A REPEAT PURCHASE DIFFUSION MODEL:
BAYESIAN ESTIMATION AND CONTROL

Gary L. Lilien, M.I.T.
Ambar G. Rao, N.Y.U.
Shlomo Kalish, M.I.T.

A REPEAT PURCHASE DIFFUSION MODEL:
BAYESIAN ESTIMATION AND CONTROL

Gary L. Lilien, M.I.T.
Ambar G. Rao, N.Y.U.
Shlomo Kalish, M.I.T.

Abstract

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 as well as a word-of-mouth effect. Bayesian estimation, with priors developed from past products, is used to update the parameters of the model. The procedure, shown to predict better and give more stable parameter estimates than classical procedures, is used to develop marketing policies for new product introduction.

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 [4] 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 [15], 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 where a word of mouth effect is significant. 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, incorporating past (pre-market) information with the data about the specific product, gives stable parameter estimates and policy guidelines.

2. Word of Mouth in New Product Diffusion

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. Examples of such situations are:

- · satisfied viewers of a movie, or users of a restaurant or resort recommending it to their friends,

- · doctors recommending a successful new drug to their colleagues,

- · women recommending a new food store to other housewives.

In each of these examples, initial users are attracted by some marketing effort -- advertising or sales promotion. Their use, then, enhances the impact of that effort on the rest of the potential user population.

In some situations it might be desirable actually 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 [1] for example, points to the importance of the word-of-mouth effect in developing advertising policies. Silk & Davis [16] review the literature dealing with influence processes in marketing situations, and stress the need for explicit understanding and measurement of these effects. Dodson and Muller [6] develop a general mathematical formulation for new product diffusion problems, both for durable and non-durable goods. They focus on advertising effects as well as word of mouth effects (although they do not treat issues of parameter estimation and control).

Thus, it appears that mathematical models of such marketing situations should explicitly consider the interaction between marketing expenditures and word-of-mouth effects, in the development of policies.

This paper hypothesizes and develops an estimation and control procedure for a model structure that explicitly includes the word-of-mouth effect. 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 [2]). Other components of the marketing mix include medical journal and magazine advertising and direct mail, but a smaller portion of the total marketing budget is devoted to these components than to detailing.

For a new product, the impact of company marketing effort is augmented by the word-of-mouth 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 [5].

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

The heart of this analysis is "trial and repeat" model structuring. A number of re-purchase models have been developed; the most popular use panel data collected at the test-market stage of new product introduction to estimate long-term rates. (Fourt and Woodlock [8], Parfitt and Collins [14], and Eskin [7].) Kalwani and Silk [9] develop some interesting insight into

the nature of repeat purchase estimation, formalizing some of Eskin's [7] hypotheses. All these models are descriptive in nature though; they focus on the forecasting issue, not the decision of controlling the level of marketing effort.

As noted earlier, Dodson and Muller [6] 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.

To our knowledge, there is no model or procedure available that focuses on the dynamic updating and control of a diffusion-type process in a marketing context. Our objective here is to develop and demonstrate the use of such a procedure.

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. Advertising and direct mail have been left out to simplify the exposition. Normally these marketing efforts are highly correlated with detailing effort so that not much information is lost by considering detailing alone in the model. The approach here differs from that developed by Montgomery, Silk and Zaragoza [13] 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 heavily dependent upon managerial judgment for calibration, i.e., a decision-calculus approach (Little [10]).

In the context of development, 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. Manage-

ment has to allocate detailmen's time across a variety of products; there-
fore 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 allo-
cation. A policy that seems to fit these marketing realities is of a pulsed
type -- i.e., a short period of high effort detailing during product intro-
duction followed by a much lower "maintenance level" detailing over the re-
mainder of the planning horizon.

Managerial use of the model in the context of a new product presents
some novel aspects. Since the key period in the 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 cal-
ibration strategy. We believe that the appropriate approach is to model a
variety of products, obtaining the model parameters for each, and using in-
formation about those parameters to develop a prior distribution of para-
meter estimates for the new product. These estimates are used to develop
initial policy decisions, which are updated as sales data become available.

3. The General 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(≤N*) may eventually prescribe the drug. We observe the number of prescriptions which we assume is closely related to the number of prescribing doctors. Note that our model derivation assumes linearity in this relationship, which is not the case in general. The most productive doctors, who write a disproportionate number of prescriptions, are also more likely to be early adopters. This is critical if we wish to make inferences about the true value of N. Our objectives, however, are to infer (a) the time path of product sales and (b) develop promotional policies. For this purpose the concept of an "average" doctor is sufficient as operationalized in the process described below.

Figure 1 describes the process we wish to study, which in its most complete form, has three states, (1) never prescribed, (2) prescribing, and (3) used to prescribe. The activities affecting the various flows are labelled. Here we have a trial structure (movement from state 1 to state 2) and a repeat structure (remaining in state 2 or movement from state 3 to state 2).

Note that early in the life of the drug, the flow will be almost entirely from state 1 to state 2. Later on, the flow switches to a state 2 and state 3 interchange.

