filtering has been criticized for several reasons (Ansari, Essegaier, and Kohli 2000; Iacobucci, Arabie, and Bodapati 2000; Melville, Mooney, and Nagarajan 2002). First, collaborative filtering uses ad hoc prediction algorithms that ignore the statistical properties of the data. Second, many collaborative-filtering techniques are based on customer preference ratings instead of actual purchases. Typically, subjects rate products very selectively. Because the subject-product ratings matrix can be very sparse, the likelihood of finding a set of subjects with significantly similar ratings is usually low. Furthermore, customers tend to choose to rate products for which they have strong positive or negative reactions. This causes sampling bias and negatively impacts the predictive performance of the product recommendation system. The recent interest in hybrid models (Balabanovic and Shoham 1997; Pazzani 1999; Melville, Mooney, and Nagarajan 2002), which combine collaborative filtering with content filtering (Mooney and Roy 2000; Ariely, Lynch, and Aparicio 2004) is a reaction to this data sparsity issue.

Marketing science literature

Collaborative filtering, content filtering, and hybrid models are ad hoc in the sense that they are not based on a statistical-error theory. Although there are few recommendation models in the marketing science literature, marketing scholars have attempted to make use of statistical theory when building recommendation models, in order to reflect uncertainty in predictions. Gershoff and West (1998) augment an individual-level multiattribute model with prediction residuals from other customers to yield more accurate forecasts for a given customer. Their model uses features found in both content filtering and collaborative filtering: The basic utility specification is similar to a conjoint utility model, but information from other customers is used to predict the preferences of a given customer. Ansari, Essegaier, and Kohli (2000) develop a hierarchical Bayesian model in which the preferences of a given customer are predicted based on both customer demographics and product attributes. Again, an individual is assumed to be characterized by a multiattribute utility model with customer-specific attribute weights; however, pooling of information across customers permits more accurate forecasts. Hence, the marketing science models use the same basic ideas as the computer science models, but express these ideas in a considerably different manner.[1]

Proposed recommendation model

Our proposed model can be regarded as a statistically based collaborative filtering model in which the probability of purchase is the outcome measure. In common with collaborative filtering, our model predicts choice by weighing the choices of customers by the degree of similarity between customers. However, because the model is based on an explicit statistical model of the correlations in choice outcomes among customers, we are able to use the underlying theory both to define an appropriate measure of customer similarity and to test different types of model specifications. Moreover, because the procedure incorporates a psychometric map based on past purchases, we are able to infer customer preference for a target product without reliance on a specific multiattribute utility model or the need to define product attributes. We show subsequently that the use of past purchase information in a spatial-choice framework is an effective forecasting system.

Spatial-Choice Recommendation Model

A product recommendation model uses information in a customer database to forecast the probability that a customer will buy a product. It is assumed that marketers are interested in cross-selling a particular target product to regular customers who have not yet been asked to buy the product (cf. Kamakura, Ramaswami, and Srivastava 1991; Kamakura et al. 2003). The purpose here is to construct a model that allows a marketer to identify customers in the database who would be good prospects for solicitation regarding a target product.

Conceptual overview

The logic of the procedure is outlined in Figure 1. We assume that our customer database consists of two groups: customers who have had an opportunity to buy the target product (calibration customers), and customers who are currently unaware of the product (holdout customers). Information on all customers' purchases of nontarget products is available. The goal is to use this customer background information (for all customers) and actual purchases of the target product (for calibration customers) to predict the probability that each of the holdout customers will buy the target product.

Figure 1
Conceptual Overview of the Model

Construct Joint-Space Map
(Data: all customers' nontarget product purchases)

↓

Calibrate Spatial-Choice Model
(Data: distances on joint-space map between calibration customers and their target product purchases)

↓

Forecast Probability of Target Product Purchase
(Data: distances on joint-space map between calibration and holdout customers and calibration customers' target product purchases)

The procedure involves three steps. In the first step, we construct a joint-space map using the nontarget product purchase histories of all the customers in the database. This map contains an ideal point for each customer and a location for each nontarget product. In the second step, we build a predictive model based on the joint-space map and the actual purchases of the calibration customers. As we will show, this model links a given customer's purchase probability to the actual purchases of other customers using a measure of customer similarity derived from the joint-space map. In the final step, we use the predictive model to forecast the probability that each holdout customer will buy the target product.

Joint-space mapping

The proposed model is based on a measure of customer similarity derived from a joint-space map in which customers' ideal points and products are simultaneously represented. A pick-any dual-scaling algorithm is used to generate a joint space based on customer purchase records (Levine 1979; Holbrook, Moore, and Winer 1982). We construct a customer-by-product matrix in which binary variables (buy/not buy) indicate whether or not a customer has purchased a particular product in the past. Since we do not have sufficient information to locate the target product, we exclude the target product in constructing the matrix. An eignenvector analysis of the matrix then creates a map in which customers are located in the center of the products that they have already purchased. A more detailed discussion of the algorithm can be found in Appendix A.

Conceptually, the pick-any procedure estimates the map position of customers on the basis of similarity in product preferences. We define the preference similarity of two customers based on the proximity of their ideal points on the map. That is, for any two customers (a and b), we can compute the customer distance between ideal points as the Euclidian metric

$$CD_{ab} = \sum_r (a_r - b_r)^2, \tag{1}$$

where the summation runs over each dimension r of the constructed joint space. The dimension of the constructed map is $R - 1$, where R is the number of products included in the customer-by-product matrix. The predictive model developed below makes use of these distance measures.

Spatial choice: AL model

The second step of our procedure is to develop a model that uses patterns in the calibration customer group to infer the probabilities that the holdout customers will choose the target product. In this step, we make use of a spatial-statistical model in which the probability that one customer will make a purchase is linked to the actual purchases of other customers. The key assumption of our spatial-choice model is that customers who are near one another on the joint-space map have similar preference structures.

The model used here is a modified version of the AL regression model found in the spatial-statistics literature (Besag 1974; Cressie 1993; Russell and Petersen 2000). Let Z_c be a binary variable indicating whether or not customer *c* has purchased the target product. The AL model assumes that the probability that customer *c* will purchase the target product is given by the conditional logit model

$$Pr(Z_c = 1 \mid Z_{c'} \text{ for } c' \neq c) = [1 + \exp[-\{\alpha + \textstyle\sum_{c' \neq c} \theta_{cc'} Z_{c'}\}]^{-1}, \tag{2}$$

where $\theta_{cc'}$ are symmetrical parameters that measure the similarity between focal customer *c* and another customer *c'* ($\theta_{cc'} = \theta_{c'c}$). In words, the AL model states that we can predict purchase probability if we know the relative similarity of customer c to other customers ($\theta_{cc'}$) and whether the other customers have purchased the target product ($Z_{c'}$).

The model in Equation 2 implies that the outcome variables Z_c are correlated across customers. In statistical terminology, Equation 2 is called a full conditional distribution: It specifies the probability distribution of Z_c (for customer *c*) given the known values $Z_{c'}$ (for all other customers *c'*). Because these full conditional distributions (one for each customer in the database) are mutually dependent, we cannot analyze this model further unless we understand the joint distribution of all the *Z*'s. Put another way, the AL expression summarized by Equation 2 is a simultaneous way of specifying a statistical model of the choice outcomes for all customers in the database.

The form of the joint distribution of all the *Z*'s is not evident from simple inspection of Equation 2. However, by applying a key theorem developed by Besag (1974) and assuming that all intercustomer parameters are symmetric ($\theta_{cc'} = \theta_{c'c}$), it can be shown that the joint distribution of the Z_c variables ($c = 1, 2, \ldots N$) is given by the multivariate logistic distribution of Cox (1972). Further information on the multivariate logistic distribution and its relationship to the autologistic expressions in Equation 2 can be found in Appendix B. Further details on this important result are found in Cressie (1993) and in Russell and Petersen (2000). It should be noted that the symmetry of the $\theta_{cc'}$ coefficients is a key element in the model specification. If the $\theta_{cc'}$ coefficients are not symmetrical, the methodology fails because no joint distribution of the Z_c variables exists.

