MANAGEMENT SCIENCE
Vol. 27, No. 5, May 1981
Printed in U.S.A.

BAYESIAN ESTIMATION AND CONTROL OF DETAILING EFFORT IN A REPEAT PURCHASE DIFFUSION ENVIRONMENT*

GARY L. LILIEN,† AMBAR G. RAO‡ AND SHLOMO KALISH†

This paper develops a model and an associated estimation procedure to forecast and control the rate of sales for a new product. A repeat-purchase diffusion model is developed, incorporating the effect of marketing variables—detailing force effects in particular—as well as a word-of-mouth effect. Bayesian estimation, with priors developed from past products is used to estimate and update the parameters of the model. The procedure is used to develop marketing policies for new product introduction.
(MARKETING; NEW PRODUCT; SALES FORCE)

1. Introduction

Early in the life of a frequently purchased product, there is often too little data available either to forecast long term sales accurately or to make proper marketing decisions. A popular procedure (see Blattberg and Golanty [5] for example) is to make direct use of model parameters from other similar products.

But all products have some uniqueness: how should experience with similar products be incorporated into an estimation and control procedure? Bayesian analysis (see Raiffa and Schlaiffer [19] for example) was developed to incorporate past experience in a systematic, formal way. We incorporate bayesian estimation for the purpose of forecasting and control into a repeat-purchase model. We show that, as sales data become available, the parameters of the model and the marketing policies can be updated in a bayesian framework. This framework, incorporation past (pre-market) information with the data about the specific product, gives stable parameter estimates and policy guidelines.

In many product-marketing situations, the impact of brand promotional efforts is enhanced by a "word-of-mouth" effect—that is, by the recommendation of the brand by current satisfied users to potential users. In some situations it might be desirable to direct some of the initial marketing effort toward "opinion leaders," people who are more likely to try the new product and whose subsequent recommendations will carry more weight than the rest of the target population. Arndt [2] for example, points to the importance of the word-of-mouth effect in developing advertising policies. Silk and Davis [20] review the literature dealing with the influence processes in marketing situations, and stress the need for explicit understanding and measurement of these effects. Dodson and Muller [7] develop a general mathematical formulation for new product diffusion problems, both for durable and nondurable goods. They focus on advertising effects as well as word of mouth effects (although they do not treat issues of parameter estimation and control).

*Accepted by Donald R. Lehmann, former Departmental Editor; received December 20, 1979. This paper has been with the authors 4 months for 1 revision.
†Massachusetts Institute of Technology.
‡New York University.

0025-1909/81/2705/0493$01.25

This paper develops a model and an estimation and control procedure for marketing effort in a repeat purchase environment. For the sake of definiteness we consider the marketing of an ethical drug, aimed at a certain specialty class of doctors. One of the most important components of the marketing mix employed by pharmaceutical companies is "detailing"—i.e., personal selling by a force of detailmen, who visit doctors and describe the portfolio of products produced by their company, provide free samples and literature, and of course, attempt to combat the efforts of detailmen from competing companies. Surveys performed over a number of years have indicated that physicians generally perceive detailmen as influential sources of information (Bauer and Wortzel [3]). Other components of the marketing mix include medical journal and magazine advertising and direct mail; however these expenditures are usually linked closely to detailing effort and receive a smaller portion of the total marketing budget. The impact of company marketing effort is augmented by the effect that occurs when doctors first prescribing the product find it satisfactory and recommend it to their colleagues. A classical study in this area was performed by Coleman, Katz, and Menzel [6].

One of the problems in testing diffusion models is that direct data on word-of-mouth is hard to collect, and is usually not collected. Therefore, our model validation will be indirect in nature—i.e., we postulate a model and then, using the observed data, check to see whether the model is consistent with the data. Data for two ethical drugs are used to demonstrate the use of the model.

The heart of this analysis is a "trial and repeat" model structure. A number of re-purchase models have been developed; the most popular use panel data collected at the test-market stage of new consumer product introduction to estimate long-term rates. (Fourt and Woodlock [9], Parfitt and Collins [18], and Eskin [8]). Kalwani and Silk [12] develop some interesting insight into the nature of repeat purchase rates, formalizing some of Eskin's [8] hypotheses, All these models are descriptive in nature, though: they focus on forecasting, not on the decision of controlling the level of marketing effort; see Mahajan and Muller [15] for a review. Dodson and Muller [7] do incorporate an advertising variable into a repeat purchase model, but give no insight on how the model might be calibrated and used for decision making.

The application developed here explicitly considers only the detailing activity on behalf of, and against a new product, and the interaction of this effort with the word-of-mouth effect. The approach here differs from that developed by Montgomery, Silk and Zaragoza [17] in that we address the impact of word-of-mouth effects in the context of developing a long-term total detailing strategy. Montgomery et al. develop a more detailed, tactical procedure that is more heavily dependent upon managerial judgment for calibration, i.e., a decision-calculus approach (Little [13]).

