MANAGEMENT SCIENCE
Vol. 50, No. 8, August 2004, pp. 1120–1133
ISSN 0025-1909 | EISSN 1526-5501 | 04 | 5008 | 1120

informs

DOI 10.1287/mnsc.1040.0256

Identifying Innovators for the Cross-Selling of New Products

Wagner A. Kamakura
Fuqua School of Business, Duke University, One Towerview Road,
Durham, North Carolina 27708, kamakura@duke.edu

Bruce S. Kossar
@Futures L.L.C., 409 Executive Drive, Langhorne, Pennsylvania 18940, bruce@futuresllc.com

Michel Wedel
University of Michigan Business School, 701 Tappan Street, Ann Arbor, Michigan 48109, wedel@bus.umich.edu

With recent advances in information technology, most companies are amassing extensive customer databases. The wealth of information in these databases can be useful in identifying those customers most likely to purchase a new product and in predicting when this adoption may take place. This can assist database marketers in determining when individuals should be targeted for the promotion of a new product, which may increase the efficiency of manufacturing and distribution, and assure a faster return on investments. For this purpose, we propose a model that considers the timing of past purchases across multiple product categories and produces estimates of each customer's propensity of ever purchasing in a particular product category and of the timing of their purchases. The model is designed to help managers identify the best prospects for a new offer in one of multiple categories based on generalizations obtained from past offers. The proposed model also provides projections of aggregate penetration for new brands within the database, based on sample estimates.

Key words: database marketing; customer relationship management; hazard model; simulated maximum likelihood

History: Accepted by Christopher Tang, supply chain management; received June 2001. This paper was with the authors 19 months for 2 revisions.

Introduction

There are good reasons why the identification of early adopters is important in the introduction of a new product. First, by targeting early adopters, the firm will assure a faster return on the investments incurred in the development of the new product. Second, most firms operate under resource constraints and will benefit from a targeted approach aimed at the most responsive consumers, leading to higher sales at lower costs. Third, influencing imitators through innovators generates free word of mouth for the new product and speeds up the adoption process, maximizing turnover. Last, these factors may increase the efficiency of manufacturing and reduce initial costs of production, distribution, and inventory.

The past decade has produced an unprecedented growth in the electronic capture and recording of customer-level purchase behavior, which greatly improves the ability to forecast demand, thus reducing uncertainty in forecasting production and logistics operations. As the variety of these databases expand and they become integrated, we see a shift in focus to micromarketing approaches whereby distribution and promotion efforts can be directed at microsegments or specific individual customers. With our increased ability to track when and what customers buy comes an increased ability to target and select specific individuals for specific products and promotions at increased speed. Customer databases are especially useful in the identification of prospects for a new product or service, because past-purchasing patterns tend to outperform other geodemographic information typically available as predictors of future purchase behavior (Schmittlein and Peterson 1994). An abundance of practitioner literature (e.g., Blackburn 1990) testifies to the importance of manufacturing speed as a strategic weapon. There is a growing literature on manufacturing lead times (see Karmarkar 1987), but as Duenyas and Hopp (1994) note, very little work in that area has focused on the customer perspective. Moreover, time has become an important strategic management tool for companies well beyond just-in-time and time-to-order. By competing on introduction time, companies enjoy advantages of higher pricing, higher market share, and productivity improvement (Helms and Ettkin 2000). Based on an analysis of past purchase patterns for related products/services, the firm may be able to identify those who are more likely

to adopt the new offer early, speed up the adoption process, and reap the resulting benefits in manufacturing, distribution, and marketing planning.

Unfortunately, past-purchase data from customer databases have some limitations for the selection of prospects for targeting new products or services that limit their usefulness for the goals outlined above. Most likely, the new product or service is not identical to existing ones. The manager is then faced with the problem of estimating customers' purchase probability for a new product based on previously observed purchase patterns for related products. In other words, when identifying prospects for a new product or service, one needs to generalize customers' purchasing patterns for existing products' response to the new offer.

The main purpose of the model we propose next is to help managers form these generalizations for new products, based on an analysis of customer response to past new-product introductions combined with their own subjective estimates of key parameters for the new product. The model is used to identify customers who are likely to adopt a new product, and are expected to do it earlier than other customers. We apply the model to a physician database containing the time of adoption for several drugs introduced in the past. We first calibrate the proposed model on a sample from the database, producing maps of these previously introduced drugs, which are then used as tools to help experts provide graphical subjective estimates regarding the nature of the new drug being introduced. Application of the model to the adoption patterns from a given physician and these subjective inferences leads to estimates of the likelihood and timing of adoption for the new drug. This may support informed prelaunch decisions about which physicians should be visited first to promote a new drug. The main feature of the model is that it generalizes from past introductions (across drugs and therapies) to identify the most innovative adopters before the new drug is introduced and before data is available for the drug. The performance of the model in that respect is investigated in detail.

The identification of early adopters is particularly important in the pharmaceutical industry. The average cost for a single sales visit to a physician, during which an average of three drugs are presented (or "detailed"), exceeds $150. There are over half a million physicians prescribing cardiovascular drugs. Therefore, the costs of a single exposure to every prospective prescriber for a new cardiovascular drug could be over $20 million. Early adopters are critical in the diffusion of a new drug, because of network effects (Coleman et al. 1957) that are influenced by the way physicians are trained (i.e., the internship system) and practice the profession (within group practices, with multiple hospital privileges, etc.). Savvy salespeople are aware of the role of innovators and opinion leaders in this industry, and use them as references in their sales presentations.

A Multivariate Split-Hazard Model with Unobserved Heterogeneity

To achieve our main purpose of identifying buyers who are likely to make early purchases, we develop a model that can later produce individual-level estimates of the purchase probability and purchase timing for a new brand. Our goal is to calibrate a model on data obtained from multiple previous product introductions and to generalize this to identify customers who are likely to adopt a new product and are likely to do it earlier. Hazard models are powerful models for describing the timing of events. The basic concept in hazard models is the probability of the occurrence of an event at random time T during a certain time interval, for example, t to $t+\Delta t$, given that the event has not occurred before t, $\Pr[t \le T \le t+\Delta t \mid T \ge t]$. By letting Δt approach zero, one obtains the hazard rate

$$\lambda = \lim_{\Delta t \downarrow 0} \frac{\Pr[t \le T \le t+\Delta t \mid T \ge t]}{\Delta t}.$$

To describe our particular formulation of the hazard model, we will use the following notation.

t_i = the time elapsed (since the introduction of a new product) until customer i first purchased the product

$\lambda\{t_i\}$ = the hazard function, or probability that customer i will make the first purchase at time t_i, given that she has not done it before

$S\{t_i\} = 1 - F\{t_i\}$ = the survivor function, or probability that customer i has not made a purchase yet, and will do so after t_i, where $F(t_i)$ is the cumulative density function of the time elapsed

θ_i = the probability that customer i will ever make a purchase of the new product

$X_i = (x_{ip}) = P$-dimensional row vector of demographic and other descriptors for customer i

O_i = the set of brands that were purchased by customer i during the sampling period

$\lambda_i = (\lambda_{ij}) = J$-dimensional column vector with the random hazard rate for products j

$\alpha = (\alpha_j) = J$-dimensional column vector with the log-baseline hazard rate for product j

$\beta = (\beta_p) = P$-dimensional vector capturing the effect of demographics on the baseline hazard

$H = ((\eta_{jk})) = (J \times K)$-dimensional matrix with K factor weights representing variance of the hazard rate for product j and defining the innovation map

$N = ((\nu_{jm})) = (J \times M)$-dimensional matrix of coordinates of the location of product j in M-dimensional penetration space

$z_i = (z_{ik}) = K$-dimensional vector of scores of customer i on the K innovation dimensions

$w_i = (w_{im}) = M$-dimensional vector of scores of customer i on the M penetration dimensions

A standard hazard model would assume that all survivors at any given time become buyers sometime in the future. The split-hazard model (Schmidt and White 1989, Sinha and Chandrashekaran 1992) assumes that the probability for customer i becoming an eventual buyer is $0 \le \theta_i \le 1$. The likelihood function for customer i is

$$L_i = \theta_i \lambda\{t_i, X_i\} S\{t_i, X_i\} = \theta_i f\{t_i, X_i\} \quad \text{if a purchase is made at } t_i, \text{ or} \tag{1a}$$

$$L_i = \theta_i S\{t^*, X_i\} + 1 - \theta_i \quad \text{if a purchase is not made before the final time point } t^*. \tag{1b}$$

Equations (1a) and (1b) reflect two different contributions to the likelihood that arise depending on whether an observation is censored (1b) or not (1a). So, for a particular customer, the contribution to the likelihood equals (1a) if a purchase is observed for that customer at $t_i < t^*$, while it is (1b) if that customer has not purchased before the end of the observation period t^*. The effects of customer background variables on hazard rates are accounted for by incorporating them as exogenous variables in the hazard function $\lambda\{t_i, X_i\}$. However, a well-known problem in the estimation of hazard models such as the one in (1a) and (1b) is that additional unobserved differences in hazard rates across individuals will lead to biased estimates of the model, unless this heterogeneity is sufficiently accounted for by the model.