Defining customer similarity

In a multivariate logistic distribution, the $\theta_{cc'}$ parameters are functions of the correlations among the Z_c variables. Intuitively, this explains why the $\theta_{cc'}$ parameters must be symmetrical. In general, due to the fact that the correlations of a multivariate logistic distribution are unrestricted, the signs of the $\theta_{cc'}$ parameters can be positive, negative, or zero. However, in our application, due to the properties of the pick-any mapping procedure, we assume that consumers near one another on the map have similar preferences. For this reason, we constrain the values of $\theta_{cc'}$ to be non-negative. This logic justifies the interpretation of the $\theta_{cc'}$ parameters as symmetrical, ratio-scaled measures of customer similarity.

Because the number of distinct $\theta_{cc'}$ parameters equals $N(N-1)/2$ where N is the number of calibration customers in the database, it is infeasible to calibrate the model without further restrictions. We solve this problem by linking $\theta_{cc'}$ to the proximity between customers on the joint-space map. Specifically, we assume that

$$\theta_{cc'} = \beta\, CD_{cc'}^{-1} / \tau \tag{3}$$

where $CD_{cc'}^{-1}$ is the inverse of the distances defined in Equation 1, τ is a normalizing constant equal to the average of the $CD_{cc'}^{-1}$ measures across all pairs of calibration customers, and β is a proportionality constant to be estimated. Logically, we are assuming that the correlation in choices (as defined by the multivariate logistic distribution of the binary purchase variables) declines as the distance between customers on the map increases.

Using Equation 3, the specification of the AL model can be written in this compact form:

$$Pr(Z_c = 1 \mid Z_{c'} \text{ for } c' \neq c) = [1 + \exp[-\{\alpha + \beta\, CS_c\}]^{-1}, \tag{4}$$

where the implied customer similarity variable *(CS)*

$$CS_c = \sum_{c' \neq c} CD_{cc'}^{-1} Z_{c'} / \tau \tag{5}$$

depends upon both the intercustomer map distances and whether or not other customers *c'* have purchased the product. It should be emphasized here that working with full conditional distributions, as specified in Equations 4 and 5, is equivalent to specifying the structure of a multivariate distribution of binary purchase variables across the calibration customer population. However, the idea behind the model—that one can link purchase probability to the purchase behavior of similar consumers—is best understood by the conditional model specification given above.

Calibrating the AL model

As noted in Appendix B, direct use of a multivariate logistic distribution for parameter estimation is mathematically intractable. Accordingly, we follow the common practice in the spatial-statistics literature of using full conditional distributions (specified in Equations 4 and 5) to calibrate the model parameters (α and β). Note that the logit expression in Equation 4 is not independent across customers because each customer's purchase variable Z_c is related to the purchase variables of all the other calibration customers. Consequently, from the standpoint of maximum likelihood estimation theory, we should not multiply the N logit likelihoods generated by Equation 4 together to compute the overall likelihood function.

Nevertheless, it can be shown that maximizing the function obtained by multiplying together the logit likelihoods represented by Equation 4 yields consistent estimates of model parameters. This procedure is known as maximum pseudo-likelihood estimation (MPLE) (Cressie 1993).[2] For the autologistic model, the procedure is computationally simple since it amounts to using standard logit software to estimate the model parameters—ignoring the fact that the response variables are actually interdependent. However, the standard errors produced by logit software are not correct for MPLE. Appropriate standard errors for the MPLE procedure can be computed using the spatial-jackknife approach developed by Lele (1991). Details on this procedure are provided in Appendix C.

Forecasting holdout customer response

After the model is calibrated, forecasting the probability that holdout customers will purchase the target product is straightforward. We first compute customer similarity (*CS*) measures for each holdout customer, as in Equation 5. In this case, however, the summation in each *CS* measure runs over all the customers in the calibration data set. (Clearly, we do not know whether or not the other holdout customers will purchase the target product.) Once all the *CS* measures are defined, the AL model from Equation 4 is used to compute choice probabilities for each customer in the holdout sample.

Application

In this section, we use the AL model to forecast the reactions of customers to a solicitation to buy an insurance policy. Because the actual choices of all customers are known in this application, we are able to use these data to study the effectiveness of the proposed model in predicting choice. Results indicate excellent performance for the proposed AL model.

Data description

The data used in this study, supplied by the Dutch data-mining company Sentient Machine Research, were taken from the database of an insurance firm. The database provides variables related to sociodemographics and various insurance policy ownerships (22 products). Sociodemographic variables are based on the postal code of the residence of the customer. That is, all the people in the same postal code area have the same sociodemographic variables. One insurance policy (private third-party insurance) was selected as the target product for recommendation in this analysis.

From the complete data set, we selected 400 customers for the calibration sample and 300 customers for the holdout sample. As noted earlier, all customers are used in creating a joint-space map, whereas only the calibration customers are used to estimate model parameters. The calibration customers were selected using a stratified sampling method based on the number of insurance policies owned by each customer. We selected more heavy buyers rather than using a random sampling to reduce the problem of data sparsity in the customer-by-product matrix. The holdout sample was selected by random sampling. The two groups had no customers in common. As noted earlier, we were able to study the accuracy of different forecasting methods because the actual decisions (to buy or not to buy the focal policy) of all customers were known.

Constructing the joint-space map

We used nine insurance products (out of a possible set of 21 products, excluding the target product) to construct the joint-space map. These nine products were selected on the basis of a stepwise logit regression for the calibration customers, in which all nontarget insurance policy variables were used to predict ownership of the target insurance policy. We retained only those insurance ownership variables that had a positive logit coefficient and were statistically significant at the (extremely liberal) .50 level. This screening procedure was necessary to ensure that the map represented a product cluster that was relevant for forecasting the target product. In other words, it is necessary to take into account the relationship of each product to the target product in preparing a joint space (Solomon and Buchanan 1991).

The 700-by-9 (customer-by-product) matrix of binary variables (buy/not buy) denoting customer product ownerships was analyzed using Levine's (1979) pick-any scaling algorithm. Appendix A describes the procedure used to generate estimates of customer ideal points. As discussed previously, intercustomer distances from this map are used to create the *CS* variables of Equation 5. Thus, this pick-any map is a key component of the AL model.[3]

Variants of the AL model

It should be noted that sociodemographic variables are not used in creating the joint-space map in the AL model of Equations 4 and 5. This original model is called AL-P because the *CS* index is based on past purchases only. To study the role of demographics in forecasting choice, we also considered two variants of the model. The AL-D model uses a *CS* index based on demographics only. In calculating customer demographic similarity, the correlation between the demographic profiles of two customers replaced the $CD_{cc'}^{-1}$ term in defining of customer similarity (Equation 5). The AL-PD model, which combines both past purchases and demographics, has the following hybrid structure:

$$Pr(Z_c = 1 \mid Z_{c'} \text{ for } c' \neq c) = [1 + \exp[-\{\alpha + \beta_1 CS_c^{(P)} + \beta_2 CS_c^{(D)}\}]]^{-1} \quad (6)$$

where $CS_c^{(P)}$ is the customer similarity measure based on the pick-any map and $CS_c^{(D)}$ is the customer similarity measure based on demographics. In this equation, each source of *CS* is weighted differentially by the size of the β parameters when we compute the probability that customer *c* will make a purchase. Because the AL-P and AL-D models are nested inside the AL-PD model, we are able to comment on the explanatory power of past purchases relative to demographics in forecasting the choices of holdout customers.