In its application, the model is used to develop "good" detailing policies. We call them "good" rather than "optimal," because they have been specified to be profit improving as well as easily implementable in the total detailing context rather than just profit maximizing. Management has to allocate detailmen's time across a variety of products; therefore a policy for a *single* product must be simple enough to be incorporated within the total portfolio. This, we believe, precludes policies that are highly state and time dependent, requiring frequent changes in effort allocation.

Managerial use of the model presents some interesting problems. Since the key period in the marketing planning horizon occurs at the beginning, when there is no marketing data on the product, even purely adaptive estimation of parameter values cannot be advocated as a model calibration strategy. Our approach is to model a variety of products, obtaining the model parameters for each, and to use information about those parameters to develop a prior distribution of parameter estimates for the new product. These estimates are used to develop initial policy decisions, which are updated as sales data become available.

2. The Model

Consider the case of ethical drug adoption where there are N^* doctors in the prescribing class (psychiatrists for anti-depressants, e.g.) of which $N(\leqslant N^*)$ may eventually prescribe the drug. We do not observe the number of prescribing physicians; rather we observe prescriptions filled (or sales). In general these two quantities will be related closely. The most productive doctors, who write a disproportionate number of prescriptions, are also, however, more likely to be early adopters. The difference in mean prescription rate per doctor is unlikely to be large during the introduction phase of the drug (the period we observe here) though, and, hence, for simplicity and to keep the number of parameters tractable, our model derivation assumes a linear relationship, constant in time, between doctors and prescriptions.

In addition, note that we may have two, convergent phenomena here: (a) early prescribing doctors prescribing more; (b) detailing effectiveness decaying over time. Both of these phenomena are linked to decreasing returns to detailing spending over time, a phenomenon we model below. The separation and estimation of these effects are important if we wish to make inferences about the "true" value of N, as is careful analysis of the time-decay in number of prescriptions per prescribing physician. However, here, as with most diffusion models, optimal policies are relatively *in*sensitive to fairly major changes in adoption potential (see Horsky and Simon [10]). Thus as our objectives are to infer the time path of product sales early in the life cycle and develop efficient policies, a simple model of detailing effectiveness as operationalized below is sufficient.

Figure 1 describes a flow model of the process, comprising four states: (1) never having prescribed, (2) trier (prescribing once), (3) repeating, and (4) past prescriber. The activities affecting the various flows are labelled. Here we have a trial structure (moves from state 1 to state 2) and a repeat structure, comprising states 2, 3, and 4. Early in the life of the drug, the flow will be mainly from states 1 to 2 (and perhaps on to state 3, for limited repeat). Later on switches from 2 to 4 and between 3 and 4 will dominate the flows.

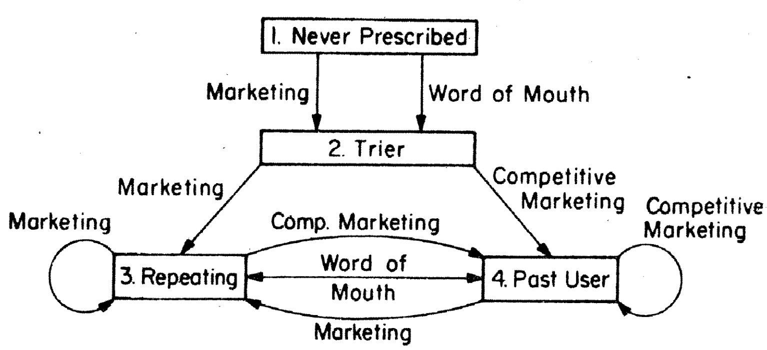


FIGURE 1. General Flow Model Describing the Process.

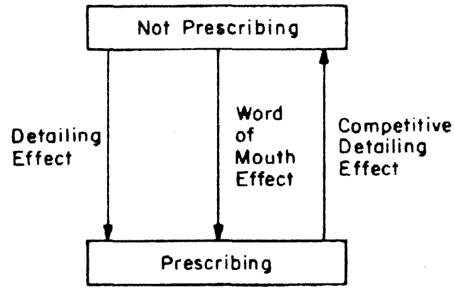


FIGURE 2. Simplified Flow Model Describing the Process.

The four state model is quite complete, but has too many parameters for efficient estimation for the data sets we have examined. In particular, Figure 1 requires that panel-type data be collected to separate out the different flow rates. We therefore use a two-state model—prescribing vs. not prescribing—as an approximation. The difference in the detailing effectiveness between the "trial" and "repeat" portion of the four state model will be handled in parameter estimation by an effectiveness decay factor, $f(t)$, applied to the coefficient of detailing for the new drug.