While representing a good starting point, the original split-hazard model is not applicable to our problem because it only provides individual-level predictions conditional on the individual's demographic background, and because it only pertains to a single product. The model we propose next accounts for unobserved individual differences in hazard rates and probability of eventual purchase, using a random-coefficients formulation. In other words, our model allows for unobserved heterogeneity, producing individual-level predictions for the probability of eventual purchase and timing of the purchase for a new product. Another important distinction between the split-hazard model and our formulation is that the former applies to multiple, possibly dependent, purchases of a single product, while ours is applied to the first purchase across a wide range of products. This requires a multivariate extension of the hazard function. This distinction is critical because our main purpose is to calibrate the model across a wide range of previously introduced brands to be used for the identification of innovative customers who might be early adopters of a new brand.

A Split-Hazard Model for Multiple Products

To attain our objectives of identifying early buyers for a new brand, our model must possess two distinct characteristics as compared to previous hazard models. First, it must allow for unobserved heterogeneity both in the hazard rates and in the probability of eventual purchase to avoid the risk of bias in case exogenous variables do not account for all the differences across customers. Second, it must simultaneously accommodate a wide range of previously introduced brands in related product categories to help managers draw generalizations from these previous experiences. These generalizations are important to obtain subjective inferences that help make predictions for a new brand without any historical information for that particular brand.

Having these two aspects in mind, we develop a multivariate random-coefficients extension of the split-hazard model. The random-coefficients formulation for both the hazard rates and the probability of eventual purchase addresses the first concern listed above by producing individual-level estimates of hazard rates. We impose a factor structure to the covariance of the random coefficients, which will prove helpful for producing inferences from the existing brands to new ones. This structure, estimated across a range of previous brand introductions, addresses the second issue discussed above. It parsimoniously describes the dependence of the adoption pattern of different products through a limited set of underlying factors. This induces a relationship between the adoption patterns of different products. That relationship can be graphically represented, as we illustrate below, and can be used by managers to form subjective inferences by generalizing to the adoption of new products.

For the estimation of our model, we either observe the time of adoption t_{ij}, or a censored observation t_j^*, if a purchase was not observed for customer i. The likelihood for customer i over all brands in the estimation sample is then

$$L_i = \prod_{j \in O_i} \theta_j f\{t_{ij}, X_i\} \prod_{j \notin O_i} [\theta_j S\{t_{ij}, X_i\} + 1 - \theta_j]. \tag{2}$$

Factor Model Formulation

We assume the parameters λ_{ij} and θ_{ij} to be randomly distributed across customers, with a certain mean and covariance structure to be estimated. We now specify the random log-hazard rate as

$$\ln(\lambda_i) = \alpha + X_i\beta + Hz_i, \tag{3}$$

where the J-dimensional vector $\alpha = (\alpha_j)$ defines a baseline hazard rate for product j, and $X_i\beta$ captures the (proportional) effect of demographics on the hazard. The term (Hz_i) is the random component of the hazard rate, where $H = (\eta_{jk})$ a $(J \times K)$ matrix. The (K-dimensional) row vector η_j can be thought of as representing K unobserved product attributes/dimensions and determines the variance of the hazard rate for product j in the population, and z_i are K-dimensional column vectors of i.i.d. standard normal distributed quantities that may be thought to represent subject-specific weights for the K unobserved attributes in question, or customers' degrees of innovativeness on those dimensions. Customers with a positive $(\eta_j z_i)$ for a particular product j will have a higher hazard rate than the average consumer and are thus likely to adopt the product earlier than average. The term $(\eta_j z_i)$ imposes a low-dimensional factor structure on the covariance of the adoption hazard across products, which describes parsimoniously the dependence of the adoption pattern across different products. The parameter matrix H defines the *Innovation* vector space, because it reflects the adoption hazard. Plots of the estimate of these parameters for the J brands in K dimensions provide a series of charts that we will call *innovation maps*. The direction defined by a brand's weights η_j in the *innovation* space identifies early adopters for the particular brand. By interpreting the directions on the *innovation* map based on previously introduced brands, managers may form inferences and provide subjective judgments that can be later used to make predictions for new brand introductions. This ability to generalize from past experience is useful when a new product cannot be clearly classified based on its physical characteristics, which is often the case for "radical innovations" that include new characteristics not present in the current market, as occurs frequently for the pharmaceutical products considered in our empirical application. In addition, in that market it is hard to identify a common set of attributes that can be measured for all existing products.[1]

We assume a multivariate distribution for the probabilities of eventual purchase (θ_{ij}), but impose a factor structure to the covariance of this multivariate distribution, so that customers and products can be mapped on a common reduced space. This map will be helpful later on, when generalizing from the experience with previous brands to new ones. We use an ideal point model[2]

$$\theta_{ij} = \frac{1}{1 + (\nu_j - w_i)'(\nu_j - w_i)}, \tag{4}$$

where ν_j is an M-dimensional vector of points that determine the covariance structure of the random coefficients across products and w_i is an M-dimensional vector of i.i.d. standardized normal subject scores. The coordinates ν_j define the location of each product j in a multidimensional space. The w_i are interpretable as subject ideal points and brands at locations ν_j that are far away from the ideal in the M-dimensional space will have low probabilities of being adopted. The set of factor weights $N = (\nu_j)$ defines the M-dimensional *penetration* idealpoint map because those weights reflect the level of penetration that a brand reaches in the population (Equation (4)). Customers positioned closer to a certain brand j on this map will be more likely to eventually adopt that drug than other customers farther from it. Conversely, brands that are likely to be adopted by the same customers will tend to cluster together in the *penetration* map. By interpreting the regions in the *penetration* map based on the location of the current brands, managers may subjectively infer the likely location of a new brand.

Model Estimation

Our random-coefficients split-hazard model now depends on $(K + M)$ unobservable random variables, which we assume to be i.i.d. standardized normals. The likelihood function for one particular customer i is

$$L_i = \iint \prod_{j \in O_i} \theta_{ij} f\{t_{ij}, X_i\} \prod_{j \notin O_i} [\theta_{ij} S\{t_{ij}, X_i\} + 1 - \theta_j] \cdot \phi^*\{z\}\phi^*\{w\}\partial z\, \partial w. \tag{5}$$

This likelihood can be difficult to compute, depending on the dimensionality of the integral. However, the model can be estimated via simulated maximum likelihood (Gourieroux and Monfort 1993) by approximating the likelihood as

$$L_i \approx \sum_r \prod_{j \in O_i} \theta_{ijr} f\{t_{ij}, X_i\} \prod_{j \notin O_i} [\theta_{ijr} S\{t_{ij}, X_i\} + 1 - \theta_{ijr}], \tag{6}$$

where $\lambda_{ijr} = e^{\alpha_j + X_i\beta + \eta_j z_{ir}}$, $\theta_{jr} = 1/(1 + (\nu_j - w_{ir})^2)$, and $\{z_{ir}, w_{ir}\}$ contains one particular draw r from the i.i.d. standardized normal distributions of these quantities

[1] In markets where such a unique set of product attributes can be identified, a "hedonic regression" on those attributes for new products may replace the subjective estimates that we use. We thank an anonymous reviewer for pointing this out.