Benchmark: Principal-components logit

We use four benchmark models to evaluate the performance of the proposed AL model. The first benchmark model, called the principal-components (PC) logit model, is a simple logit regression in which principal-component variables are used to forecast the probability of target product choice. To construct this model, we first analyzed all sociodemographics and nontarget insurance policy ownership variables using principal components. In doing this, we retained 23 principal components, accounting for 70% of the total variation of the original 82 raw explanatory variables. Using these 23 principal-component variables, we estimate the coefficients of a logit regression model using data from the calibration customers only. To forecast holdout customers' choice probability, we inserted the values of each holdout customer's principal-component variables (factor scores) into this logit model specification.

Benchmark: Moore-Winer model

The second benchmark is based on a marketing science model proposed by Moore and Winer (1987). Using data on customers' choice of ground caffeinated coffee, these authors measured the distance between the customer ideal point and the product location on a pick-any map, and then used these distance measures in a market share response model. Analogous to this approach, we define a product attractiveness (PA) variable that is the inverse of the distance between the target product and customer *c* on the joint-space map. This variable is a measure of the attractiveness of the target product to each customer. (See Gruca, Sudharshan, and Kumar 2002 for a similar model.) This benchmark model, which we call the Moore-Winer (MW) model, has the following specification:

$$Pr(Z_c = 1) = [1 + \exp[-\{\alpha^* + \beta^*\, PA_c\}]^{-1} \qquad (7)$$

where α^* and β^* are parameters to be estimated. We expect the β^* parameter to be positive.

In the MW model, the target product must be included on the pick-any map (in addition to the previously selected nine insurance products) because we need to compute the distance between the target product and each customer. This requirement creates a technical problem because, in practice, we do not know whether the holdout customers will choose the target product. For this reason, we first use the PC model to forecast the probabilities that the holdout customers will choose the target product. Then, we create a 700-by-10 (customers-by-products) matrix. In the matrix, the holdout customers' target product binary variables are replaced by forecasts generated from the PC model. This matrix is then analyzed using the pick-any algorithm to generate a joint space containing all customer ideal points and the location of the target product.

Benchmark: Collaborative filtering

The third benchmark model, collaborative filtering (CF), is drawn from the computer science literature (Schafer, Konstan, and Riedl 2001). This benchmark model is especially important because it is similar to models adopted by many Internet retailers. In this model, a Pearson correlation coefficient is used to compute customer similarity based on available customer information. Specifically, the model takes the form

$$\text{Forecast } Z_c = b + \textstyle\sum_{c' \neq c} r_{cc'}(Z_{c'} - b) / K_c \qquad (8)$$

where $r_{cc'}$ is the correlation between customers *c* and *c'*, $K_c = \sum_{c' \neq c} r_{cc'}$, and *b* is the base rate of the response. We define the base rate to be the overall proportion of calibration customers who buy the target product.

Two aspects of the CF model should be noted here. First, the CF model is not estimated by a statistical procedure. Once correlations are computed, Equation 8 is used to compute the expected value of Z_c. Since Z_c is a binary (0-1) variable in this study, we regard Forecast Z_c as an estimate of the probability that customer *c* will purchase the target product. Second, the summation over customers *c'* is interpreted to mean summation over a small neighborhood of customers with high correlations relative to customer *c*. For this study, we defined the neighborhood as the 20 customers closest (as measured by correlation) to the target customer. This definition, developed empirically, maximizes the performance of the CF procedure in our data set.

Table 2
Parameter Estimates of Various Models

	AL-P	AL-D	AL-PD	MW	PC
Intercept	.4287 (.1161)*	−.4072 (.1028)*	−.4306 (.1162)*	.0243 (.1569)	−2.0543 (.2415)*
Customer preference similarity	2.5205 (.2822)*	---	2.5008 (.2825)*	---	---
Customer demographics similarity	---	1.4124 (.8541)	.8281 (.9525)	---	---
Product attractiveness	---	---	---	.6230 (.0723)*	---
Principal-component variables	---	---	---	---	13 of 23 variables statistically significant at the .05 level
Log Likelihood	−227.1#	−267.9#	−226.7#	−168.0	−124.5
BIC	466.2#	547.8#	471.4#	348.0	392.8

AL-P = autologistic regression based on purchases, AL-D = autologistic regression based on demographics, AL-PD = autologistic regression based on both purchases and demographics, MW = Moore-Winer model, and PC = principal-components logit.
Standard errors are shown in parentheses. Estimates significant at the .05 level are denoted by *. The estimates of the principal-component variables used in PC are not shown. However, all principal-component variables are linear combinations of customer demographics and insurance purchases other than the target product.
Although log likelihood and BIC values are provided for the AL models, values denoted by # are *not* comparable to values for the other models due to the use of a pseudolikelihood procedure to calibrate the AL parameters. Comparison of values denoted by # only makes sense within the group of AL models.

As we did for the AL model, we examined three different CF variants. CF-P is the collaborative filtering model in which the customer's vector of past purchases is used to compute the correlation coefficients among customers. Note that past purchases include binary variables indicating purchase status of the nine nontarget products. CF-D is the collaborative filtering model in which the customer's vector of demographic variables is used to compute correlation coefficients. CF-PD is a hybrid procedure in which correlations are computed using a vector containing both past purchase indicators and demographics. In contrast to the autologistic hybrid procedure, we cannot differentially weigh the two sources of information. In all cases, forecasts are made by inserting the appropriate correlations into Equation 8.

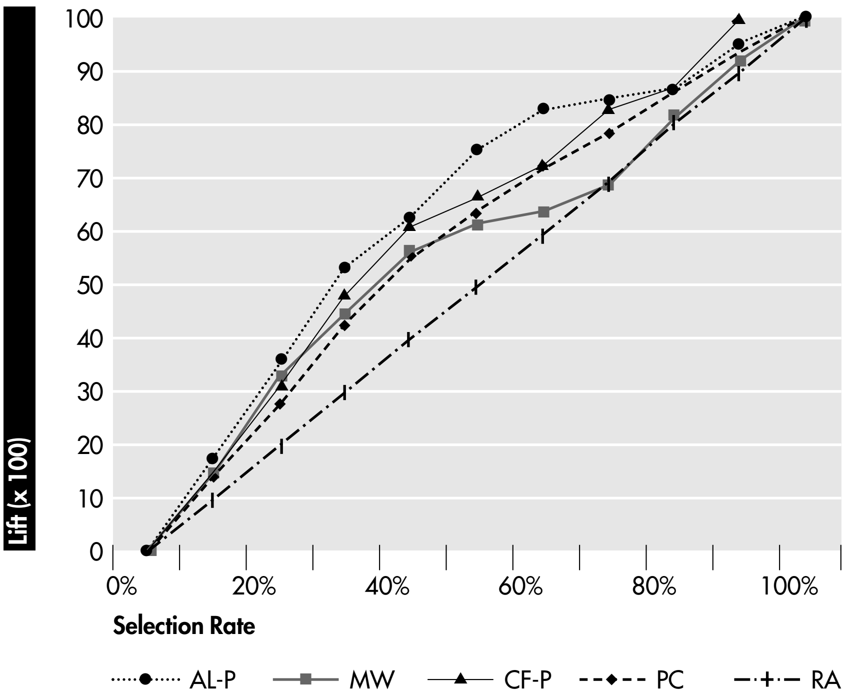
Benchmark: Neural networks

Neural-network (NN) models are known for superior predictive performance on problems involving many variables and extreme curvature in the response function (Hastie, Tibshirani, and Friedman 2001). For the purposes of comparison with the AL model, we calibrated an NN model using the nine nontarget products used for the pick-any map. Due to the relatively small number of inputs, we used an NN architecture consisting of one hidden layer with four nodes. All functions linking nodes were assumed to be in logistic form. An extensive discussion of neural-network architecture, estimation approaches, and forecasting accuracy can be found in Soucek (1991).