Figure 2 outlines the two-state model. Here we consider two classes of doctors, prescribing and not prescribing. The flow between these classes is controlled by company and competitive detailing and product experience, positive or negative.

Formally, the model becomes:

$$C_2(t+1) = C_2(t) + \lambda_1(d(t))C_1(t) + \lambda_2(\Delta C_2(t))C_1(t) - \lambda_3(\bar{d}(t))C_2(t) \tag{1a}$$

and

$$C_1(t) + C_2(t) = N \quad \text{for all } t. \tag{1b}$$

where

$C_1(t)$ = number of doctors at t not prescribing, $t = 1, 2, \ldots$

$C_2(t)$ = number of doctors prescribing the drug at t.

$K_1(t)$ = number of new (i.e. initial) prescriptions for the drug observed at t.

$K_2(t)$ = number of prescription renewals for the drug at t.

W = random variable, the number of patients actually using the drug class that a randomly chosen doctor has.

(Recall that, although the model is structured in terms of prescribing doctors, we observe the number of prescriptions. We estimate C_2 as follows:)

$K_1(t) + K_2(t) = \rho C_2(t) E(W)$, where ρ is the average percentage of a prescribing doctor's patients who receive the drug, or

$C_2(t) = (K_1(t) + K_2(t))/E(W)\rho$

$\bar{d}(t)$ = competitive detailing level at t.

$d(t)$ = level of detailing at t.

$\lambda_1(d(t))$ = company detailing effectiveness.

$\lambda_2(\Delta C_2(t))$ = word of mouth impact, where $\Delta C_2(t) = C_2(t) - C_2(t-1)$.

$\lambda_3(\bar{d}(t))$ = competitive detailing effect.

$f(t)$ = decay factor for detailing effect.

The decay factor, $f(t)$, models two phenomena: early prescribing doctors prescribing more and decaying of detailing effectiveness. These are simultaneous correlates of a decline in $\lambda_1(t)$ over time; our data will not permit a separate look at these phenomena so they are combined in one decay function. Normally, we expect $f(0) = 1$ and $f(t)$ to be non-increasing. For simplicity of exposition we will exclude $f(t)$ from the discussion; operationally, $f(t)$ multiplies the true value of $d(t)$ to get "effective" detailing at that point in time. This term is analogous to a discount factor in financial calculations and is also similar to Little's [14] copy-effectiveness factor.

The form and estimation of the time-varying factor, $f(t)$ can be handled in several ways. See [1] for a review of time-varying parameter structures, and Wildt and Winer [22] for some approaches and problems in a marketing framework. One seemingly appropriate method is to assume a linear relationship between $f(t)$ and $f(t-1)$, possibly including random noise, and to estimate that relationship using Kalman filter methods. We did not choose this tack as we wish to keep the number of parameters small and to estimate them early in the product introduction process, before a significant decline in $f(t)$ could be expected to be observable. We thus keep the problem simple by judgmentally calibrating $f(t)$; management usually is willing to provide educated guesses about the rate of decline in $f(t)$.

As structured, this model has several important simplifying assumptions. The first is that N, the number of doctors in the class, is assumed fixed. Kalish [11] suggests how this assumption can be relaxed.

The second assumption is that all doctors are in the same class (psychiatrists versus general practitioners, for example). It is not difficult to amend the model to eliminate this assumption by constructing a series of parallel processes, such as that in Figure 2, for each class of doctors.

A third assumption is that detailing effectiveness is not related to the current number of prescribing doctors. This could be handled in the model through an interaction term between simple detailing effectiveness and the word-of-mouth effect.

These modifications are beyond the scope of our current objectives however and data needed to attempt such extensions are not available.

3. Estimation and Validation

In the previous section, we proposed a model structure for the detailing decision. Now we must answer two questions:

(a) Is the model good, i.e. does it perform better as a forecaster than alternative, naive models?

(b) How does one use the model in the typical new product situation when either no or very little data are available for the product?

The parameter estimation issues involved in (a) and (b) are different because in validating the model we can use a substantial amount of historical data on a product. In this section we focus on (a). We propose functional forms for the response $\lambda_i(\,)$ and show how the parameters of these forms are estimated using part of the data for a particular product. The model is then used to forecast sales of the product, and the forecasts are compared to actual sales achieved. Two naive models—one a polynomial in time and one an autoregressive scheme—are also estimated and used for forecasting. These forecasts are also compared to actual sales, and the resulting root mean square errors are used to test the validity of the proposed model. The issues raised in (b) are discussed in the next section.

Consider now the specification of functional forms for our response models. Although linear response functions are tempting to use from the estimation viewpoint they are clearly unsatisfactory for policy development purposes since they always imply that marketing efforts should be either zero or as large as possible. The possibility of non-linearity of response to detailing effort is therefore essential.