[2] We could also specify a vector map as in Equation (3), but this vector formulation is less useful to represent adoption rates because the only way it can represent probabilities of adoption close to zero or one is through very large (positive or negative) location parameters ν_j.

for subject i. The dimensionality of η_j and ν_j is determined empirically by estimating the model with various combinations of M and K, respectively, and using information criteria such as Akaike's consistent information criterion (Bozdogan 1987). We use a bootstrap procedure (Efron and Tibshirani 1994) to obtain the standard errors of the estimates.

Computing Factor Scores

Once the parameters from the random-coefficients model have been estimated, each customer in the database may be scored along the *innovation* and *penetration* spaces. This scoring involves the same likelihood function in (6), except that the parameters of the model are known, and the optimization is done on the factor scores z_i and w_i. The factor scores for customer i, based on the estimated model and the subject's adoption behavior for previously introduced brands, are the values of z_i and w_i that maximize the likelihood:

$$L_i = \prod_{j\in O_i} \frac{\exp(\alpha_j + X_i\beta + \eta_j z_i - t_{ij} e^{\alpha_j + X_i\beta + \eta_j z_i})}{1+(\nu_j - w_i)'(\nu_j - w_i)} \cdot \prod_{j\notin O_i} \frac{\exp(-t_j^* e^{\alpha_j + X_i\beta + \eta_j z_i}) + (\nu_j - w_i)'(\nu_j - w_i)}{1+(\nu_j - w_i)'(\nu_j - w_i)}. \quad (7)$$

We use a sampling-importance-resampling (Smith and Gelfand 1992) procedure to compute these factor scores.

Identifying Early Buyers for a New Brand

Once the innovation (z_i) and penetration (w_i) scores are computed for all customers in the database, the proposed model can be used to estimate the time of adoption and probability of eventual adoption for a new brand. However, to do so the analyst must know the parameters α_j, ν_j, and η_j for the new brand j. The first parameter (α_j) determines the baseline hazard rate for the new brand. The factor weights ν_j and η_j, combined with a customer's factor scores, will predict whether the customer is likely to adopt the new brand and whether she is likely to do so early.

Because the new brand is yet to be introduced, no historical data is available for it. As noted above, when new products cannot be easily classified based on physical characteristics of products currently in the market (e.g., for radical innovations that include new characteristics), the analyst must rely on past experience to produce subjective estimates for the parameters. Fortunately, as mentioned earlier, one may form generalizations from the parameter estimates obtained from previous brands, which then can be used to produce subjective estimates for the key parameters of the model. The factor maps produced by the proposed model provide a powerful tool to graphically elicit subjective estimates of ν_j and η_j (because the main goal is to identify early adopters for the new brand, an estimate for the average hazard rate α_j for the new brand is not essential). The best prospects for the new brand, indexed by q, are customers who have a high probability of eventual adoption and a higher hazard rate,

$$\theta_{iq}\lambda_{iq} = \frac{e^{\alpha_q + X_i\beta + \eta_q z_i}}{1+(\nu_q - w_i)'(\nu_q - w_i)}. \quad (8)$$

Therefore, to decide whether customer a is a better prospect for the new brand q than customer b, it suffices to know if $\theta_{aq}\lambda_{aq} > \theta_{bq}\lambda_{bq}$.

Forecasting the Penetration of Existing Brands

The parameters of the model (ν_j, η_j, and α_j) and factor scores (z_i, w_i) are all estimated on the basis of the observed adoption history of the existing brands. Once the parameters and factor scores are estimated, they can be used to project the penetration of existing brands beyond their observed life. The penetration of brand j at any period t is given by

$$F\{t|\alpha_j, \eta_j, \nu_j, z, w\} = \sum_{i=1}^{I}\left[\frac{1-\exp(-t e^{\alpha_j + X_i\beta + \eta_j z_i})}{1+(\nu_j - w_i)'(\nu_j - w_i)}\right]. \quad (9)$$

Predictions for a new brand could be made as shown above for existing brands, except that one must first obtain subjective estimates for the new brand parameters. As discussed earlier, the innovation and penetration maps for existing brands may prove helpful in obtaining estimates for ν_j and η_j, but not in determining the average hazard rate α_j for a new brand, which is important for projecting its penetration.[3]

Identifying Early Adopters for a New Drug

As an illustration of the proposed model, we present an application to a sample from the prescription database supplied by National Data Corporation (NDC) Health Information Services. NDC is one of the primary providers of retail pharmacy prescription data. Their database contains the prescribing behavior of over one million U.S. physicians. While the data is available in a variety of forms, we were given monthly prescriptions (TRx's) for all of the new products (ethical drugs) introduced during a specific time period for a sample of physicians. The sample used

[3] One reviewer pointed out that subjective estimates of α_j can be obtained by plotting its values for existing products in the sample on a line and asking judges to place the new products among them. Alternatively, an estimate of the average hazard rate for the new brand can only be obtained after the brand introduction, when some historical data is already available.

in this application consisted of 16,166 physicians in the United States. For each of these physicians, we have data on the time (months since introduction) of the first filled prescription for 40 new drugs introduced after October 1992, as well as some background variables for the physician. We treat the time of the first filled prescription of the new drug by a physician as an indicator of the time of adoption because physicians typically try out a new drug by first giving free samples to their patients, and later adopt the drug with prescriptions to be filled at pharmacies. Thus, we have multivariate adoption data for multiple drugs. Physicians typically will differ strongly in the rate of adopting a particular drug, while not all are even likely to ultimately adopt a specific new drug. These effects of opinion leadership are well documented in this industry (Bauer and Wortzel 1966, Colemanet et al. 1957, Lilien et al. 1981). Those features of the data render it particularly well suited for the application of our model.

Of the 40 new drugs, 35 were chosen for the calibration of the proposed model, and 5 more recent introductions were held for out-of-sample predictions. The 35 brands used for model calibration have at least 16 months and at most 34 months of adoption data. The 5 newer brands used for predictive tests have a short observed history of only 7 to 9 months. These brands are listed in Table 1, along with their cumulative observed penetration among all 16,166 physicians in our sample.