Empirical Findings

Results of model calibration for the three AL model variants, the MW model, and the PC model are shown in Table 2. Parameter estimates are not available for the CF model variants because the procedure is defined determin-

Figure 2
Lift Curves for Various Models



Note: AL-P = autologistic regression based on purchases, MW = Moore-Winer, CF-P = collaborative filtering using the 20 nearest neighbors in purchases, PC = principal-components Logit, and RA = random assignment. The lift statistic is defined as the proportion (%) of buyers in the holdout sample who belong to the selected group for product recommendation. The selected group is defined as the top *x%* of prospects selected by the forecasting procedure, where *x%* equals the selection rate predetermined by marketers.

istically. The NN parameter values are not shown because NN parameters are not readily interpretable (Cui and Wong 2004).[4] Although we present log likelihood and BIC (Bayesian information criterion) values for the models in Table 2, we caution the reader that the values for the AL models are not comparable to those for the other models. Because the AL models use a maximum pseudolikelihood estimation approach, the reported values do not correspond to the values that would be obtained by maximum likelihood analysis of the underlying multivariate logistic binary response model. All coefficients on customer similarity in the AL model and product attractiveness in the MW model in Table 2 have logically correct (positive) signs.

We take the ability of each model to forecast whether holdout customers will purchase the target product as an indicator of model performance. Adopting industry practice, we examined lift statistics on the holdout customer sample (sample size = 300) to evaluate the predictive validity of the models. We computed lift statistics using the following procedure: After estimating the parameters of a particular model, we used the model to predict the probability that each customer in the holdout sample would purchase the target product. Then, we sorted all the customers in the holdout sample in descending order by the probability that they would make the purchase. Finally, we selected the top x% of customers according to a predetermined selection rate. Specifically, we tested nine different selection rate levels (10%, 20%, …, 90%). A 20% selection rate means that we selected those customers that were in the top 20% with regard to the probability that they would buy the target product; a 30% selection rate means that we selected those customers that were in the top 30%, and so on.

Lift is defined as the proportion of target product customers within the holdout sample who are included in the selected group. Notice that in Figure 2 the lift statistics increase monotonically with the selection rate. This is due to the fact that a 100% selection rate (selecting all the customers in the holdout sample) must yield a lift statistic equal to one. We also included the lift statistics that would be expected if we selected customers at random (random assignment = RA). In the lift curve in Figure 2, the lift of random selection forms a 45-degree angle through the origin of the graph. Within a particular selection rate condition, higher values of lift statistics indicate a better procedure. For this reason, forecasting methods with higher lines in Figure 2 are more desirable.

The overall predictive validity is summarized by the Gini coefficient. The Gini coefficient, defined as the area between each model's cumulative lift curve and the random-assignment lift curve, varies between 0 and 1, with 0 indicating the same performance as random assignment (Curry 1992). These coefficients, along with the

Table 3
Lift Statistics for Various Models

Model	Selection Rate (%) 10	20	30	40	50	60	70	80	90	Gini Coefficient
NN	.175	.357	.540	.635	.738	.873	.897	.905	.944	.673
AL-PD	**.198**	**.357**	**.508**	**.635**	**.738**	**.825**	**.849**	**.905**	**.944**	**.618**
AL-P	**.175**	**.365**	**.532**	**.627**	**.754**	**.833**	**.841**	**.865**	**.952**	**.613**
CF-P	.143	.310	.484	.611	.667	.730	.833	.873	1.000	.389
PC	.143	.278	.429	.563	.635	.722	.786	.857	.937	.365
MW	.151	.325	.444	.563	.611	.635	.690	.810	.921	.252
CF-D	.143	.254	.341	.421	.524	.603	.683	.802	.889	.062
AL-D	.103	.191	.325	.429	.516	.627	.706	.825	.905	.058
CF-PD	.095	.206	.286	.389	.508	.611	.698	.778	.905	.002
RA	.100	.200	.300	.400	.500	.600	.700	.800	.900	.000

Comparison models: NN = neural network based on purchases, AL-PD = autologistic regression based on both purchases and demographics, AL-P = autologistic regression based on purchases, CF-P = collaborative filtering using the 20 nearest neighbors in purchases, PC = principal-components logit, MW = Moore-Winer model, CF-D = collaborative filtering using the 20 nearest neighbors in demographics, AL-D = autologistic regression based on demographics, CF-PD = collaborative filtering using the 20 nearest neighbors in both purchases and demographics, and RA = random assignment (null model).
Lift and selection rate: Lift is defined as the percentage of all buyers in the holdout data set who fall into the group of selected customers. For each model, the probabilities that the holdout customers would purchase the target product were forecasted. Holdout customers were then grouped according to their forecasted probability of choosing the product, and the top *x*% of the holdout customers were selected, where x% is the selection rate. We varied our selection rate; within each selection rate group, higher lift indicates a better procedure. By definition, as the selection rate approaches 100%, lift approaches unity.
Gini coefficient: The Gini coefficient is determined by the area between each model's cumulative lift curve and the random lift curve. The Gini coefficients in this table have been normalized so that a value of 1.0 corresponds to the best possible lift curve. Higher values of the Gini coefficient imply a better model. Models in this table are ranked in terms of the Gini coefficient.

lift statistics for each procedure, are shown in Table 3. In this table, procedures have been sorted by the Gini coefficient to bring out the patterns in the results.[5]

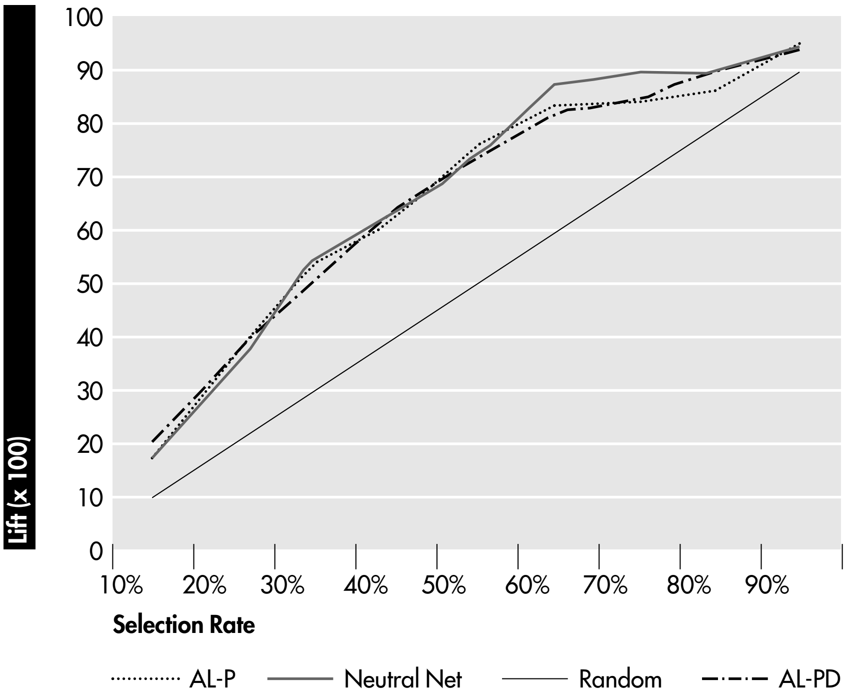
Model comparison

The AL-P model (based only upon past purchases) shows superior performance relative to the RA model and the non-NN benchmark models (MW, CF, and PC) in lift statistics.[6] This pattern is consistent across all selection rate groups except at the 80% and 90% selection rates, where the CF-P model slightly outperforms the AL-P model. (These exceptional cases for the CF model are unimportant in practice because managers will not find it economically justifiable to have extremely high selection rates.) Overall, the superior performance of the AL-P model is demonstrated by the fact that its Gini coefficient (.613) exceeds the Gini coefficients of all non-NN benchmark models (Table 3).