Consider the following form. It is a simple form that contains nonlinearity. Let

$$\lambda_1(t) = a_1 f(t)d(t) + a_2 f(t)d^2(t), \tag{2a}$$

$$\lambda_2(\Delta C_2(t)) = a_4(C_2(t) - C_2(t-1)), \tag{2b}$$

$$\lambda_3(\bar{d}(t)) = a_3\bar{d}(t). \tag{2c}$$

Substituting $C_1(t) = N - C_2(t)$, and the above in (1a) we get

$$C_2(t+1) - C_2(t) = \left(a_1 f(t)d(t) + a_2 f(t)d^2(t)\right)(N - C_2(t))$$
$$- a_3\bar{d}(t)C_2(t) + a_4(C_2(t) - C_2(t-1))(N - C_2(t)). \tag{3}$$

Note that this model-structure handles market feedback in a different way than the interactions in most diffusion models. This formulation replaces the $(N - X)X = NX - X^2$ interaction term used in Bass [4] and other formulations by a time-based difference $(X(t) - X(t-1))(N - X(t-1))$. Thus, positive sales increases are a positive accelerator. A decline in sales (bad product experience, e.g.) leads to negative feedback.

In general, the positive and negative effects will not be symmetric (see Midgley [16] for empirical evidence). A slight modification of this formulation incorporating indicator variables will allow separate estimation of a_4^+ and a_4^- (positive and negative feedback coefficients, respectively). The data sets we have examined have only incorporated positive feedback so the symmetric form in 2b is retained here.

We refer to model (3) as the "full model." This equation contains five unknown parameters: a_1, a_2, a_3, a_4 and N (assuming $f(t)$ known), with N appearing in a way that makes it impossible to use conventional linear estimation procedures. However, if N is known, then (3) becomes linear in its parameters.

There are at least two approaches to estimating these parameters. Note that, although N is critical for forecasting, it is not involved in determining the detailing policy in the introductory period (see Section 5). Thus we consider two options:

(a) Direct estimation of all parameters using non-linear estimation methods and

(b) Two-stage estimation—estimating N from a simpler model and then estimating the other four parameters, given knowledge of N.

In efforts to implement strategy (a), we were unable to obtain satisfactory results due to multicollinearity in d, d^2 and $\bar{d}$, present in all the data sets we have examined. Thus we explore strategy (b) as an estimation approach. Specifically, the approach we follow to estimate N is as follows: *Assume:*

A1. Detailing response can be approximated locally by a linear function.

A2. Company and competitive detailing effectiveness are approximately equal.

A3. The population that the word-of-mouth term affects early in the life of the drug can be approximated by N (i.e., $C_2(t)$ is small relative to N).

Under these assumptions we get the following model from equations 2 and 3:

$$a_1 = a_3, \quad \text{and}$$

$$C_2(t+1) - C_2(t) = a_1 d(t)(N - C_2(t)) - a_1 \bar{d}(t) C_2(t) + a_4 \Delta C_2(t) N$$

or

$$\begin{aligned} C_2(t+1) - C_2(t) = {} & A\, d(t) && \text{where} \quad A = a_1 N \\ & - B\big(d(t) + \bar{d}(t)\big) C_2(t) && B = a_1 = a_3 \\ & + C\,\Delta C_2(t) && C = a_4 N. \end{aligned} \tag{4}$$

We will call this model the "intermediate model." We estimate the parameters A, B, and C using ordinary least squares and estimate N from the fact that $N = A/B$. This value (A/B) gives the maximum likelihood estimate of N. We need the distribution of N as well, in the next two sections, (a) to construct a prediction interval around expected sales and (b) in deriving a detailing policy to maximize the expected number of prescribers, where that expectation is over the distribution of N. As the estimates of A and B ($\hat{A}$ and $\hat{B}$ respectively) are approximately bivariate normal, the distribution of N can be developed analytically, although not in closed form. Rather than perform the numerical integrations needed to develop this distribution, it is simpler to obtain the distribution by simulation as follows: If X and Y are independent identically distributed (0, 1) normal random variables then it is easy to show that

$$\hat{A} = \sigma_1 X + \mu_1 \quad \text{and}$$

$$\hat{B} = \frac{\rho}{\sigma_1} X + Y \sqrt{\sigma_2^2 - \left(\frac{\rho}{\sigma_1}\right)^2} + \mu_2$$

are distributed as bivariate normal with mean (μ_1, μ_2) and covariance matrix

$$\begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}$$

The maximum likelihood estimate of N is the mode of the simulated frequency distribution of A/B.

Table 1 gives the key pieces of data for two cases of ethical drugs introduced into two different markets. The data, obtained through a cooperating firm from IMS America, has been disguised by multiplication by an arbitrary constant, to protect company confidentiality. Case 1 is used to validate the model structure and Case 2 to illustrate model use. The parameter estimates for the intermediate model are shown in Table 2.