Table 1 Data Description

		Cumulative observed adoptions								
Drug	Product life(mo.)	All physicians	Family practice	Internal medicine	Allergists	Cardiologists	Dermatologists	Neurologists	OB/Gyn	Psychiatrists
allergy	28	0.653	0.854	0.771	0.934	0.541	0.794	0.341	0.449	0.270
angina	34	0.443	0.683	0.661	0.116	0.698	0.070	0.188	0.104	0.072
antibio1	34	0.439	0.758	0.534	0.626	0.255	0.238	0.196	0.243	0.136
antibio2	34	0.358	0.651	0.425	0.562	0.167	0.144	0.126	0.161	0.099
antibio3	34	0.312	0.595	0.381	0.516	0.172	0.097	0.091	0.126	0.050
antibio4	34	0.596	0.822	0.714	0.744	0.494	0.550	0.261	0.605	0.159
arthritis1	31	0.493	0.752	0.642	0.240	0.307	0.147	0.497	0.224	0.137
arthritis2	22	0.313	0.543	0.407	0.111	0.148	0.053	0.197	0.072	0.045
asthma	17	0.349	0.590	0.498	0.833	0.229	0.064	0.066	0.064	0.034
choleste1	16	0.303	0.512	0.490	0.071	0.601	0.026	0.049	0.037	0.021
choleste2	34	0.465	0.717	0.699	0.160	0.794	0.090	0.155	0.123	0.099
choleste3	34	0.447	0.682	0.663	0.158	0.795	0.087	0.150	0.109	0.083
contracept1	31	0.217	0.407	0.192	0.096	0.064	0.137	0.071	0.831	0.065
contracept2	34	0.225	0.496	0.186	0.075	0.060	0.073	0.055	0.835	0.044
depression1	17	0.294	0.520	0.341	0.088	0.104	0.047	0.289	0.087	0.659
depression2	31	0.569	0.795	0.665	0.278	0.402	0.183	0.646	0.340	0.848
dermato1	34	0.172	0.254	0.183	0.138	0.043	0.710	0.041	0.055	0.026
dermato2	32	0.045	0.041	0.019	0.012	0.007	0.570	0.002	0.005	0.003
endocrin	29	0.147	0.308	0.160	0.034	0.034	0.013	0.022	0.511	0.016
epilepsy1	24	0.074	0.073	0.062	0.011	0.015	0.009	0.701	0.015	0.033
epilepsy2	18	0.073	0.064	0.055	0.016	0.019	0.009	0.709	0.008	0.034
heartburn	34	0.518	0.708	0.683	0.298	0.503	0.097	0.263	0.135	0.168
hiv	34	0.028	0.029	0.058	0.007	0.010	0.013	0.006	0.010	0.009
hypertens1	34	0.145	0.213	0.245	0.042	0.344	0.020	0.050	0.014	0.010
hypertens2	34	0.520	0.741	0.722	0.206	0.873	0.132	0.286	0.173	0.157
insomnia	28	0.547	0.718	0.668	0.252	0.520	0.155	0.533	0.312	0.680
migraine	29	0.445	0.725	0.569	0.379	0.149	0.102	0.807	0.299	0.163
ophtal1	31	0.279	0.449	0.263	0.773	0.064	0.088	0.064	0.084	0.035
ophtal2	18	0.103	0.063	0.056	0.581	0.020	0.020	0.018	0.023	0.020
ophtal3	21	0.067	0.062	0.047	0.036	0.020	0.013	0.017	0.019	0.019
ophtal4	34	0.051	0.024	0.031	0.009	0.012	0.008	0.012	0.013	0.009
pain	34	0.275	0.351	0.394	0.049	0.131	0.016	0.168	0.031	0.033
prostate	34	0.285	0.458	0.456	0.104	0.290	0.076	0.079	0.054	0.060
psycho1	34	0.031	0.021	0.016	0.011	0.009	0.007	0.095	0.005	0.295
psycho2	18	0.114	0.132	0.124	0.043	0.041	0.025	0.214	0.022	0.637
DRUG A	7	0.042	0.064	0.033	0.008	0.005	0.006	0.034	0.008	0.324
DRUG B	7	0.109	0.207	0.118	0.023	0.024	0.006	0.083	0.041	0.400
DRUG C	8	0.244	0.407	0.334	0.795	0.081	0.056	0.048	0.059	0.031
DRUG D	7	0.014	0.005	0.002	0.000	0.001	0.000	0.242	0.000	0.004
DRUG E	9	0.358	0.598	0.476	0.828	0.150	0.195	0.088	0.126	0.060

Out of the total sample of 16,166 physicians, we used one third (5,387 cases) to calibrate our model and left the remaining (10,779) to simulate the out-of-sample calculations for the database, based on sample estimates. Therefore, our predictive tests are performed out of the sample of physicians, as well as for new drugs not used for calibration. In our application, instead of including specialty as a variable in our model, we use it to post hoc validate our findings. Although in principle this variable can be included directly in the hazard formulation as shown in Equation (3), we prefer to use them here to establish face validity of our findings.

Model Calibration

The main purpose of the innovation and penetration maps produced by the proposed model is to help experts form generalizations so that they can make informed graphical subjective judgments about new brands. We test for the appropriate numbers of dimensions in the innovation and penetration maps by estimating models with two and three dimensions and comparing the Consistent Akaike Information Criterion (CAIC) statistics, which reveal that the model with a two-dimensional penetration map and a two-dimensional innovation map presents the best fit.

Insights from the Penetration Map

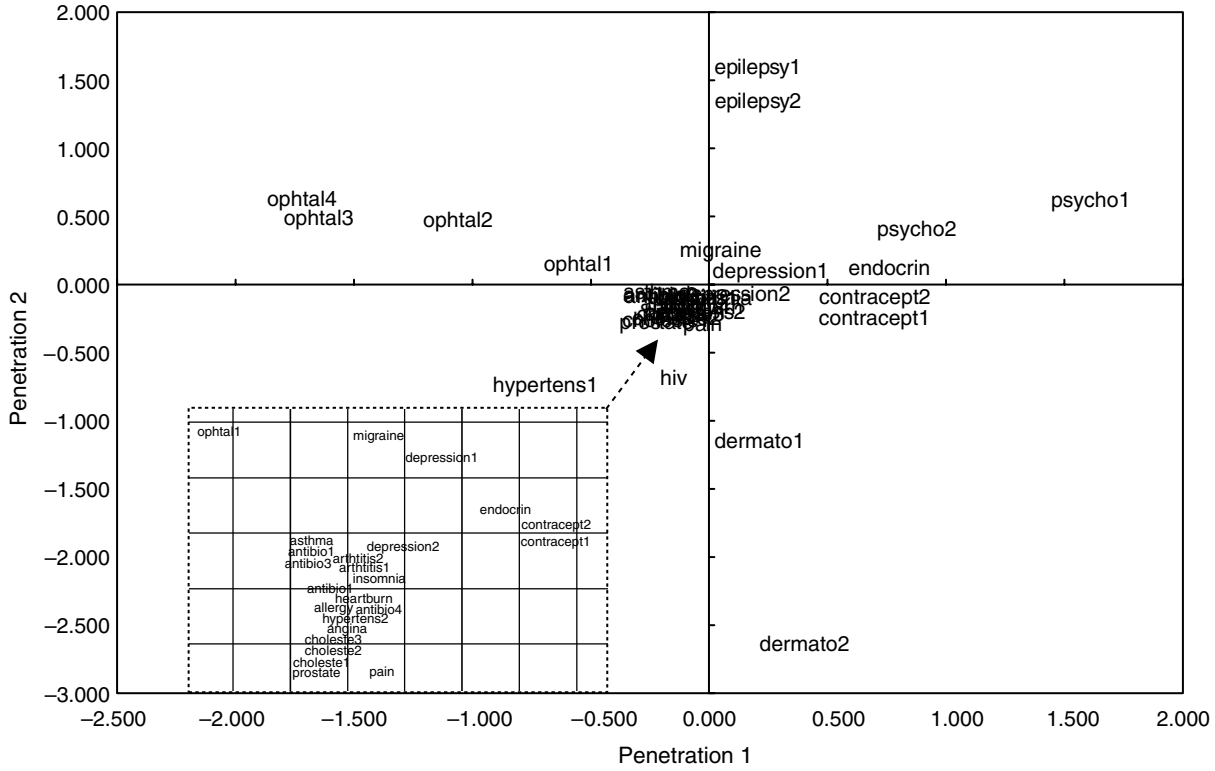
Estimates of the intercept (α_j), factor loadings for the hazard function (η_j), and the map coordinates for the probability of eventual adoption (ν_j) are listed in Table 2 for each of the 35 existing drugs considered in our study along with their standard errors estimated with the bootstrap procedure. Given the large number of parameter estimates, looking at the graphs that represent these estimates in a two-dimensional space is more informative. Figure 1 shows the penetration map (i.e., a plot of the coordinates (ν_j) for each drug