Observe that the AL-P model is better than the MW model even though both models use the same joint-space mapping methodology. We argue that this performance difference arises from the fact that the MW model is forced to use estimated target product ownership values for holdout customers to locate the target product on the map. The measurement error introduced by the use of estimates in the mapping procedure is apt to substantially degrade the performance of the MW model. (An experiment in which known values are inserted into the mapping procedure supports this conclusion.) The AL-P model escapes this difficulty because the approach does not require the researcher to use the target category in constructing the psychometric map.

In general, the results bring out the fact that the manner in which customer information enters the model plays a key role in forecasting per-

Figure 3
Lift Curves for AL and NN Models



formance. The CF-P model is inferior to the AL-P model despite the fact that both models employ some type of customer similarity index based on past purchases. Evidently, a similarity index based on proximity in a psychometric map (AL-P) provides a better approximation to true customer similarity than an index based on the correlation coefficient (CF-P). In addition, the AL-P model performs better than the PC model despite the fact that the PC model uses variables that summarize past purchase information. Again, we argue that this is due to the superior specification of the AL-P model. In the PC model, past purchase variables simply shift the mean purchase probability. In contrast, the AL-P model uses past purchase information to infer similarity of preferences across customers. The results presented in Table 3 strongly suggest that better predictions follow from models that use past purchase information to identify clusters of customers with similar preferences.

Predictive Power of Demographics. The explanatory power of demographics in this application is quite poor. Models that use only demographics (AL-D and CF-D) have extremely small Gini coefficients. In contrast, models that combine past purchases and demographics in some fashion (AL-PD, PC, and CF-PD) exhibit a wide variety of Gini coefficients. The models that perform relatively well (AL-PD and PC) provide for a differential weighting of purchase and demographic information. The performance of the CF-PD procedure is exceptionally poor. We argue that this is due to the fact that the CF-PD methodology combines customer information on both past purchases and demographics into one common index, ignoring the relative impact of each information source on the purchase probability. In general, the results in Table 3 are consistent with the view that past purchase information is considerably more valuable than demographics in predicting choice. When demographics are added to the AL-P model according to Equation 6's differential weighting formula, we obtain a slightly higher Gini coefficient (.618 for AL-PD versus .613 for AL-P). Clearly, some information is present in the demographic variables. However, this information is very weak relative to the information found in past purchase variables.

Comparison with the Neural-Network Model. As we expected, the Gini coefficients in Table 3 indicate that the NN model is globally superior to the AL-P and AL-PD models in terms of lift. However, careful examination of the lift curves in Figure 3 provides a more complex and interesting comparison of the forecasting ability of the models. As in Figure 2, we also show the curve corresponding to random assignment.

Although the NN is globally better than the AL models, all models have very similar lift profiles for selection rates of 50% or less—that is, for those customers who were in the top 50% with regard to the probability that they would buy the target product. Intuitively, this means that the top half of the customer rankings produced by the AL and NN models is similar. Only for extremely high selection rates is the NN model better than the AL models. From a practical perspective, this global dominance is not of interest. In most settings, profitability considerations will require a low selection rate. Hence, the AL and NN procedures produce remarkably similar forecasts in the region of the selection rate space that is relevant to marketing managers.

The AL models, in fact, are superior to the NN model in a number of ways. First, in contrast to the NN model, the AL models are based on an error theory that allows statistical testing of different model specifications. Second, the AL procedure is easily understood because forecasts are based on a pick-any map that may be viewed and discussed by managers. NN models, in contrast, are essentially "black box" forecasting tools (Hastie, Tibshirani, and Friedman 2001). Third, the AL model can be adapted to holdout data sets that differ in terms of base rate from the calibration data set. This can be easily accomplished by adjusting the α intercept parameter of the AL model higher or lower depending upon how the base rates differ from calibration to holdout. In contrast, the "black box" nature of the NN makes such base rate adjustments very difficult.

These considerations, along with the lift profiles in Figure 3, provide a very strong argument for adopting the AL spatial model in this application. The AL model provides a statistical error theory, managerial interpretability, and model flexibility without sacrificing forecasting ability over the relevant region of the selection space.

Summary

Our study provides strong evidence in support of the spatial-model approach in developing a product recommendation model. In theory, the space used in spatial modeling is a replacement for missing variables that actually drive response behavior (Cressie 1993; Bronnenberg and Mahajan 2001, Bronnenberg and Sismeiro 2002). For example, in ecological studies of waterways, physical geography is used in spatial models to represent the (unknown) sources of pollution. In the same way, the joint-space map obtained from past purchase behavior actually represents clusters of (unknown) variables that determine purchase behavior (e.g., word-of-mouth, network externalities, lifestyle). By relying on this map, the AL spatial model offers better predictive performance using the same customer background information as other procedures.

Conclusions

This research develops a product recommendation system based on a spatial-choice model. Customers are assumed to be located on a joint-space map in which individuals who are near one another share similar product preferences. By merging the AL spatial-choice model with this map, we are able to construct a statistically based method of forecasting product choices for customers in a firm's database. We show empirically that the AL model provides excellent forecasts relative to existing recommendation procedures.

Managerial contribution

The AL model assists database managers in two ways. First, the approach promises to be very effective in capturing the effects of variables that drive choice behavior but are not explicitly included in the customer records available to the researcher. For example, variables such as life-style, psychographics, financial resources, and product features often determine customer choices, but are typically absent from a firm's database. As long as a map can be created in which locational proximity is related to product preference, the spatial methodology allows the analyst to construct a model that implicitly contains more information about customer behavior than is apparent from the available data. In a certain sense, the spatial-choice methodology can be interpreted as a representation of unobserved heterogeneity in customer preferences. However, instead of using statistical distributions to represent heterogeneity, the spatial-choice methodology weights the choice behavior of other customers.

Second, the proposed model is flexible: It can integrate various sources of information on customer similarity and differentially weight this information to capture its effect on choice behavior. In our empirical work, we showed that the differential weighting of information is critical in developing useful forecasts. Methods such as collaborative filtering, which do not allow for differential weighting, perform worse than procedures that do allow for it. Moreover, because our procedure allows for statistical testing, the researcher can easily examine the usefulness of different data sources in improving forecasting. For example, we found that the coefficients on the demographic similarity index for the AL models were statistically insignificant (Table 2), a finding that was entirely consistent with the predictive performance of demographic information among holdout customers (Table 3). This ability to assess the information content of different information sources allows managers to determine the minimal set of variables needed to effectively forecast choice behavior.

Working with large customer databases

A key question is whether the proposed AL model can be used when the firm is processing large amounts of customer data. This issue, known in the data-mining literature as scalability, is usually interpreted to mean that the data-mining model must be able to be calibrated on a large data set (Soucek 1991). However, Kim et al. (in press) provide empirical evidence that data-mining models can also be made scalable in a "sampling, then roll-out" (SRO) fashion; that is, the model can first be calibrated on a representative sample of customers and then used to forecast to a larger customer database. In fact, the proposed AL model can be made SRO scalable.

Briefly, a SRO-scalable AL model could be constructed as follows. First, the researcher would develop a pick-any map using a sample of customer records. This is not necessarily a random selection of customers. In fact, a much better approach would be to use a stratified random sample in which sampling strata are based on customer response profiles for the nontarget products. The idea is to span the profile space without doing an enumeration of all possible nontarget category buying profiles. Second, the pick-any map and the AL model parameters would be calibrated using this (small) sample of customers. In essence, the researcher would need a selection of customers that provides a large amount of information on the model structure. If done properly, this need not translate into a large number of customer records.