Figure 3 shows forecasts using the full model, assuming $N = 10{,}700$, the maximum likelihood estimate (standard deviation = 2280), using the first 12 points for fitting.

TABLE 1

Analysis Data

	Case 1			Case 2		
Quarter	Detailing	Competitive Detailing	Sales*	Detailing	Competitive Detailing	Sales
1	53	308	0	84	725	0
2	47	417	65	69	846	35
3	55	383	133	91	834	71
4	57	396	213	58	1023	109
5	53	411	280	63	953	138
6	46	417	311	67	837	153
7	56	462	410	63	924	160
8	61	467	578	84	953	170
9	44	498	710	84	736	192
10	53	488	775	81	992	223
11	51	523	782	72	776	260
12	49	581	783	81	662	307
13	44	611	785	67	822	348
14	43	581	786	47	1024	384
15	40	585	793	45	989	414
16	38	493	796	47	777	438
17	41	505	775	72	992	464
18	35	516	730	65	756	492
19	32	485	667	79	852	506
20	27	444	604	66	1103	506
21	28	463	557	80	946	506
22	26	427	537			
23	24	466	536			
24	22	472	536			

*Sales $\equiv (K_1(t) + K_2(t))/E(W)p$; or Sales $= C_2(t)$, an *estimate* of the number of prescribing doctors. Here sales data have been transformed to represent the approximate number of prescribing doctors, C_2.

TABLE 2

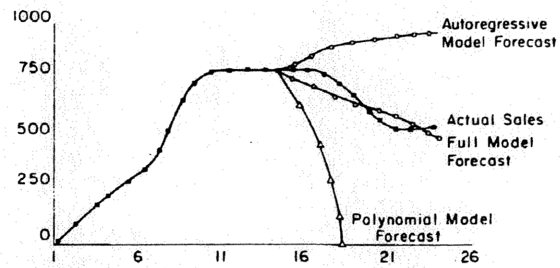
Parameter Estimates: Intermediate Model

Coefficient	Value	t-Stat
A	0.800	2.07
B	7.49×10^{-5}	2.21
C	0.626	4.03

$F(2; 19) = 522$
Corrected R-Square = 0.99

Here the function $f(t)$ was modeled as $f(t) = 1$, $t < 12$; $= 0.6$, $t \geqslant 12$. This form, consistent with historical decay patterns in this market, works satisfactorily for case 1. Several alternatives were tried (exponential decay, varying times for shift, varying levels for shift) and this one worked adequately both in terms of fit and prediction. Operationally, more historical analysis will lead to greater confidence in an appropriate form for $f(t)$. In section 5 we review why, for policy development (our main focus here) the precise form of $f(t)$ is not of major concern.

Figure 3 compares the forecasts obtained using the nonlinear model, and other models with actual sales. We show parameter estimates for a third order polynomial that was fit to the data using the first 14 points, as well as a third order autoregressive



FULL MODEL

Coefficient	Value	t-stat	
a_1	3.17×10^{-5}	0.13	F(3;10) = 197.4
a_2	7.85×10^{-7}	0.17	Corrected R-Square = .98
a_3	6.05×10^{-5}	1.10	RMSE% = 5.9%
a_4	5.48×10^{-5}	2.21	

AUTOREGRESSIVE MODEL: X(t) = A+BX(t-1) + CX(t-2) + CX(t-3)

Coefficient	Value	t-stat	
A	64.90	2.37	F(3;10) = 206.8
B	1.99	7.47	Corrected R-Square = .98
C	-1.65	-3.55	
D	0.60	2.37	RMSE% = 39.8%

POLYNOMIAL MODEL: $X(t) + A + BT + CT^2 + DT^3$

Coefficient	Value	t-stat	
A	61.1	0.32	F(3;10) = 112
B	-23.9	-0.31	Corrected R-Square = .96
C	17.7	1.94	
D	-0.86	-2.59	RMSE% = not relevant

FIGURE 3. Model Forecasting Comparison: Case 1.

scheme, together with the resulting forecasts. In each of these cases, the order of the model was selected as having the same number of parameters as the full model.

The forecasts from the polynomial model are obviously unsatisfactory; they become negative. The autoregressive model does better, but the RMS error in this case is 7 times greater than the RMS error for the full model.

Based on the results from this data set, and its comparison with other models, we have some confidence in the full model.

4. Using the Model

Now we turn to the question, how does one use the model in a typical new product situation when limited data are available? We suggest that a bayesian procedure, using a prior for the parameters, based on a combination of managerial judgment and historial experience with similar products, is an appropriate way to incorporate past information. Again, we have two alternatives—we can use a complete, nonlinear estimation procedure or we can adapt the two-stage procedure employed in the last section. Our experience with the two-stage procedure has been more satisfactory and it has the added advantage of permitting bayesian regression with a natural conjugate. (See Raiffa and Schlaiffer [19]).