The three state model, however complete, has too many parameters for efficient estimation for any of the data sets we have examined. 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 three 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.
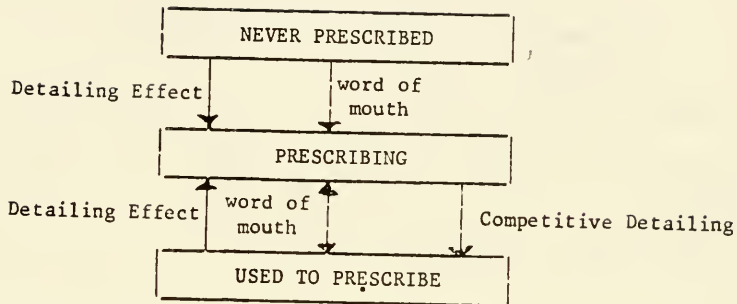
Figure 1:

Complete Flow Model Describing the Process

Let

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

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

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

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

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

Note that although the model is structured in terms of the number of prescribing doctors, the data we observe are the number of prescriptions. Hence, we assume that

$$K_1(t) + K_2(t) = C_2(t) E(W)$$

or

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

We describe the flows between these two classes of doctors (1 = not prescribing and 2 = prescribing) as follows:

The flow from C_1 to C_2 is affected

a) by level of detailing

b) by word-of-mouth effect related to the change in the number of prescribing doctors.

The flow from C_2 to C_1 is affected by

a) competitive detailing.
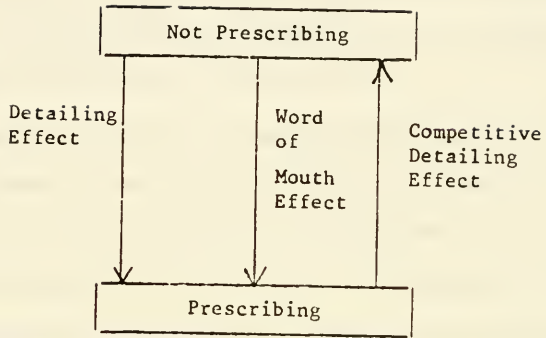b) possible word of mouth.

Figure 2 describes this process.

Figure 2:   Simplified Flow Model Describing the Process

Let

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

$$d(t) \qquad\qquad = \text{ level of detailing at t}$$

$$\lambda_2(C_2(t)-C_2(t-1)) = \text{ word of mouth effect}$$

$$\lambda_3(\bar{d}(t)) \qquad\quad = \text{ competitive detailing effect}$$

$$f(t) \qquad\qquad = \text{ decay factor for detailing effect}$$

The decay factor $f(t)$ allows us to consider the possibility that the same amount of detailing effort might have varying effectiveness at different stages in the life of the drug. Normally, we would expect $f(0) = 1$ and $f(t)$ to be non-increasing -- i.e. as the product becomes more established, detailing becomes less effective, unless a new communication strategy is developed. We now define

$$\lambda_1(d(t),t) = \lambda_1(d(t) \cdot f(t))$$

to be detailing effectiveness. This formulation is similar to Little's copy effectiveness factor in BRANDAID [11]. An alternative formulation -- making λ_1 itself a function of time was rejected as overly complex. For ease of notation, we will use the term $\lambda_1(d(t))$ to refer to the $\lambda_1(d(t),t)$ above. Now,

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

and

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

Note that this model-structure handles the word-of-mouth term (λ_2) in a different way than the interactions in most diffusion models. The advantage

of this formulation is that it approximates the $(N-X)X = NX-X^2$ interaction
term used in Bass [3] and other formulations by a time-based difference
$(X(t)-X(t-1))$, which permits a <u>negative</u> word-of-mouth effect (lost sales) due
to competitive activity or to bad product experience, for example. This model
is thus symmetric in that the model structure handles competitive word-of-
mouth explicitly.

This model does have several important simplifying assumptions. The first
is that N, the number of doctors in the class, is assumed fixed. Mahajan and
Peterson [12] show how this assumption can be relaxed.

The second assumption is that all doctors are in the same class (psychi-
atrists 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 how-
ever and data needed to attempt such extensions are not available.

4. 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 is 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 responses $\lambda_i(\ )$ 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. Thus sales 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 imply that marketing efforts should be either zero or as large as possible. Non-linearity of response for determining detailing effort for a brand is essential.

Consider the following model form. It is a simple form that contains non-linearity.

Let

$$\lambda_1(d(t)) = a_1 d(t) + a_2 d^2(t)$$

$$\lambda_2(C_2(t) - C_2(t-1)) = a_4(C_2(t) - C_2(t-1))$$

$$\lambda_3(\bar{d}(t)) = a_3 \bar{d}(t)$$

Substituting $C_1(t) = N - C_2(t)$ in (1a) we get

$$
\begin{aligned}
(2) \quad C_2(t+1) - C_2(t) &= \lambda_1(d(t)) \ (N - C_2(t)) \\
&\quad + \lambda_2(C_2(t) - C_2(t-1)) \\
&\quad - \lambda_3(\bar{d}(t)) \cdot C_2(t) \\
\\
&= N\lambda_1(d(t)) \\
&\quad - \lambda_1(d(t)) \ C_2(t) \\
&\quad + \lambda_2(C_2(t) - C_2(t-1)) \\
&\quad - \lambda_3(\bar{d}(t)) \cdot C_2(t) \quad ,
\end{aligned}
$$

and plugging in the proposed functional forms for the λ_i's we obtain:

$$
\begin{aligned}
(3) \quad C_2(t+1) - C_2(t) &= (a_1 d(t) + a_2 d^2(t))(N - C_2(t)) \\
&\quad - a_3 \bar{d}(t) C_2(t) + a_4(C_2(t) - C_2(t-1))
\end{aligned}
$$

This equation contains five unknown parameters: a_1, a_2, a_3, a_4 and N, with N appearing in a way that makes it impossible to use conventional linear estimation procedures. Direct estimation of the parameters using nonlinear estimation methods leads to unstable results due to multicollinearity, present in all the data sets we examined. However, if N is known, then (3) becomes linear in its parameters.