Table 2 Parameter Estimates

			Innovation Map				Penetration Map			
Drug	Intercept	SE	Factor 1	SE	Factor 2	SE	Factor 1	SE	Factor 2	SE
Pain	**−4.146**	0.002	**0.746**	0.022	**0.089**	0.025	**−0.140**	0.025	**−0.250**	0.017
Migraine	**−3.060**	0.000	**0.731**	0.022	**−0.546**	0.020	**−0.163**	0.003	**0.176**	0.005
arthritis1	**−3.139**	0.001	**0.722**	0.011	**−0.469**	0.003	**−0.244**	0.004	**−0.061**	0.005
arthritis2	**−3.719**	0.002	**0.876**	0.030	**−0.628**	0.013	**−0.287**	0.009	**−0.055**	0.006
Angina	**−3.567**	0.001	**0.712**	0.021	**−0.195**	0.020	**−0.300**	0.010	**−0.173**	0.006
hypertens1	**−5.093**	0.003	**1.687**	0.114	**−0.678**	0.056	**−0.943**	0.094	**−0.704**	0.092
hypertens2	**−3.212**	0.001	**0.651**	0.014	**−0.291**	0.008	**−0.302**	0.008	**−0.168**	0.003
depression1	**−3.560**	0.001	**1.088**	0.035	**−0.413**	0.016	**−0.015**	0.012	**0.137**	0.011
depression2	**−2.943**	0.000	**0.542**	0.011	**−0.240**	0.004	**−0.159**	0.004	**−0.034**	0.005
dermato1	**−4.371**	0.005	−0.017	0.047	**−0.853**	0.037	−0.043	0.036	**−1.116**	0.023
antibio1	**−3.463**	0.003	**0.278**	0.033	**−1.304**	0.035	**−0.319**	0.013	**−0.100**	0.007
antibio2	**−3.961**	0.003	**0.248**	0.047	**−1.325**	0.053	**−0.393**	0.016	**−0.043**	0.016
antibio3	**−4.250**	0.004	**0.263**	0.030	**−1.225**	0.033	**−0.387**	0.017	**−0.051**	0.011
antibio4	**−2.932**	0.000	**0.251**	0.011	**−0.664**	0.011	**−0.246**	0.010	**−0.134**	0.012
hiv	**−6.378**	0.004	**0.217**	0.051	**−0.117**	0.024	**−0.236**	0.051	**−0.650**	0.032
asthma	**−3.209**	0.002	0.084	0.042	**−0.980**	0.031	**−0.394**	0.023	**−0.017**	0.007
dermato2	**−4.809**	0.036	**−1.554**	0.104	**−1.642**	0.215	0.183	0.114	**−2.603**	0.054
epilepsy1	**−5.287**	0.003	**2.455**	0.056	0.180	0.102	−0.006	0.042	**1.631**	0.010
epilepsy2	**−5.015**	0.009	**1.826**	0.080	**0.331**	0.060	−0.037	0.042	**1.382**	0.012
contracept1	**−4.225**	0.002	**−0.154**	0.014	**−0.771**	0.042	**0.462**	0.019	−0.014	0.042
contracept2	**−4.346**	0.001	**−0.074**	0.015	**−1.125**	0.041	**0.434**	0.023	0.016	0.044
endocrin	**−5.018**	0.003	**0.395**	0.029	**−1.165**	0.035	**0.332**	0.016	0.026	0.035
prostate	**−4.295**	0.002	**0.945**	0.015	**−0.540**	0.005	**−0.408**	0.010	**−0.241**	0.005
heartburn	**−2.765**	0.001	**0.592**	0.010	**−0.262**	0.003	**−0.267**	0.005	**−0.127**	0.002
allergy	**−2.432**	0.001	**0.102**	0.012	**−0.652**	0.013	**−0.255**	0.007	**−0.129**	0.009
choleste1	**−3.501**	0.002	**1.146**	0.040	**−0.421**	0.007	**−0.396**	0.021	**−0.227**	0.009
choleste2	**−3.145**	0.001	**1.352**	0.027	**−0.535**	0.007	**−0.355**	0.018	**−0.209**	0.008
choleste3	**−3.381**	0.001	**1.092**	0.024	**−0.542**	0.012	**−0.337**	0.011	**−0.182**	0.003
ophtal1	**−4.173**	0.003	**−0.477**	0.045	**−1.064**	0.070	**−0.729**	0.009	**0.185**	0.022
ophtal2	**−4.786**	0.009	**−0.777**	0.102	**−1.039**	0.101	**−1.236**	0.020	**0.507**	0.051
ophtal3	**−4.809**	0.007	**−0.598**	0.076	**−0.840**	0.091	**−1.834**	0.014	**0.581**	0.037
ophtal4	**−5.388**	0.007	**−0.681**	0.047	**−0.530**	0.082	**−1.897**	0.019	**0.661**	0.034
insomnia	**−3.047**	0.000	**0.538**	0.011	**−0.217**	0.002	**−0.197**	0.003	**−0.066**	0.004
psycho1	**−6.858**	0.014	**2.154**	0.226	**0.524**	0.102	**1.382**	0.064	**0.670**	0.049
psycho2	**−4.405**	0.004	**1.149**	0.038	**0.130**	0.025	**0.680**	0.022	**0.444**	0.036

Note. Boldface type indicates a statistically significant loading.

Figure 1 Penetration Map with the Insert Showing the Center Location



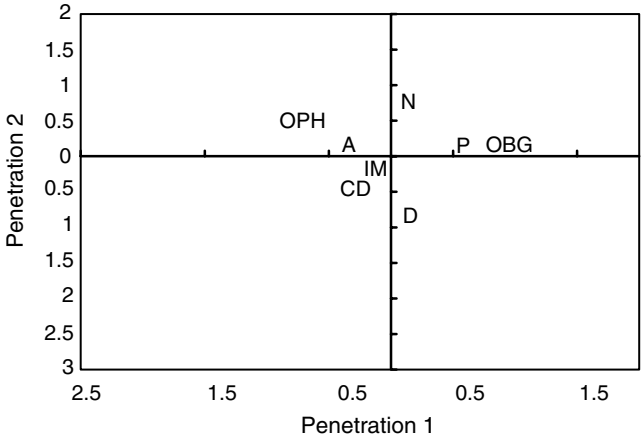
in the estimation sample). The innovation map will be useful later to elicit subjective prelaunch estimates for new drugs (Figure 4a).

To facilitate interpretation of the penetration map, we replace the brand names with the main application (therapy) for each drug. Note that in this map, drugs with similar applications tend to cluster together because they have high probability of eventually being adopted by the same physicians. The drugs clumped together near the origin in Figure 1 are likely to be adopted by the majority of physicians (i.e., nonspecialists). The insert in the bottom left quadrant of Figure 1 presents a closer view of that cluster.

The probability of eventual adoption for a new brand for a particular physician is related to the Euclidean distance between the brand and the physician in the penetration map in Figure 1. Therefore, one should expect physicians to be closer to the drugs more directly related to their specialty. We set aside physician specialty as a descriptor variable to investigate that assumption post hoc. Figure 2 shows the average locations for physicians in different specialties. Note that these specialties were not included in the model as parameters, thus differences among the specialists provide much evidence of the face validity of the estimates of the model. A comparison between Figures 1 and 2 shows that neurologists have ideal points near the neurological drugs, psychiatrists are close to the psychoactive drugs, and so on—a strong evidence of face validity for the penetration map. Note, however, that there is substantial overlap between the specialties with respect to drugs not directly related to a particular specialty. This overlap is because most physicians prescribe a broad spectrum of drugs, as can also be seen from the cumulative adoptions listed in Table 1.

Thus, the penetration map provides important managerial insights. It reveals what products/brands tend to be similar in terms of their adoption pattern. These groups of brands are located close together in the map. For example, Figure 1 reveals clusters