Our procedure is similar to that developed in the previous section—i.e. using an intermediate model to obtain an estimate of N, and then estimating the parameters of the full model. The differences are that

(i) a smaller number of data points (4 to 8) are employed in the estimation,

(ii) priors for the parameters, A, B, C, and $a_1, \ldots, a_4$, derived from other "similar" products and modified, if necessary, to reflect unique characteristics of the product class, are used together with these data points in a bayesian procedure,

(iii) the parameters are updated as more data become available.

This procedure assumes that knowledge of other drugs in the class is available, either quantitatively or intuitively.

Our procedure has four steps:

Step 1: Estimate parameters of the intermediate model, using ordinary least squares or bayesian regression (if past data are available).

Step 2: Derive the distribution of N from the assumption that (A, B) are bivariate normal.

Step 3: Sample several values of N; $N_1, \ldots, N_k$, from the distribution derived in Step 2. Incorporating prior estimates of $a_1, \ldots, a_4$ from previous data, for each N_i, develop posterior estimates of $a_1, \ldots, a_4$.

Step 4: Develop a detailing policy to maximize expected long-term per period profit for each value of N_i and optimize across the distribution of N_i.

To summarize and motivate this procedure, we use bayesian methods to estimate N. But, as the uncertainty in N can affect forecasts as well as detailing policy, information about the entire distribution of N is used in estimating the full model; i.e., the parameters of the full model are re-estimated for each of a sample of values from the probability distribution of N. A detailing policy is then developed for each sample value of N, and an optimal policy, minimizing the loss due to uncertainty in N is derived. We illustrate this procedure in this and the next section, using case 2 data from Table 1.

In developing a prior for case 2, parameters from case 1 on the entire data stream were used, adjusted, as follows, to reflect the slower diffusion rate management expected for drugs in this (second) class. In particular, $-a_1/2a_2$ (the optimal short-term detailing level as derived in the next section) was set initially to 90, consistent with historical detailing levels in this class, and the variance-covariance matrix was used from case 1, with the first two diagonal elements strengthened to be consistent with these detailing level assumptions. Note that this assumes that the two drugs have similar market characteristics; greater experience with historical cases should lead to more realistic priors. The updated (posterior) coefficients for the intermediate model, using priors from case 1 and the first eight data points are given in Table 3.

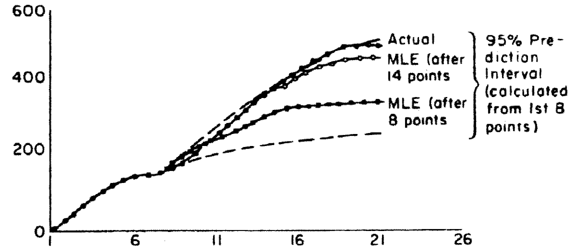
The density of N, again obtained by simulation, has a mean of 5400 and a standard deviation of 1150. Figure 4 shows the bayesian full model forecasts based on the first eight data points, giving the forecasts obtained using the maximum likelihood estimate of N, together with a 95% prediction interval. That figure also shows the improved fit and prediction as 6 more points are added. Note that this bayesian procedure allows us to use the full model, even with a small number of data points available. The figure suggests that the bayesian procedure is useful as a forecasting procedure, with a RMSE of only 6%.

For the purpose of comparison, OLS estimates from the first 8 data points were also developed and predictions were made. Forecasts were made similarly for the intermediate model. Although all models fit the first 8 points equally well, differences in predictive accuracy were considerable: 23% RMSE for the intermediate model, 45%

TABLE 3

Posterior Coefficients, Intermediate Model (Case 2)

	Updated Coefficients (mean)	t-values
A	0.171	0.77
B	3.17×10^{-5}	0.90
C	0.805	4.52



Posterior Etimates of the Coefficients

Variable	After 8 points Value	After 8 points t-stat	After 14 points Value	After 14 points t-stat
a_1	1.01×10^{-4}	6.13	1.01×10^{-4}	8.4
a_2	5.08×10^{-7}	-1.30	-5.13×10^{-7}	-1.38
a_3	3.40×10^{-5}	0.88	3.44×10^{-5}	0.92
a_4	8.41×10^{-5}	1.93	8.47×10^{-5}	2.68

FIGURE 4. Case 2: Bayesian Estimates, Using First 8 and then 14 Points Plus Case 1 Data as Prior. Prediction (and Prediction Interval) for Rest of Series.

for the OLS forecasts compared to 6% for the full model. These differences were largely due to the fact that the full model picked up the second up-turn in the data series much better than either of the other models, pointing to the value of this procedure.