At this point, the model would be fully calibrated. To forecast the probability that a holdout customer would make a purchase requires three additional steps. First, the researcher would place the customer on the existing pick-any map. (The pick-any map and the AL parameters are not recomputed.) This is possible because the pick-any model equations (see Equation A3 in Appendix A) state that each customer's ideal point is a weighted sum of the elements of the nontarget category buying profile. That is, a simple linear weighting of the nontarget buying profile will provide an estimate of the holdout customer's ideal point on the existing pick-any map. Second, using the pick-any ideal points of the set of calibration customers, the researcher would then compute the distances between the holdout customer and each of the calibration customers. (There are N computations required, where N is the number of customers in the calibration data set.) These distances provide the researcher with sufficient information to compute the customer similarity variable in Equation 5. Finally, using the known AL model parameters and the customer similarity value, the researcher would then compute the holdout customer purchase probability using Equation 4.

The algorithm outlined above is clearly SRO scalable. Given the model structure, it is possible to calibrate the model (pick-any map and AL model parameters) using a relatively small number of customers. Once this is done, the remaining steps are not computationally burdensome. Accordingly, the AL model can be implemented in CRM settings with a large number of customer records.

Model extensions

The AL forecasting model can be extended in a variety of ways. The most obvious extension would be a recommendation system in which forecasts are prepared simultaneously for multiple target products. In this setting, the researcher is assumed to be interested in predicting the likely purchase probabilities of a portfolio of products. This entails considering a model of binary variables Z_{cp} in which the subscript c denotes a customer and the subscript p denotes a target product. Embedding this structure in the multivariate logistic model requires the researcher to consider the correlation linking multiple customers and multiple products. In practical terms, such a model could improve forecasting by taking into account how the likelihood of purchasing in one category affects the likelihood of purchasing in another category. This approach could prove especially useful if the products in the portfolio of purchases tend to be strong complements. Further generalizations of the model to incorporate temporal dynamics are also desirable. This would entail considering a model of binary variables Z_{cpt} in which the subscript t denotes time. This step would be important for repeatedly purchased product categories. ■

Acknowledgements

The authors thank Sentient Machine Research, a Dutch data-mining company, for providing access to the data used in this research. They also thank Wagner Kamakura for providing detailed comments on an earlier draft of this paper. Financial support from the Marketing Science Institute, the North Carolina State University Edwin Gill Research Grant, and the Teradata Center for Customer Relationship Management at Duke University is gratefully acknowledged. The first author (Moon) was awarded first prize in the 2002 MSI Clayton Doctoral Dissertation Proposal Competition for the work reported in this article.

Appendix A. Generating a Joint Space Using Pick-Any Scaling

Both the AL regression model and the Moore-Winer (MW) model employed in this research make use of a joint space constructed with pick-any scaling. This appendix briefly reviews the procedure used to create a pick-any map.

Pick-any scaling was originally developed in the area of psychometrics (Levine 1979) and has seen numerous applications in marketing science (Holbrook, Moore, and Winer 1982; Moore and Winer 1987; Gruca, Sudharshan, and Kumar 2002). The method is designed around a data collection procedure in which respondents (e.g., customers) are allowed to select as many objects (e.g., products) as they desire out of a set of objects presented by the researcher. In marketing, the objects are typically products, and customers are asked to select products that match certain criteria. In the research reported here, we defined the pick-any set of products as those products that the customer had already purchased in the past.

Pick-any scaling (Levine 1979) creates a map with two properties: customers are located in the center of the products that they select, and products are located in the center of customers who select them. The mapping algorithm can be formulated as an eignenvalue problem as follows: Let N be the number of customers and K be the number of products. Define E as a symmetric matrix of dimension $N + K$ with the form

$$E = \begin{bmatrix} \phi_N & B \\ B' & \phi_K \end{bmatrix}, \tag{A1}$$

where B is an N-by-K matrix of zeros and ones with elements b_{ij} indicating the choice of product j by respondent i. Both ϕ_N and ϕ_K are N-by-N and K-by-K null matrices. D is defined as a diagonal matrix of order $N + K$ with diagonal elements

$$d_{ii} = \sum_{j=1}^{N+K} e_{ij}. \tag{A2}$$

For $i \le N$, d_{ii} represents the number of products chosen by respondent i, and for $N + 1 \le i \le N + K$, d_{ii} is the number of respondents choosing product $i - N$. Finally, let $\boldsymbol{x}_r$ be a column vector of $N + K$ coordinates for the N respondents (first N rows) and K products (last K rows) of the rth dimension of the joint space.

Using this notation, Levine's (1979) procedure solves the eigenvalue problem

$$\lambda_r x_r = D^{-1} E \boldsymbol{x_r}, \tag{A3}$$

where λ_r and $\boldsymbol{x}_r$ are the rth eigenvalue and eigenvector (respectively) of the nonsymmetric matrix $D^{-1}E$. In words, this states that a customer's coordinate on dimension k of the map must be proportional (through λ_r) to the centroid of the products he or she selected. Simultaneously, a product's coordinate on dimension r of the map must be proportional (through λ_r) to the centroid of the customers who selected the product.

For computational simplicity, this problem is restated as

$$\lambda_r(\boldsymbol{D^{1/2} x_r}) = D^{1/2}(D^{-1}E)\boldsymbol{x_r} = W(\boldsymbol{D^{1/2} x_r}), \tag{A4}$$

where $W = D^{-1/2}ED^{-1/2}$. This formulation is a symmetric eigenvalue problem where λ_r and $\boldsymbol{D^{1/2} x_r}$ are eigenvalues and eigenvectors (respectively) of the symmetric matrix W. The coordinates of the products and the customers can therefore be obtained by calculating the eigenvectors of W, and then premultiplying them by $D^{-1/2}$ to obtain $D^{-1/2}(\boldsymbol{D^{1/2} x_r}) = \boldsymbol{x_r}$.

Appendix B. Multivariate Logistic Distribution

Multivariate logistic distribution describes the probabilities associated with a vector of N binary responses $\boldsymbol{Z} = [Z_1, Z_2, \ldots Z_N]$. Each of these binary responses can take on two values (0 or 1) without restriction. That is, the fact that Z_i has taken on particular value does not restrict the possible values of any other variable Z_j. For this reason, there exist 2^N possible values of the $\boldsymbol{Z}$ vector: Any string of length N containing only 0's and 1's is allowed.

Let $\boldsymbol{X} = [X_1, X_2, \ldots X_N]$ be a realization of $\boldsymbol{Z}$. Then Cox's (1972) multivariate logistic distribution is defined as

$$Pr(\boldsymbol{Z} = \boldsymbol{X}) = \exp\{\boldsymbol{\Phi}(\boldsymbol{X})\} / \Sigma_{\boldsymbol{X}^*} \exp\{\boldsymbol{\Phi}(\boldsymbol{X}^*)\}, \tag{B1}$$

where $\boldsymbol{\Phi}(\boldsymbol{X}) = \Sigma_i\, \alpha_i \boldsymbol{X}_i + \Sigma_{i<j}\, \theta_{ij} \boldsymbol{X}_i \boldsymbol{X}_j$ and the summation in the denominator runs over all 2^N possible values of the Z vector. The cross-effect coefficients θ_{ij} are symmetrical in i and j due (in part) to the fact that the correlations among the $\boldsymbol{Z}$'s are determined by these parameters.

The moments of this distribution are complex. However, we can make some sense of the situation by observing that the structure of Equation B1 is analogous to that of a log linear analysis of an N-way contingency table consisting of main effects (α_i) and (symmetric) two-way interactions (θ_{ij}). One special case is worth noting. When all the cross-effect parameters θ_{ij} are zero, then the $\boldsymbol{Z}_i$'s are independent. In our work, this special case implies that choices by one customer cannot provide information about another customer.