Thus, estimation is simplified if we can develop an estimate of N. We do so by fitting a model that is linear in the response to detailing and deriving N from this model. The linear model assumes:

$$\lambda_1(\ ) = \lambda_3(\ ) = B,$$
$$\lambda_2(\ ) = C$$

and that $\qquad BN = A.$

Thus, equation (2) reduces to

$$(4) \qquad C_2(t+1) - C_2(t) = A \ d^2(t)$$
$$- B(d(t)+\bar{d}(t))C_2(t)$$
$$+ C(C_2(t)-C_2(t-1)).$$

We estimate the parameters A, B, and C using ordinary least squares and estimate N from the fact that N = A/B. Since the estimates of A and B, $\hat{A}$ and $\hat{B}$ respectively are approximately bivariate normal, the distribution of N can be developed analytically; however it is simpler to obtain this 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

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

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

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 $\hat{A}/\hat{B}$.

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 of the linear model are shown in Table 2, and the distribution of N in Figure 3. Table 3 shows the parameter estimates for the nonlinear model, assuming N = 10,700, the maximum likelihood estimate, using only the first 12 points for fitting.

The function f(t) was modeled as f(t) = 1, t<12, = .6, t$\geq$12. This form, consistent with historical decay patterns in the market, works 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).

Table 4 shows the forecasts obtained using the nonlinear model, together with the actual sales, and Figure 4 graphs these series. The forecasts are excellent, with a root mean square error of 43.86.

Table 5 shows the parameter estimates of a third order polynomial that was fit to the data using the first 15 points, and Figure 5 the resultant forecasts. Similarly, Table 6 and Figure 6 show the parameter estimates of a third order autoregressive scheme and the resulting forecasts. In each of these cases, the order of the model was selected as having the same number of parameters as our model.

TABLE 1: ANALYSIS DATA

Quarter	Case 1			Case 2		
	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	131	91	834	70
4	57	396	213	58	1023	111
5	53	411	302	63	953	150
6	46	417	303	67	837	173
7	56	462	410	63	924	160
8	61	467	668	84	953	168
9	44	498	610	84	736	195
10	53	488	775	81	992	223
11	51	523	797	72	776	259
12	49	581	672	81	662	307
13	44	611	697	67	822	348
14	43	581	829	47	1024	397
15	40	585	803	45	989	405
16	38	493	798	47	777	442
17	41	505	764	72	992	448
18	35	516	648	65	756	501
19	32	485	746	79	852	506
20	27	444	609	66	1103	489
21	28	463	553	80	946	530
22	26	427	505			446
23	24	466	549			458
24	22	472	522			471

Table 2:

Parameters of the Linear Model,

First Set of Data

Variable	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 = .98
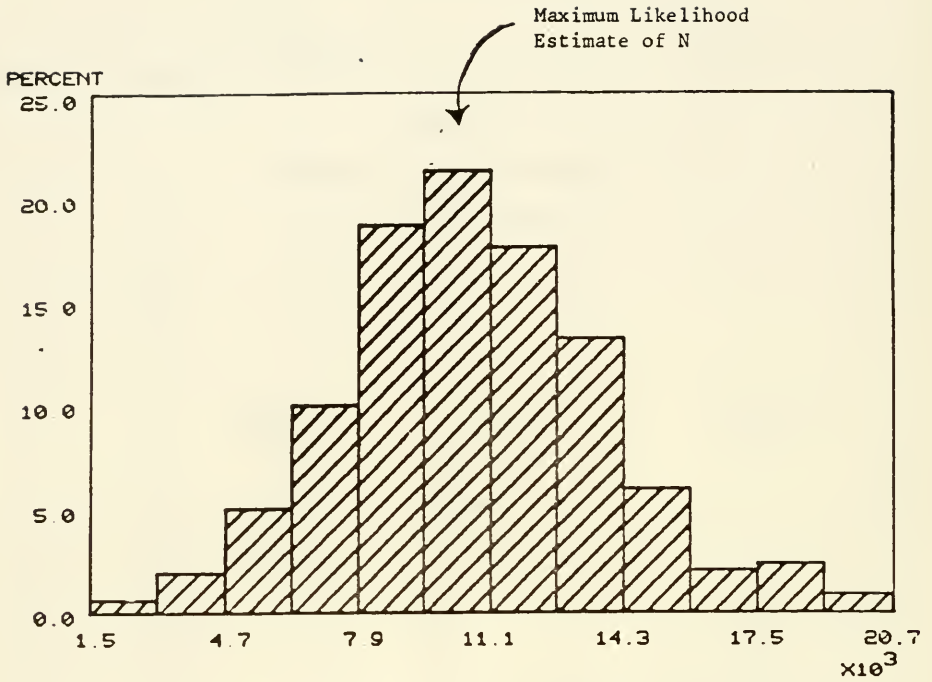
Figure 3: Simulated Distribution of N