5. Determination and Updating of Detailing Policies

In principle, the profit maximizing policy over a planning horizon T periods long can be obtained by solving a dynamic programming problem with one state variable, C_2 (see equations 1a–1c). Computation of such a policy requires some assumptions about competitive detailing activity during the planning period, but these assumptions can probably be made, and the sensitivity of the policy to these assumptions examined.

We believe, however, that this approach will lead to policies that are complicated to implement and also unrealistic. For competitive reasons it is usually desirable to drive the market share of the new product up as quickly as possible, and then to maintain it at that level. As will be shown below, this would imply a pulse of detailing activity during the introductory phase of the detailing campaign, followed by a (perhaps) reduced "maintenance" level of detailing during the life of the product. More highly time dependent policies, calling for a different amount of effort on each drug in *each* period are difficult to implement or control; these are the types of policies that are likely to be yielded by a dynamic programming, profit maximization formulation.

In view of the above we develop the parameters of a policy of the following type: "Drive the market share of the product up to some level m, and then maintain it at this level."

With this type of policy, the introductory phase goal is to reach a desired share m as quickly as possible. We can operationalize this by computing policies that maximize m at the end of t periods, where t can be a variable to be selected to provide the desired m.

Setting $t = 1$, it is easy to show from equation 3 that the optimal detailing level $d_1^* = -a_1/2a_2$. Because the objective function as now set up is separable between periods, we can show that $d_j^* = -a_1/2a_2$, $j = 1, 2, \ldots, t$, maximizes m_t, the market share at the end of t periods. Thus, during the introductory phase, the detailing level should be maintained at $-a_1/2a_2$ until the desired or target share is achieved. In order to compute the value of m_t, assumptions must be made regarding competitive detailing levels.

In the long run, a reasonable objective is to maximize steady state per period profit. For a fixed N, the per-period number of prescribing doctors is:

$$C_2 = \frac{N\left(a_1 f(\infty)d + a_2 f(\infty)d^2\right)}{a_3\bar{d} + a_1 f(\infty)d + a_2 f(\infty)d^2}$$

(assuming $\bar{d}$ = constant and $f(\infty)$ = long-term decay value for λ_1).

A simple, linear model of annual profit combines margin per prescribing doctor (b_0) and the incremental cost of detailing (b_1) as follows. Per period profit is, then:

$$\Pi_s(N) = b_0 C_2 - b_1 d.$$

We may wish to choose a policy d that maximizes expected profit, as follows:

Find d to max $\int_n \Pi_s(n) f_N(n)\,dn$ where $f_N(n)$ refers to the distribution of the number of potential prescribers, calculated from the procedure described in section 4.

For our case, the company determined its average margin per prescribing doctor per 8 week period as \$66 (b_0) and its average cost of a detailman's visit as \$95 (b_1). (This latter figure is consistent with recent McGraw-Hill estimates). Using these figures, the optimal detailing effort was calculated as 52 visits per 8 week period.

As indicated earlier, a short-run policy is to drive the share up as fast as possible. In our case, this is done by setting $d = -a_1/2a_2$, where we use the posterior estimates of a_1 and a_2. For the value of N associated with the optimal long-term policy, this level of effort is 93.

If we assume that $\bar{d}$ is approximately constant at 900 and $d = 52$, then we get that the steady-state share, assuming $f(\infty) = 0.6$, is

$$C_2/N = \frac{a_1 f(\infty)d + a_2 f(\infty)d^2}{a_1 f(\infty)d + a_2 f(\infty)d^2 + a_3\bar{d}} \simeq 0.067.$$

By our reasoning, then, the suggested policy is to set a detailing level at 93 until a share of about 7% is reached and then back down to around 52.

One of the powers of the bayesian approach is that updating of policies is natural as more data are collected. In a manner identical to that above, the updated, optimal long-term policy after 6 more points are available, was calculated as 71. (In practice, updating would occur each time new data were received from the field.)

The impact of the form of the $f(t)$ function could be of concern here. Note, however,

(a) the short-term policy uses $f(t) = 1$, so that policy is not affected by $f(t)$ at all; and

(b) the steady state policy is affected only by the level of the shift. If the level (from $f(t) = 1$ to 0.6 in our case) is biased, our updating procedure will compensate for the bias in the updated estimates of a_1 and a_2.

Thus, the policy development aspect of the procedure, our main focus here, is relatively insensitive to the choice of $f(t)$, and, although more research might be directed here, the nature of the derived optimal policy is unlikely to be affected.

6. Discussion and Conclusion

This paper has developed an approach for modeling and controlling a market penetration program when a word-of-mouth effect is present. An aspect of the procedure, applicable in many other product areas, is that it uses a bayesian procedure, developed on other, similar products, to permit parameter estimates earlier in the life of the product. This updating procedure is in marked contrast to other judgmental methods in that it:

(1) specifically and systematically accounts for information available in similar product-areas, and

(2) allows for updating of parameter estimates for purposes of forecasting and control, gradually improving the estimates as data come in.