Autologistic Model

Given the structure of Equation B1, it is easily seen that the full conditional distributions of the multivariate logistic distribution have the form

$$Pr(\boldsymbol{Z}_i = 1 \mid \boldsymbol{Z}_j \text{ for } j \neq i) = [1 + \exp[-\{\alpha_i + \Sigma_{j \neq i}\, \theta_{ij} \boldsymbol{Z}_j\}]^{-1}, \tag{B2}$$

where again we note that the cross-effect θ_{ij} parameters are symmetrical. Equation B2 is called an AL regression model because we use a logit function and the realizations of other $\boldsymbol{Z}$'s to predict a particular $\boldsymbol{Z}_i$. By construction, the AL model is also the set of N full conditional distributions implied by the multivariate logistic distribution.

Although Equation B1 implies Equation B2, it is not clear from inspection that the converse is also true. However, using a seminal theorem developed by Besag (1974), it can be shown that Equation B2, along with the symmetry of the cross-effect parameters θ_{ij}, implies that the $\boldsymbol{Z}$'s have the multivariate logistic distribution shown in Equation B1. Details are provided in Cressie (1993).

Model Calibration Using the Pseudolikelihood Approach

In our research, we specify the model by assuming that the full conditional distributions of the $\boldsymbol{Z}$'s are given by Equation B2, with the constraint that all α_i parameters have the same value α. We also constrain the θ_{ij} parameters to be proportional to the inverse of the distances between customers on the pick-any map (see Equation 3). The logic presented here then implies that the $\boldsymbol{Z}$'s collectively have the multivariate logistic distribution given by Equation B1, again with the parameter constraints noted above.

If the parameters of the model were to be estimated using maximum likelihood methods, we would ideally like to make use of Equation B1. Note, however, that the denominator of Equation B1 involves 2^N terms, one for each possible value of the $\boldsymbol{Z}$ vector. (Recall that N is equal to the number of customers in the calibration data set.) Because the dimension of the denominator creates computational problems, Cressie (1993) recommends that the parameters of Equation B1 be estimated using the full conditional distributions in Equation B2. The method of pseudolikelihood entails multiplying the N likelihood functions implied by Equation B2 together and regarding this product as the likelihood to be maximized. It can be shown that this approach, though not efficient, yields consistent estimates of model parameters. We use the pseudolikelihood approach for model calibration in our empirical work.

Appendix C. Jackknifing Spatially Dependent Observations

The maximum pseudolikelihood estimation procedure used to estimate the AL model parameters does not provide correct values of standard errors. For this reason, we use a jackknife method developed for spatially dependent observations (Lele 1991) to compute standard errors. This section outlines the details of the procedure.

Let $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_N)$ be a vector of N spatially dependent random variables having a joint distribution $F(\boldsymbol{Z}, \boldsymbol{\theta})$. Let

$$G(\boldsymbol{Z}, \boldsymbol{\theta}) = \sum_{c=1}^{N} g_c(Z,\theta) = 0 \qquad \text{(C1)}$$

be an estimating equation for $\boldsymbol{\theta}$. That is, a vector $\boldsymbol{\theta}$ satisfying Equation C1 is a consistent estimator of the parameters of $F(\boldsymbol{Z}, \boldsymbol{\theta})$.

In the usual jackknife procedure, we delete one observation at a time and estimate $\boldsymbol{\theta}$ from the remaining observations. However, removal of data segments from a serially correlated sequence of observations causes statistical difficulties. Instead, Lele's (1991) procedure deletes one component of a linear estimating equation at a time. Let θ_N be the estimator implied by Equation C1. In addition, let $\theta_{N,-j}$ be the estimator produced by solving

$$G^j(\boldsymbol{Z}, \boldsymbol{\theta}) = \sum_{c \neq 1} g_c(Z,\theta) = 0, \qquad \text{(C2)}$$

an expression identical to Equation C1 with the exception of the jth term. That is, $\theta_{N,-j}$ is an estimator constructed by deleting the jth term of estimating Equation C1.

Using this notation, we can define the jackknife estimate of $\boldsymbol{\theta}$ as

$$JK\,\theta_N(Z_1, Z_2, \ldots, Z_N) = \theta_N - \frac{N-1}{N} \sum_j R_{Nj}, \qquad \text{(C3)}$$

where $R_{Nj} = \theta_{N,-j} - \theta_N$. Also, we can define the jackknife estimate of variance as

$$JKV\,\theta_N = (N-1) \sum_{c=1}^{N} \sum_{j \in N(c)} (R_{Nc} - \bar{R}_N)(R_{Nj} - \bar{R}_N). \qquad \text{(C4)}$$

In this equation, $\bar{R}_N = (1/N) \sum_c R_{Nc}$ and $N(c)$ denotes a neighborhood of data point c in which correlations among observations are expected to be nonzero. Finally, the standard error of the jackknife estimator for the $\boldsymbol{\theta}$ estimator is

$$SE(\boldsymbol{\theta}) = \sqrt{JKV\theta_N}\big/N \qquad \text{(C5)}$$

where N is the number of data points.

To implement this procedure with the AL model, we recall that the pseudolikelihood expression involves N terms from the logit likelihood, one corresponding to each customer. Obtaining parameter estimates using all N logit likelihood terms corresponds to Equation C1. Obtaining parameter estimates after deleting the logit term corresponding to the jth customer corresponds to Equation C2. Accordingly, by running a standard logit maximum likelihood procedure N times and dropping one of the logit terms each time, we can construct N estimates of model parameters.

The variation among these N estimates is then used to construct standard errors. The definition of each customer's neighborhood is determined by how similar other customers are relative to the focal customer. In our model, the similarity of a customer pair is proportional to the inverse of the distance between them on the joint-space map. We defined neighbors as those whose similarity weight relative to the focal customer is larger than 3% when we set the sum of all similarity weights to unity. This conservative definition of neighborhood was used in Equation C4.

Appendix D. Glossary of Technical Terms

The *autologistic spatial model (AL model)* is a generalized logit choice model that allows the observed purchases of other customers to impact the probability that a given customer will make a purchase. In this study, we suggest that this cross-customer dependence arises from the fact that similar customers share similar product preferences.

Calibration customers are those customers who are selected to estimate model parameters. In our empirical analysis, calibration customers were selected using a stratified sampling method based on the number of insurance policies they owned.

Customer demographics similarity is an index of customer similarity based on customers' demographic characteristics. In this study, the index is the correlation between customers' demographic variables, taken two customers at a time.

Customer preference similarity is an index of the likelihood that a given customer will buy a target product given his or her similarity to other customers. In this study, similarity is defined with respect to proximity of customer ideal points on a pick-any psychometric map based on the profile of past purchases.

The *Gini coefficient* is determined by the area between each model's cumulative lift curve and the random lift curve. Higher values of the Gini coefficient imply a better model.

Holdout customers were those customers whom we selected by random sampling to measure the accuracy of the different forecasting methods.

Joint-space mapping places customers and products on the same psychometric map. The map used in this research, known as a pick-any map, places customers near other customers with similar purchase profiles.

Lift is the percentage of all buyers in the holdout data set who fall into the group of selected customers. For each model, the probabilities that the holdout customers will purchase the target product are forecasted. Holdout customers are then ordered in terms of the probability that they will choose to purchase the product. The top x% of the holdout customers are selected, where x% is the selection rate.

A *product recommendation model* forecasts the likelihood that a customer will buy a product that he or she has not previously considered.

Scalability is the ability of a predictive modeling procedure to easily process large amounts of customer data.

The *selection rate* is the proportion of customers selected for product solicitation out of the all customers originally considered for the solicitation. It is usually determined by a profitability analysis that balances probability of purchase against the costs of contacting the customer.

A *spatial-choice model* allows observed choice outcomes to be correlated across customers. The AL model used in this study implies that customers who are near one another on the joint-space map will also have similar preferences.

Notes

1. Recently, Ansari and Mela (2003) developed a statistical model of collaborative filtering using a Bayesian semiparametric probit analysis with a Dirichlet process prior. However, the purpose of their model was to increase website traffic, not to make product recommendations.