Table 3:

Nonlinear Model Parameter Estimates

Coef	Value	T-Stat	
a_1	2.28×10^{-5}	0.11	F(3;9) : 168
a_2	7.99×10^{-7}	0.21	Corrected R-Square = .98
a_3	7.19×10^{-5}	1.17	
a_4	0.56	2.06	

* Using MLE for N = 10,700

Table 4:

Forecasts of Sales Data, Case 1,
Using the Nonlinear Model*

	Series Values	Forecast
1	0	
2	64	
3	133	
4	213	
5	272	
6	311	
7	410	
8	578	
9	718	
10	775	
11	582	
12	783	
13	785	
14	786	
15	793	772
16	796	749
17	775	725
18	730	702
19	667	678
20	604	652
21	557	626
22	537	601
23	536	577
24	536	552

Root Mean Square Error = 43.86

* Using MLE for N = 10,700

Figure 4:

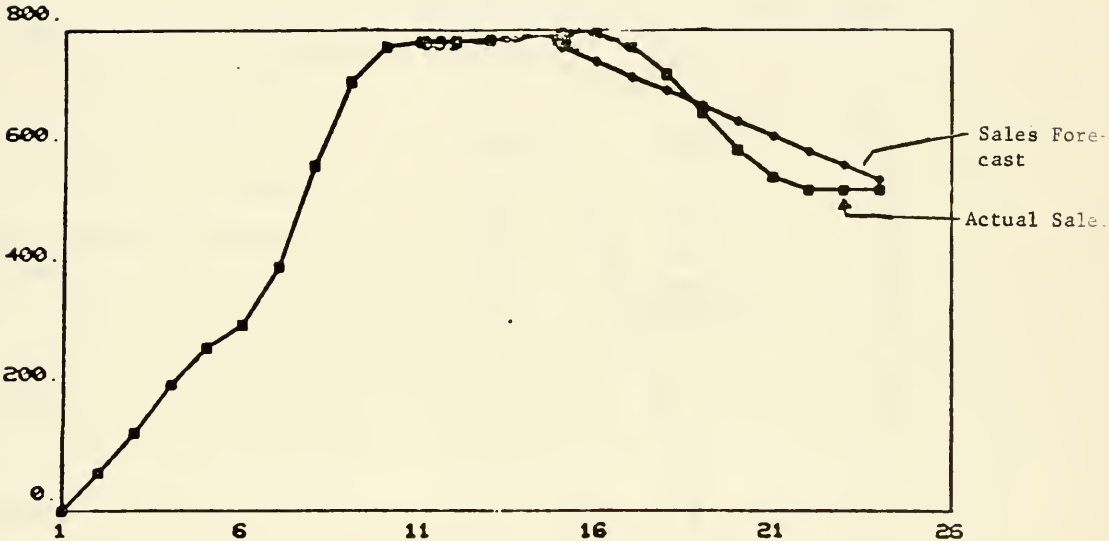Forecasts from the Nonlinear Model and Actual Sales, Case 1

Table 5:

Polynomial Model Coefficients

Model: $X(t) = A + BT + CT^2 + DT^3$

Coef	Value	T-Stat
A	61.1	0.32
B	−23.9	−0.31
C	17.7	1.94
D	−0.86	−2.59

$F(3;9) = 112$

Corrected R-Square = .965
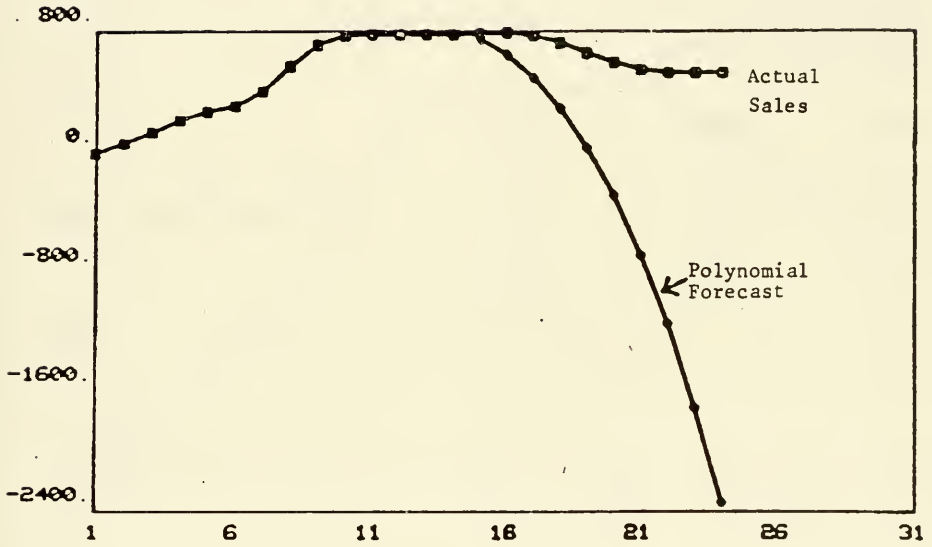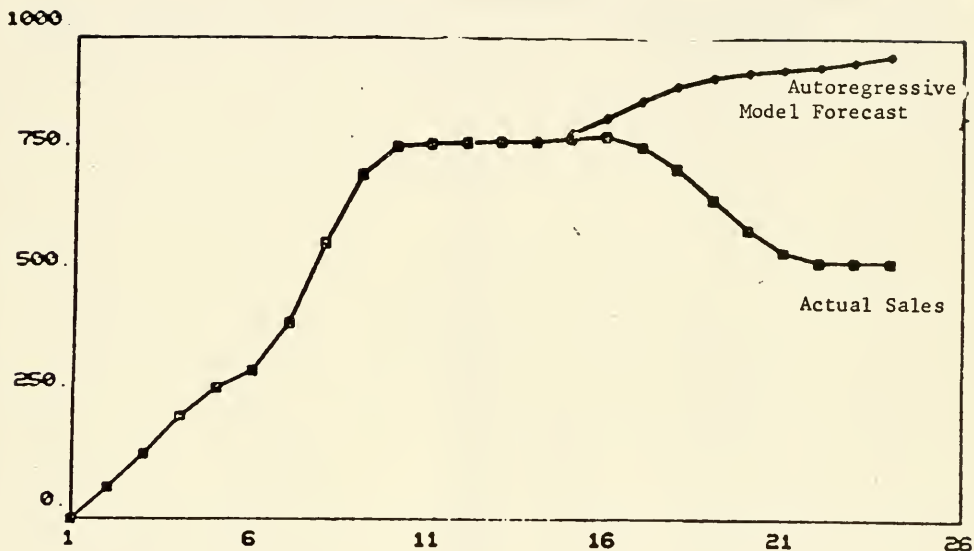
Figure 5:

Forecasts from the Polynomial Model vs. Actual Sales

Table 6:

Third Order Autoregressive Model Parameter Estimates

Model: $X(t) = A + BX(t-1) + CX(t-2) + DX(t-3)$

Coef	Value	T-Stat	
A	64.90	2.37	$F(3;8) = 206.8$
B	1.99	7.47	Corrected R-Square = .98
C	-1.65	-3.55	
D	0.60	2.37	

Figure 6:

Forecasts Using the Autoregressive Model

The forecasts from the polynomial model are obviously unsatisfactory, since they become negative. The autoregressive model does better, but the RMS error in this case is 295.89 (see Table 7), 7 times greater than the RMSE for our model.

Based on the results from these data, we have some confidence in the model.

Table 7:

Forecasts Using Autoregressive Model, and Actual Sales

	Actual Sales	Autoregressive Model Forecast
1	0	NA
2	64	NA
3	133	NA
4	213	NA
5	272	NA
6	311	NA
7	410	NA
8	578	NA
9	718	NA
10	775	NA
11	782	NA
12	783	NA
13	785	NA
14	786	NA
15	793	801
16	796	832
17	775	869
18	730	899
19	667	919
20	604	928
21	557	934
22	537	941
23	536	952
24	536	964

RMS = 295.89

5. Using the Model

Now we turn to the question, how does one use the model in the typical new product situation when little data is available? The proposed procedure is similar to that developed in the previous section -- i.e. using a linear model to obtain an estimate of N, and then estimating the parameters of the nonlinear model. The only differences are that

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

(ii) priors for the parameters, A, B, C and a_1, ..., 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 becomes available.

The use of this procedure assumes that the structure of sales growth will be similar from drug class to drug class, although the target population might be different.

Thus our procedure is as follows:

(a) Estimate parameters of the linear model, using ordinary least squares or bayesian regression (if past data are available).

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

(c) Pick several values of N, $N_1 \ldots N_k$, from the distribution derived in (b), and, incorporating prior estimates of $a_1, \ldots a_4$ from previous data, develop posterior estimates of $a_1 \ldots a_4$.

(d) Develop a detailing policy to maximize expected long-term per period profit from the distribution of N.

We illustrate this procedure in this and the next sections, 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, modified to reflect the slower diffusion rate expected for drugs in this (second) class. In particular $-a_1/2a_2$ was set initially equal to 90, consistent with historical detailing levels in this class. The variance-covariance matrix was used directly from case 1. Note that this assumes that the two drugs have identical market characteristics, their covariances differing only due to sampling variation. Greater experience with historical cases will lead to more realistic priors; as we will see this level is quite close to optimal even after adding 16 data points. The updated (posterior) coefficients for the linear model, using priors from case 1 and the first eight data points are given in Table 8.

Table 8: Posterior Coefficients, Linear Model (Case 2)

	Updated Coefficients (mean)	t-values
A	0.171	4.18
B	3.17×10^{-5}	5.42
C	0.805	27.12