The model developed here forecasts quite well in the cases studied. Most importantly, the model allows for calculation and dynamic updating of optimal marketing policies at a point in a product's life when sufficient historical data are not available to make clear "classical" inferences.

We also show that it is feasible both to estimate the effect of marketing variables in a trial/repeat framework and to dynamically update the derived policy. A modified version of the procedure appears applicable to a variety of similar new product marketing problems.

Extension of the model to incorporate other elements of the marketing mix are straightforward (given appropriate data). The model can also be used to examine how detailing effort can be allocated across a line of products.

The major source of statistical weakness and variability in estimation and forecasting surrounds knowledge of N, the market size. Incorporation of more precise market knowledge will simplify the estimation and improve the fit. This could happen either judgmentally or through the gathering of more than the single case used here in the development of a prior.

References

1. *Annals of Economics and Social Measurement* (October 1973).
2. Arndt, J., *Word of Mouth Advertising*, Advertising Research Foundation, New York, 1967.
3. Bauer, R. A. and Wertzel, L. H., "Doctor's Choice: The Physician and his Sources of Information," *J. Marketing Res.*, Vol. 3, No. 1 (1966).
4. Bass, Frank M., "A New Product Growth Model for Consumer Durables," *Management Sci.*, Vol. 15, No. 5 (1969).
5. Blattberg, Robert and Golanty, John, "TRACKER: An Early Test Market Forecasting and Diagnostic Model for New Product Planning," *J. Marketing Res.*, Vol. 15, No. 2 (1978).
6. Coleman, James, Katz, E. and Menzel, H., "The Diffusion of an Innovation Among Physicians," *Sociometry*, Vol. 20 (1957), pp. 253–270.
7. Dodson, J. A. and Muller, E., "Models of New Product Diffusion through Advertising and Word of Mouth," *Management Sci.*, Vol. 15, No. 4 (1978).
8. Eskin, Gerald J., "Dynamic Forecasts of New Product Demand Using a Depth of Repeat Model," *J. Marketing Res.*, Vol. 10 (1973).
9. Fourt, Louis A. and Woodlock, Joseph W., "Early Prediction of Market Success for New Grocery Products," *J. Marketing* (October 1960).
10. Horsky, Dan and Simon, L, S., "Advertising in a Model of New Product Diffusiom," TIMS/ORSA /EURO Marketing Conference, France, June 1980.
11. Kalish, Shlomo, "Models of the Diffusion of Innovations and their Implications for Government Policies," MIT Sloan School Masters Thesis, May 1980.
12. Kalwani, Manohar U. and Silk, Alvin J., "Structure of Repeat Buying for New Packaged Goods," *J. Marketing Res.*, Vol. 17 (1980), pp. 316–322.
13. Little, J. D. C., "Models and Managers: The Concept of a Decision Calculus," *Management Sci.*, Vol. 16, No. 8 (1970).
14. ———, "BRANDAID: A Marketing Mix Model, Part 2: Implementation, Calibration and Case Study," *Operations Res.*, Vol. 23, No. 4 (1975).

15. MAHAJAN, VIJAY AND MULLER, EITAN, "Innovation Diffusion and New Product Growth Models in Marketing," *J. Marketing*, Vol. 43, No. 4 (1979), pp. 55–68.
16. MIDGLEY, DAVID F., "A Simple Mathematical Theory of Innovative Behavior," *J. Consumer Res.*, Vol. 3 (1976), pp. 31–41.
17. MONTGOMERY, DAVID B., SILK, A. J. AND ZARAGOZA, C. E., "A Multiple Product Sales Force Allocation Model," *Management Sci.*, Vol. 18, No. 4, Part II (1971).
18. PARFITT, J. H. AND COLLINS, B. J. K., "The Use of Consumer Panels for Brand-Share Prediction," *J. Marketing Res.* (May 1968).
19. RAIFFA, HOWARD AND SCHLAIFFER, ROBERT, *Applied Statistical Decision Theory*, Harvard University Division of Research, Boston, Mass., 1961.
20. SILK, A. J. AND DAVIS, H., "Small Group Theory," in R. Ferber, ed., *Handbook of Marketing Research*, McGraw-Hill, New York, 1974.
21. ——— AND URBAN, G. L., "Pre-Test Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," *J. Marketing Res.*, Vol. 15, No. 2 (1978).
22. WILDT, ALBERT R. AND WINER, RUSSELL S., "Modeling Structural Shifts in Market Response: An Overview," *Proceedings, AMA Educators' Conference*, 1978, pp. 96–101.