Figure 2 Average Location of Specialists in the Penetration Map

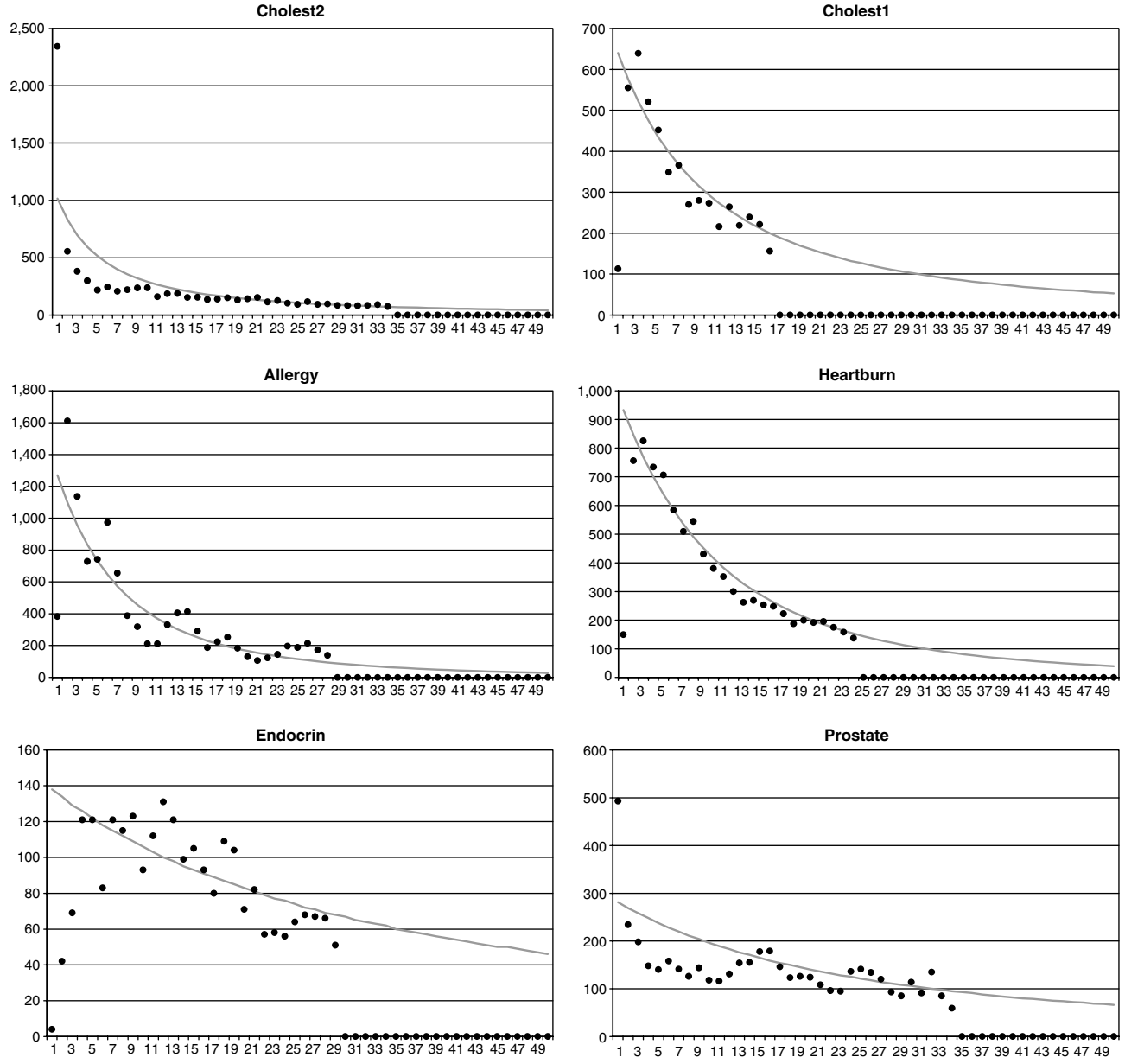


of drugs prescribed for typical medical conditions, such as neurological, dermatological, and psychological disorders. It appears that such drugs have a similar adoption pattern in terms of the physicians that adopt them. However, the boundaries or overlap between such groups of drugs may be even more interesting. These boundaries tend to indicate different types of drugs that are adopted together by the same group of physicians. Figure 1 shows, for example, that such an overlap occurs for antibiotics and asthma-related drugs, and that while some drugs for depression overlap with drugs for insomnia, others are adopted by physicians who also adopt drugs for migraine. These joint adoption patterns may provide important implications for bundling products in detailing physician visits, which would provide substantial cost reductions if done more effectively. Here it is of interest to note that our results in Figures 1 and 2 refute the conventional wisdom that the solution to detailing is to visit specialists most likely to prescribe each particular drug. Our results show that most prescriptions are written by generalists, because there are more generalists than specialists, although these generalists are harder to target accurately with new brands. This is exactly where our approach proves to be of use.

Forecasting Adoptions for Existing Drugs

As shown in Equation (8), once the parameters of the model are estimated from a sample and the fac-

Figure 3 Projected Number of Adoptions for a Sample of Brands



tor scores are computed for each customer in the database, it is possible to project the adoptions for existing drugs. The projections are made for all 16,166 customers in our selected database and compared with actual adoptions in Figure 3 for some of the drugs (which are representative of the 35 drugs used in our study). These plots show a fairly good fit to the actual adoptions in our selected database, especially when one considers that these predictions were based on parameters estimated from a subsample of 5,387 physicians and the fact that the model does not take seasonal effects into account. The estimation of the model on a small sample and further implementation to a larger database is typical in database marketing because the large databases make it impractical to calibrate a model to the population of customers. Note that the number of observed adoptions tends to decrease over time, indicating that the adoption or hazard rates are either constant or decline over time. This diffusion pattern for each brand happens when the process is dominated by innovation rather than imitation.

Identifying Early Adopters for New Drugs

The main purpose of the model is to identify early adopters for a new drug for which no historical data is available. To illustrate this feature, we combine the factor scores for each physician with graphical judgment-based estimates of the vector in the innovation map and location in the penetration map for five new drugs (A, B, C, D, and E) not included in our model calibration. We assume that these five drugs are entirely new, use the proposed model and subjective estimates to identify the physicians most likely to be early adopters of the new drug, and track the adoptions in the first nine months after product introduction to verify our predictions.

The penetration and innovation maps were presented to five consultants in the pharmaceutical industry, who were each (independently) asked to position the five new drugs based on the drugs' therapeutic characteristics and their similarities to the previously introduced ones. Their judgments, depicted in Figures 4a and 4b, provided the parameters η_q and ν_q for the $q = 1, \ldots, 5$ new drugs to be combined with the physicians' factor scores to identify early adopters for the new drugs. These figures show a fairly high level of interjudge consistency across all five experts, with the exception of the penetration for Drug A. These maps also show that Drugs A and B are judged to be similar to

Figure 4a **Subjective Inferences for the Innovation Parameters (η_j) for New Drugs (A–E)**