The density of N, obtained by simulation, is shown in Figure 7. Figure 8 shows the Bayesian nonlinear model forecasts based on the first eight data points, showing the forecasts obtained using the maximum likelihood estimate
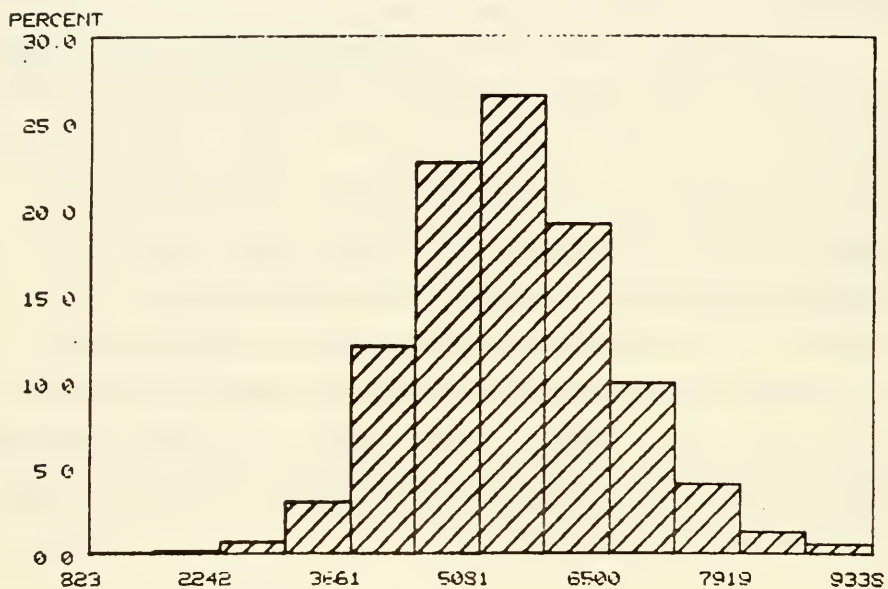
Figure 7: A Simulated Density of N, Case 2

Figure 8:

Case 2: Bayesian Estimate, Using First 8 Points
Plus Case 1 data as Prior, Prediction (and Pre-
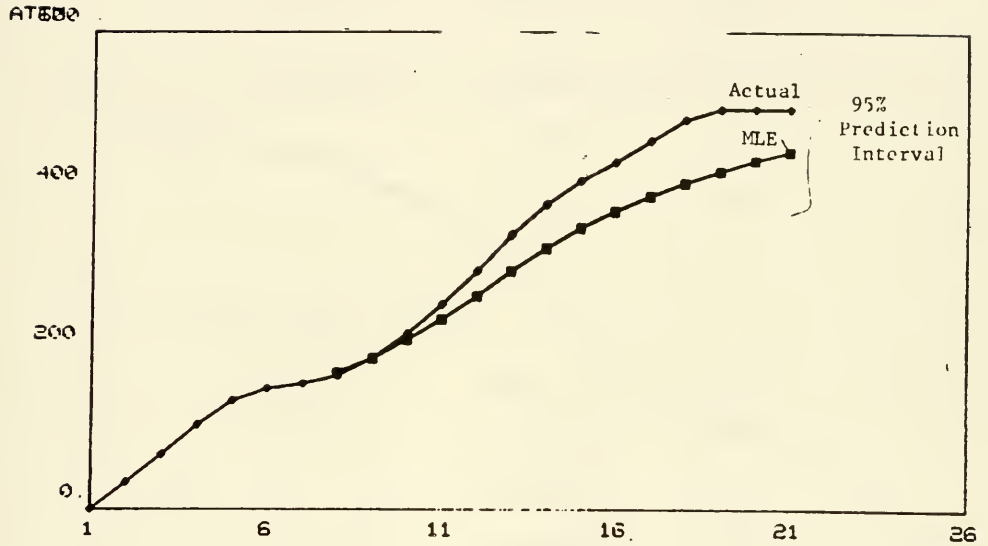diction Interval) on Rest of Data

Figure 9:

Updated Forecast, Adding 8 More Points
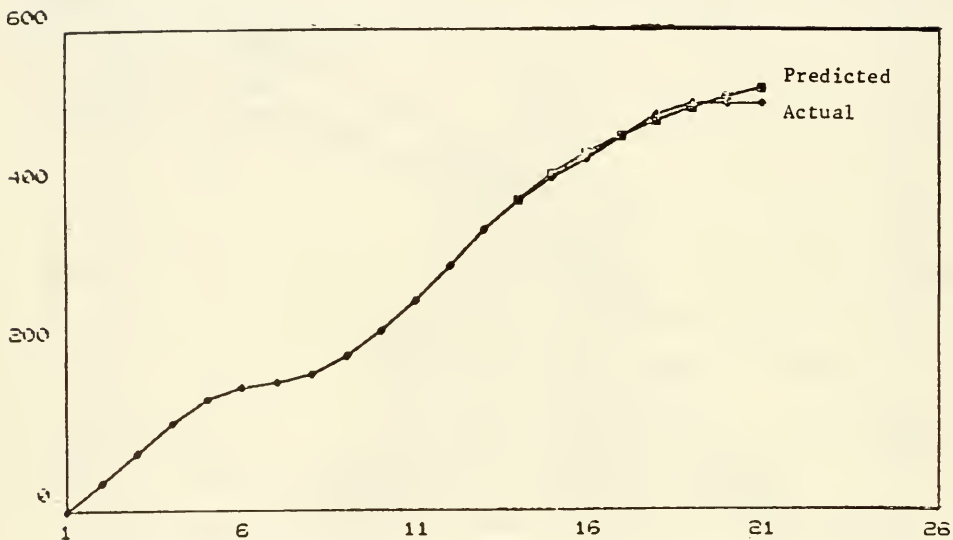Second Data Set, Nonlinear Model

-33-

Table 9:

Posterior Estimates of the Coefficients

a) After 8 points:

Variable	Value	t-stat
a_1	$3.41 \cdot 10^{-5}$	15.5
a_2	$-1.82 \cdot 10^{-7}$	-2.80
a_3	$1.11 \cdot 10^{-5}$	1.71
a_4	0.81	4.50

b) After 6 additional points:

Variable	Value	t-stat
a_1	$3.40 \cdot 10^{-5}$	17.4
a_2	$-1.83 \cdot 10^{-7}$	-3.05
a_3	$1.10 \cdot 10^{-5}$	2.00
a_4	0.82	4.82

of N, together with a 95% prediction interval. Figure 9 shows the improved fit and prediction as 6 more points are added. Table 9 summarizes the estimates of the a_i for the models used to produce these two sets of forecasts. We should note that the bayesian procedure actually allows us to use the five-parameter nonlinear model, even with a small number of data points available. Without the bayesian approach, we would be forced to run the linear model in the early part of the life of the product -- highly undesirable both from a policy, and a forecasting viewpoint. Figure 10 shows the forecasts obtained with 8 data points using the linear model. A comparison with Figure 8 indicates the tremendous improvement made possible through the use of a prior. Figure 11, the linear model with priors, does better, but still is far worse than the nonlinear bayesian model. Table 10 summarizes the root mean square errors obtained for each of the three forecasting methods.

Table 10

Summary of Forecasting Accuracy

	RMSE
OLS (Figure 10)	237.5
Bayesian Estimates, Linear Model (Figure 11)	120.6
Bayesian Estimate, Nonlinear Model (Figure 8)	31.1

Thus, the nonlinear model with bayesian estimates predicts best. As a forecasting procedure, it seems useful; next we move to issues of policy development.
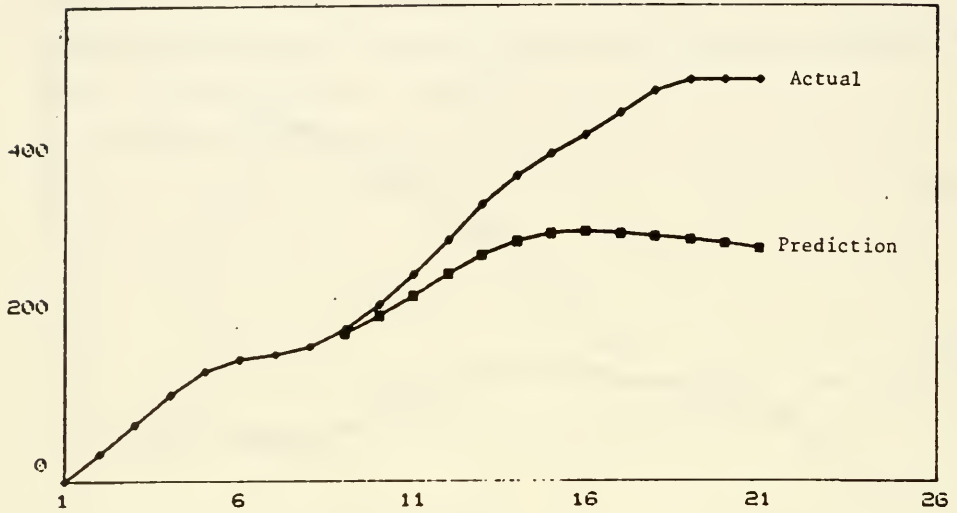
Figure 10: Case 2: OLS Estimation Using First 8 Points
Prediction of Rest of Data (Linear Model)
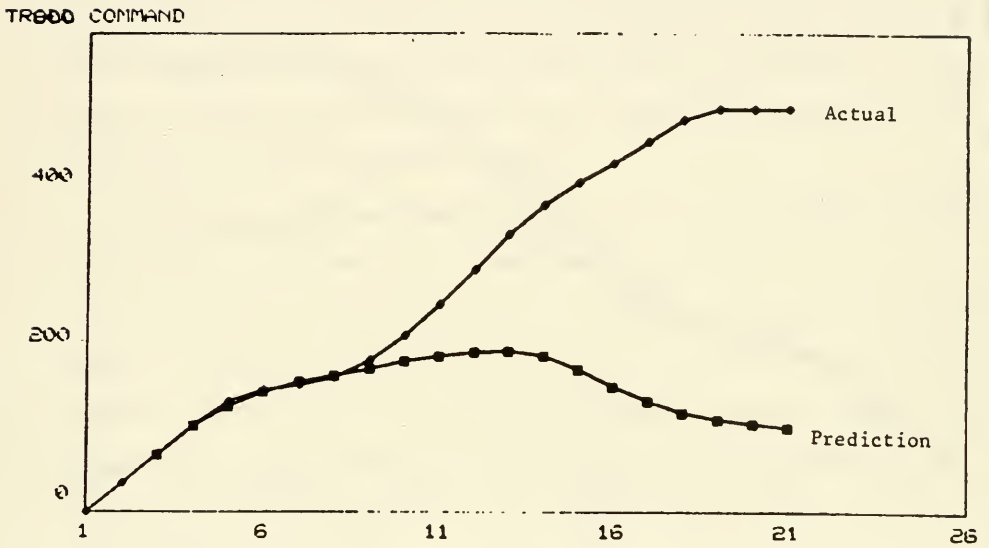
	Coefficient	T-Stat
A	0.241	1.76
B	9.66×10^{-5}	2.11
C	0.614	2.70

Corrected R-square = 0.977
F-Ratio (2,3) = 105.4

A

Figure 11: Case 2: Bayesian Estimate, Using First 8 Points
Case 1 Data as Prior, Prediction of Rest of Data (Linear Model)

	Updated, Posterior Coefficients (mean)	t-stat
A	0.171	4.18
B	3.17×10^{-5}	5.42
C	0.805	27.12

6. 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, as follows:

a. Because of 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.

b. Product management is dealing with a portfolio of drugs, all of which are promoted by the same detailing force. 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. Assuming a sequence of new product introductions by the company, an approximate "steady state" policy for the detailing force would be to devote a certain fraction of its effort to new products and the balance to "maintenance detailing."