2. When the correlation among customer response is high, the statistical efficiency of the MPLE procedure can be low (Cressie 1993). In such cases, a Markov chain Monte Carlo (MCMC) algorithm may be used to improve the efficiency of the estimation procedure (e.g., Ward and Gleditsch 2002). Because MPLE is much easier for a marketing consultant to implement than a nonstandard MCMC algorithm, we decided not to pursue the MCMC approach in this research.

3. The maximum number of dimensions of a pick-any map is R-1 where R is the number of brands. In this research, we decided to use all R-1 dimensions in computing customer distances. There exists some evidence that choosing a lower dimensional solution can increase the forecasting ability of predictive models that are based on pick-any maps (Kwak 2004). Accordingly, use of all R-1 dimensions can be viewed as a conservative test of the AL recommendation model.

4. Most NN models are statistically unidentified. Although this feature has no impact on model fit, lack of identifiability implies that estimates of model parameters are strongly dependent on the values selected to start the NN search algorithm. Even if an NN model is identified, the complex nonlinearities of an NN makes it difficult for the researcher to understand how changes in variables impact model forecasts. For these reasons, an NN is best regarded as a "black box" forecasting tool (Hastie, Tibshirani, and Friedman 2001).

5. The best possible lift curve varies as a function of the overall purchase rate among customers in the holdout sample. We have normalized the Gini coefficients in Table 3 so that a value of 1.0 corresponds to the best possible lift curve.

6. For expositional reasons, we delay the comparison the AL and NN models to a later section.

References

Ansari, Asim, Skander Essegaier, and Rajeev Kohli (2000), "Internet Recommendation Systems." *Journal of Marketing Research* 37 (August), 363–75.

_________, and Carl F. Mela (2003), "E-Customization." *Journal of Marketing Research* 40 (May), 131–45.

Ariely, Dan, John G. Lynch Jr., and Manuel Aparicio IV (2004), "Learning by Collaborative and Individual-Based Recommendation Agents." *Journal of Consumer Psychology* 14 (1-2), 81–95.

Balabanovic, Marko, and Yoav Shoham (1997), "Fab: Content-Based, Collaborative Recommendation." *Communications of the Association for Computing Machinery* 40 (3), 66–72.

Besag, Julian (1974), "Spatial Interaction and the Statistical Analysis of Lattice System." *Journal of the Royal Statistical Society, Series B (Methodological)* 36, 192–236.

Bronnenberg, Bart J., and Vijay Mahajan (2001), "Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables." *Marketing Science* 20 (Summer), 265–83.
__________, and Catarina Sismeiro (2002), "Using Multi-Market Data To Predict Brand Performance in Markets for Which No or Poor Data Exist." *Journal of Marketing Research* 39 (February), 1–17.

Cox, David R. (1972), "The Analysis of Multivariate Binary Data." *Applied Statistics (Journal of the Royal Statistical Society, Series C)* 21 (2), 113–20.

Cressie, Noel A. C. (1993), *Statistics For Spatial Data*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc.

Cui, Geng, and Man Leung Wong (2004), "Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming." Hong Kong: Lingnan University, Working Paper.

Curry, David (1992), *The New Marketing Research Systems: How to Use Strategic Database Information for Better Marketing Decisions*, New York: John Wiley and Sons.

Gershoff, Andrew D., and Patricia M. West (1998), "Using a Community of Knowledge to Build Intelligent Agents." *Marketing Letters* 9 (1), 79–91.

Gruca, Thomas S., D. Sudharshan, and K. Ravi Kumar (2002), "Sibling Brands, Multiple Objectives and Responses to Entry: The Case of the Marion Retail Coffee Market." *Journal of the Academy of Marketing Science* 30 (1), 59–69.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2001), *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York, N.Y.: Springer-Verlag.

Herlocker, Jonathan, Joseph Konstan, Al Borchers, and John Riedl (1999), "An Algorithmic Framework for Performing Collaborative Filtering." In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 230–37. New York, N.Y.: ACM Press.

Holbrook, Morris B., William L. Moore, and Russell S. Winer (1982), "Constructing Joint Spaces from Pick-Any Data: A New Tool for Customer Analysis." *Journal of Customer Research* 9 (June), 99–105.

Iacobucci, Dawn, Phipps Arabie, and Anand Bodapati (2000), "Recommendation Agents on the Internet." *Journal of Interactive Marketing* 14 (3), 2–11.

Kamakura, Wagner A., Sridhar N. Ramaswami, and Rajendra K. Srivastava (1991), "Applying Latent Trait Analysis in the Evaluation of Prospects for Cross-Selling of Financial Services." International Journal of Research in Marketing 8, 329–49.

__________, Michel Wedel, Fernando de Rosa, and Jose Afonso Mazzon (2003), "Cross-Selling through Database Marketing." *International Journal of Research in Marketing* 20, 45–65.

Kapteyn, Arie, Sara Van De Geer, Huib Van De Stadt, and Tom Wansbeek (1997), "Interdependent Preferences: An Econometric Analysis." *Journal of Applied Econometrics* 12 (6), 665–86.

Kim, Yong Seong, W. Nick Street, Gary J. Russell, and Fillipo Menczer (in press), "Customer Targeting: A Neural Network Approach Guided by Genetic Algorithms." *Management Science*.

Kwak, Kirk (2004), "Investigating the Applicability of Joint-Space Mapping to New Product Consideration and Choice Prediction." Iowa City, Iowa: Henry B. Tippie College of Business, University of Iowa, Working Paper.

Lele, Subhash (1991), "Jackknifing Linear Estimating Equations: Asymptotic Theory and Applications in Stochastic Processes." *Journal of the Royal Statistical Society B* 53 (1), 253–67.

Levine, Joel H. (1979), "Joint-Space Analysis of "Pick-Any" Data: Analysis of Choices from an Unconstrained Set of Alternatives." *Psychometrika* 44 (1), 85–92.

Melville, Prem, Raymond J. Mooney, and Ramadass Nagarajan (2002), "Content-Boosted Collaborative Filtering for Improved Recommendations." In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, 187–92. Cambridge, Mass.: AAAI Press/MIT Press.

Mooney, Raymond J., and Loriene Roy (2000), "Content-Based Book Recommending Using Learning for Text Categorization." In *Proceedings of the Fifth ACM Conference on Digital Libraries*, 195–204. New York, N.Y.: ACM Press.

Moore, William, and Russell S. Winer (1987), "A Panel-Data Based Method For Merging Joint Space And Market Response Function Estimation." *Marketing Science* 6 (Winter), 25–47.

Pazzani, Michael J. (1999), "A Framework for Collaborative, Content-Based and Demographic Filtering." *Artificial Intelligence Review* 13 (5-6), 393–408.

Pollak, Robert A. (1976), "Interdependent Preferences." American Economic Review 66 (June), 309–20.

Russell, Gary J., and Ann Petersen (2000), "Analysis of Cross Category Dependence in Market Basket Selection." *Journal of Retailing* 76 (3), 367–92.

Schafer, J. Ben, Joseph A. Konstan, and John Riedl (2001), "E-Commerce Recommendation Applications." *Data Mining and Knowledge Discovery* 5 (1-2), 115–53.

Solomon, Michael R., and Bruce Buchanan (1991), "A Role-Theoretic Approach to Product Symbolism: Mapping a Consumption Constellation." *Journal of Business Research* 22 (2), 95–109.

Soucek, Branko (1991), *Neural and Intelligent Systems Integration*. New York, N.Y.: John Wiley and Sons.

Ward, Michael D., and Kristian S. Gleditsch (2002), "Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace." *Political Analysis* 10 (3), 244–60.

Yang, Sha, and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences." *Journal of Marketing Research* 40 (August), 282–94.

Report No. 04-120
"A Spatial-Choice Model for Product Recommendations"