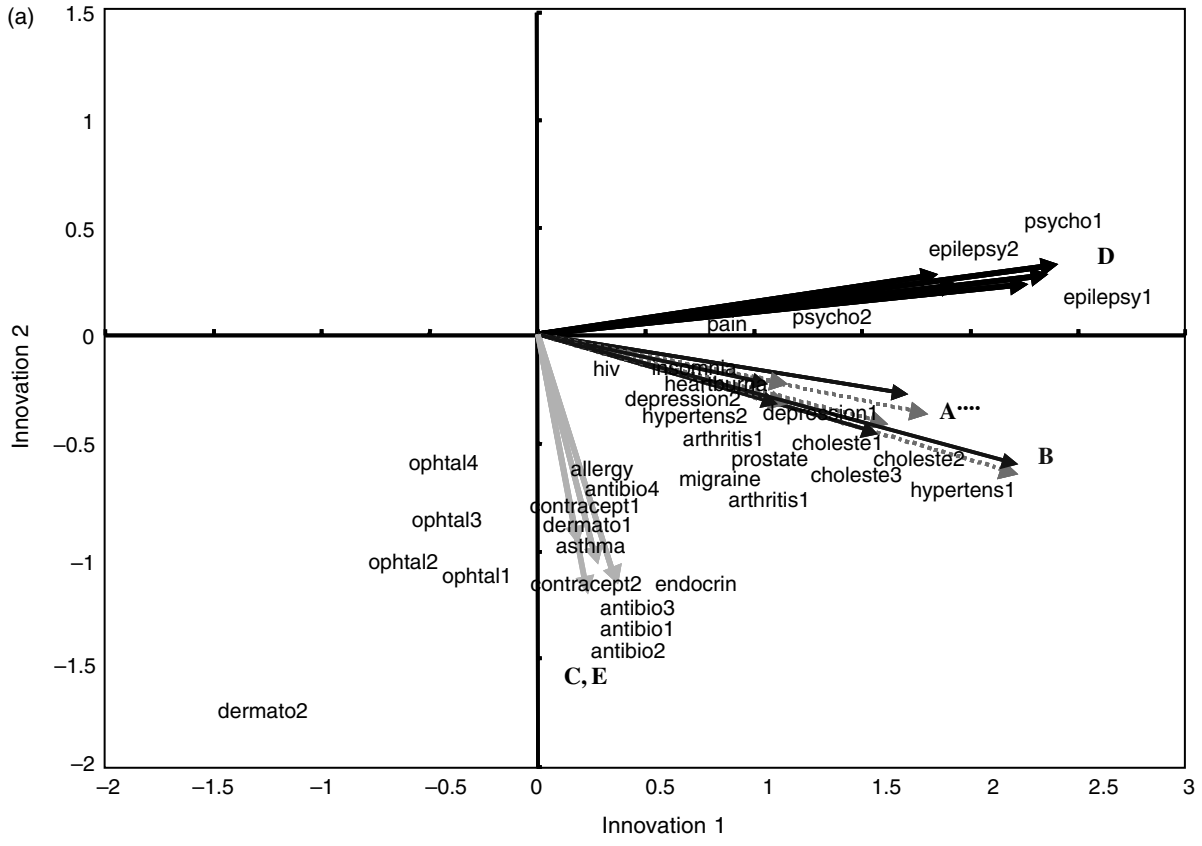
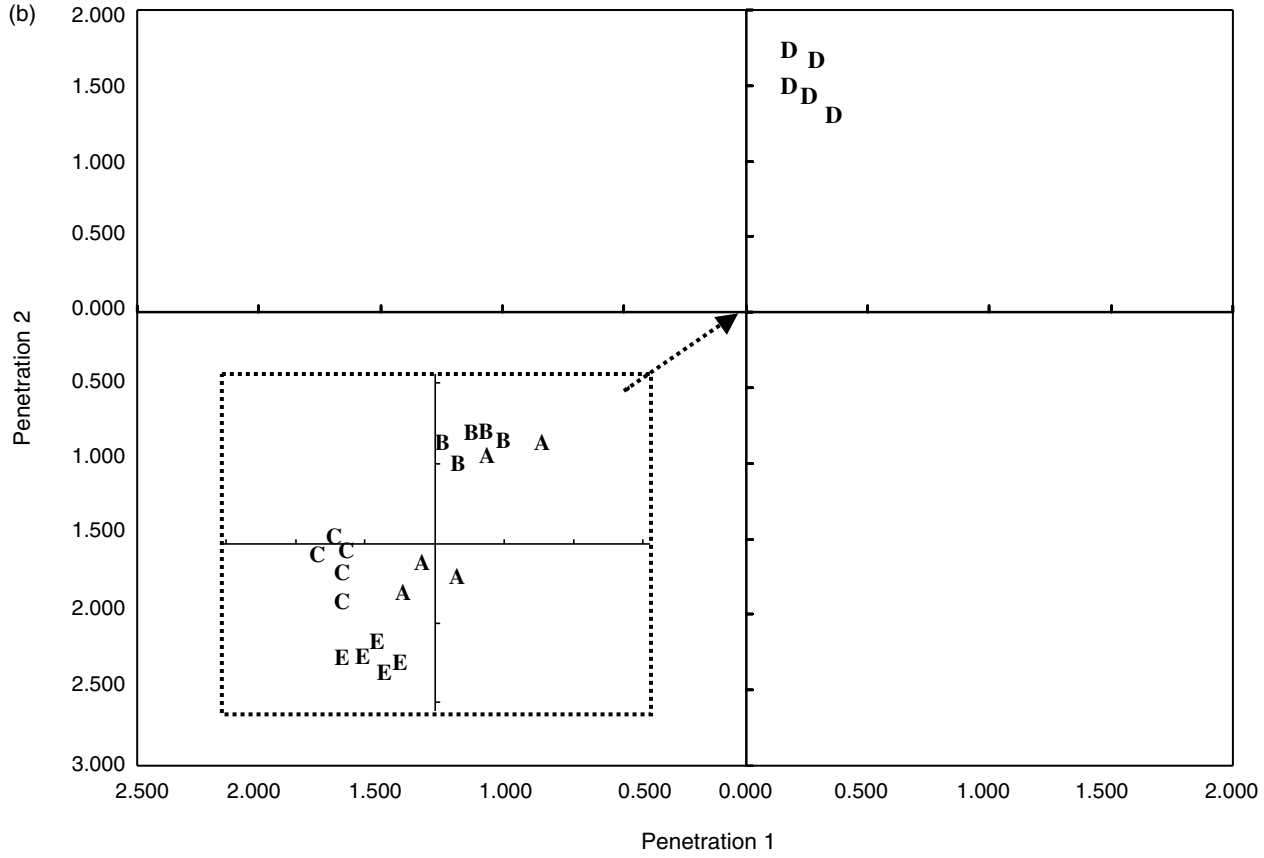


Figure 4b Subjective Inferences for the Penetration Parameters (v_j) for New Drugs (A–E)



previous antidepressants, while Drugs C and E are viewed as similar to previous antibiotics and allergy-related drugs. Drug D is clearly seen as a neurological drug.

Application of Equation (8) to every physician in our sample database produces an estimate of the expected hazard rate for that physician and drug type, relative to the average physician. This estimate may be used to identify physicians in the database who are most likely to adopt the new drug earlier than others. To have a frame of reference to evaluate the performance of the proposed model in identifying early adopters of a new drug, we compare it with a commonly used approach, based on the prescription history of the physicians in our sample. In this benchmark approach, thereby named *adopt time*, physicians are ranked based on the average time taken to prescribe the 35 drugs introduced in the past.

To demonstrate the selectivity of the proposed model in identifying early adopters for the five new drugs, we produced the "lift" curves displayed in Figure 5, plotting the cumulative number of adopters of the new product (at the last month of observed life of the product) against the cumulative number of customers in the database, sorted by the predicted probability of adoption. A random sort of the physicians would produce a 45° line as the lift curve. The lift curves in Figure 5 show that the judgments from the five experts lead to very similar results, as one would expect from the consistency of their judgments. These lift curves also show that the proposed model is consistently better and more selective than the *adopt time* benchmark.

Next, we assume that the manager is interested in identifying the top 25% physicians most likely to adopt the new drug early. Table 3 compares the profile of the target physicians in the top quartile (based on the judgments by Expert 1) for each drug with the total sample in our database, highlighting the most distinguishing characteristics in each group. Some of the differences among these groups are rather obvious. For example, the target group for Drugs C and E, which were identified by Expert 1 as similar to previously introduced allergy and asthma drugs, has a higher proportion of allergists than the total sample. Similarly, the target group for Drug D, judged to be similar to neurological drugs introduced in the past, has a higher proportion of neurologists and psychiatrists than the overall sample. However, Table 3 also shows that the target groups contain a considerable proportion of physicians in other specialties such

Figure 5 Lift Curves for the New Brands

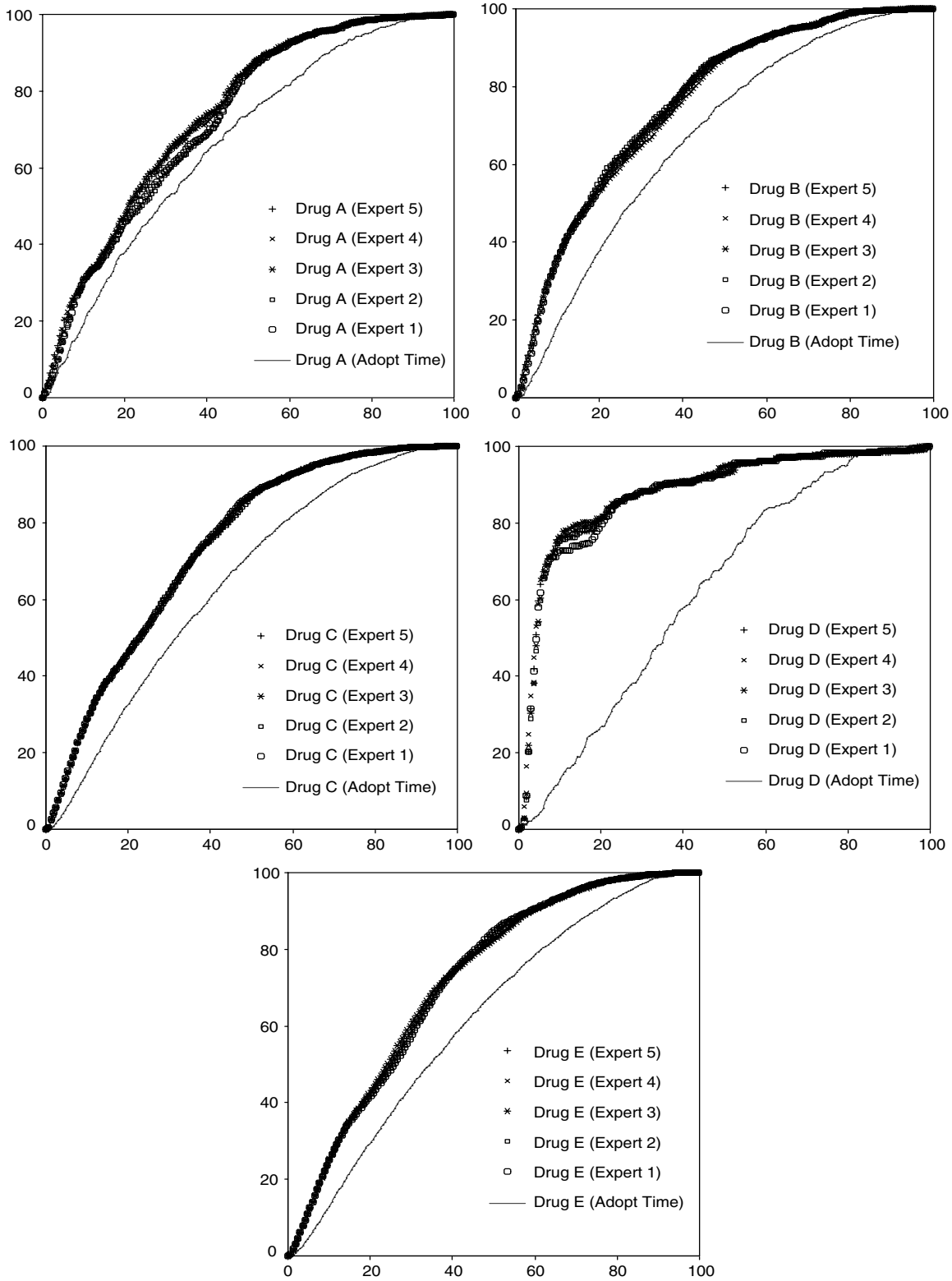


Table 3 Profile of Target Physicians (Column Percentages)

	Target physicians					Total sample
	Drug A	Drug B	Drug C	Drug D	Drug E	
Female	10.0%	10.2%	12.0%	10.7%	12.1%	15.1%
Allergist	0.5%	0.5%	**8.4**%	0.4%	**8.3**%	5.6%
Cardiologist	**7.2**%	**6.8**%	1.6%	**6.7**%	1.6%	5.2%
Dermatologist	0.1%	0.2%	2.1%	0.3%	2.6%	5.0%
Family practice	**34.7**%	**34.1**%	**38.9**%	**28.2**%	**38.8**%	19.4%
Gastroenterologist	3.7%	3.7%	2.3%	4.0%	2.3%	5.2%
General practice	6.6%	6.4%	8.0%	5.3%	8.0%	5.6%
Internal medicine	**30.7**%	**29.2**%	**28.1**%	**24.2**%	**27.6**%	23.8%
Neurologist	3.7%	5.2%	0.9%	**13.4**%	0.9%	4.9%
OB/Gyn	0.4%	0.5%	**4.1**%	0.2%	**4.5**%	5.1%
Oncologist	2.3%	2.2%	0.7%	2.7%	0.8%	5.1%
Ophtalmologist	0.2%	0.2%	1.0%	0.1%	0.9%	5.1%
Psychiatrist	3.7%	4.7%	0.9%	**8.5**%	0.9%	4.7%
Rheumatologist	6.2%	6.3%	3.0%	6.0%	2.9%	5.2%

as internal medicine, family practice, and rheumatology. These results indicate that selection by specialty alone is not sufficient for the identification of innovators for a new drug. Once again, this happens because physicians prescribe a broad range of drugs, as shown in Table 1.

Conclusions and Directions for Future Research

The model developed and illustrated here considers the nature of previous offers and facilitates generalizations based on customer response to them, so that managers can make informed decisions regarding a new offering. As with any other generalization, the innovation and penetration maps are only valid to the extent that the brands considered during calibration are representative of new brands. The 35 brands used in our illustration are only a subset of the many drugs introduced in the past five years. Better generalizations are possible with a larger sample of brands. Our application was confined to introductions of drugs. While we consider this an insightful illustration of our model, it is by no means limited to it. As customer transaction data are more widely compiled in different industries, our model can be applied in these new contexts. For example, in the online music industry, our model may be used to forecast the downloading of new tunes, based on past downloading behavior.

An important set of predictors missing in our illustration are marketing factors in effect during the introduction of the 35 brands in our calibration sample. Certain marketing activities such as detailing (sales visits to physicians) and sampling (distribution of free samples of new drugs) are believed to affect a physician's adoption behavior. Because our illustration of the model does not include these factors as exogenous variables in the hazard function and in the penetration component, the innovation and penetration scores are potentially biased by customers' response to marketing effort. In other words, some customers might have been identified as innovators, when in fact their early adoptions in the past could have been due to their responsiveness to detailing and/or sampling. The absence of marketing variables in our model is due to practical limitations. Data on the detailing and sampling of each new drug at the physician level are not widely available, and certainly not available for the whole customer database. At best, a manufacturer would have information about its own efforts, or a syndicated service could provide data from a small sample and a subset of products. Marketing variables are exogenous to each customer's behavior and, therefore, can be easily implemented as exogenous time-varying covariates in a discrete-time hazards framework whenever they are available. This would enable one to forecast the effects of marketing efforts for new brands, which appears to be an important avenue for future research.

Because the model was applied to data on the time of adoption for previously introduced brands, it can only produce estimates of the likelihood and expected time of adoption. Therefore, decisions on the optimal cut point for list segmentation are not easily implementable because such decisions would require estimates of the long-term value of each customer, which in turn depend on estimates of repeat purchases and customer loyalty. For such a purpose, the model should be extended and applied to panel data containing subsequent purchases for each brand, as multiple spells in the split-hazard model. Because the hazard rates for adoption and repeat are likely to be different for the same customer and brand, an extension of the model to dual hazard functions will be needed.

A final route of future research that is worth exploring is to embed our estimation procedure in an optimization model that determines optimal production quantities based on the profit implications on delayed introduction and the cost implications of production and inventory. This would enable optimal production and distribution decisions to be based on a model of customer behavior. However, to accomplish that our model needs to be extended to account for sales volume, where now it is tailored to describing the timing of adoptions.

References

Bauer, R. A., L. H. Wortzel. 1966. Doctor's choice: The physician and his sources of information. *J. Marketing Res.* **3**(1) 40–47.

Blackburn, J. 1990. *Time Based Competition.* Richard D. Irwin, ed. Homewood, IL.

Bozdogan, H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **52** 345–370.

Coleman, J., E. Katz, H. Menzel. 1957. The diffusion of an innovation among physicians. *Sociometry* **20** 253–270.

Duenyas, I., W. J. Hopp. 1994. Quoting customer lead times. *Management Sci.* **41**(1) 43–57.

Efron, B., R. J. Tibshirani. 1994. *An Introduction to the Bootstrap.* Chapman and Hall, London, U.K.

Gourieroux, C., A. Monfort. 1993. Simulation-based inference. *J. Econometrics* **59** 5–33.

Helms, M. M., L. P. Ettkin. 2002. Time-based competitiveness: A stretegic perspective. *Competitiveness Rev.* **10**(2) 1–14.

Karmarkar, U. S. 1987. Lot sizes, lead times and in-process inventories. *Management Sci.* **33**(3) 409–418.

Lilien, G. I., A. G. Rao, Shlomo Kalish. 1981. Bayesian estimation and control of detailing effort in a repeat purchase diffusion environment. *Management Sci.* **27**(5) 493–506.

Schmidt, P., A. D. White. 1989. Predicting criminal recidivism using 'split population' survival time models. *J. Econometrics* **40** 141–159.

Schmittlein, D., R. A. Peterson. 1994. Customer base analysis: An industrial purchase process application. *Marketing Sci.* **13** 41–67.

Sinha, R. K., M. Chandrashekaran. 1992. A split hazard model for analyzing the diffusion of innovations. *J. Marketing Res.* **29** (February) 116–127.

Smith, A. F. M., A. E. Gelfand. 1992. Bayesian statistics without tears. *Amer. Statist.* **46** 84–88.