In view of the above we shall 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."

The introductory phase goal then 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 4 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 values 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:

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

(assuming $\bar{d}$ = constant).

Per period profit is, then:

$$\Pi_s(N) = K_o C_2 - K_1 d$$

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

$$\int \Pi_s(N) f_N(n)$$

where $f_N(n)$ refers to the distribution of the number of potential prescribers, calculated from the procedure described in Section 5.

For our case, the values of K_o and K_1 are 66 and 95, respectively. Table 11 gives optimal policies as a function of N and the expected profit associated with the different policies. The optimal policy is roughly 80-83 per 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 $d = \frac{-a_1}{2a_2}$, 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 and d then we get that the steady-state share is:

$$C_2/N = \frac{a_1 d + a_2 d^2}{a_1 d + a_2 d^2 + a_3 \bar{d}} \simeq .47$$

By our reasoning, then, the suggested policy is to set a detailing level at 93 until a share of about .47 is reached and then back down to around 80.

One of the powers of the bayesian approach is that updating of policies is natural as more data is collected. In a manner identical to that above, the updated, optimal long term policy after 6 more points are available, was calculated as 79.8, quite close to the one calculated previously. In this case, even after 6 additional periods, the optimal policy remains stable. (In practice updating would occur each time new data were received from the field.)

Table 11: Optimal Policy Development

N_i	$P(N_i)$	Optimal Per Period Detailing Level, Given N_i	Expected Per Period Profit ($1000's)
2000	.001	47.6	1.58
3500	.002	65.3	1.76
5000	.007	75.6	1.86
6500	.030	80.0	1.87
8000	.120	82.6	1.87
9500	.225	84.1	1.86
11000	.265	84.6	1.84
12500	.191	86.7	1.84
14000	.099	87.2	1.84
15500	.040	87.8	1.84
17000	.013	88.3	1.83
18500	.006	88.7	1.62

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 .6 in our case) is biased, our updating procedure will compensate for the bias in the updated estimate of a_1 and a_2. This occurs in the results reported in Figure 10.

Thus, the policy development aspect of the procedure, our main focus here, is relatively insensitive to the choice of $f(t)$.

7. Detailing Force Implications

In the last section we showed how detailing policies can be computed for a single product line. If we can assume that the different product lines constituting the portfolio of product offerings are independent of one another, then portfolio profit maximization can be achieved by selecting the optimal market share for each product line individually, so long as the total number of detailers required does not exceed the available force.

In general however, the portfolio maximization problem, given a fixed detailing force D can be addressed as a lagrangian problem. If $P_{1t}(m)$ is the profit associated with the ith product line in period t with a steady state share m, and $d_{1t}(m)$ is the detailing force required in the same period (note that given m and our policy as in the previous section, d_{1t} can take only one of two possible values), we wish to

Maximize $\sum_{i,t} P_{1t}(m)$

subject to $\sum d_{1t} \leq d.$

Detailing manpower and detailing cost will be assumed to be linearly related, a reasonable assumption given that some detailing will always be done. Market share is a concave function of detailing activity both for the introductory phase and the maintenance phase, as is illustrated in Figures 3 and 4. Therefore $P_{1t}(m)$ is convex in d_{1t}. This implies that solutions to the lagrangian problem

(10) $$\mathcal{L}(d) = \sum_{it} P_{1t}(m) - \sum_{t} \lambda_t (\sum_{i} d_{1t} - d)$$

will be unique. In addition λ_t will provide us the marginal value of additional detailers.

8. Discussion and Conclusion

This paper has developed an approach toward 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 judgemental methods in that it:

(1) specifically, 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 test demonstrated, and the bayesian model works much better than a more standard procedure. 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.

APPENDIX: A NOTE ON THE BAYESIAN REGRESSION PROCESS WITH A NATURAL CONJUGATE

Suppose our regression model is:

$$v_i = x_i \beta + \epsilon_i$$

Then, the density of y is:

$$f_N (\gamma / x^1 \beta, h) = (2\Pi)^{-\frac{1}{2}} e^{-\frac{1}{2}h(y_i - x_i^1 \beta)^2} h^{\frac{1}{2}}$$

where $h = \frac{1}{\sigma} 2$

The likelihood of a sample $y_1 \cdots y_n$ is:

$$2\Pi^{-\frac{h}{2}} e^{-\frac{1}{2}h} \Sigma(y_i - x_i^1 \beta)^2 \cdot h^{\frac{n}{2}}$$

With the kernal (in matrix notation) of

$$e^{-\frac{1}{2} h} (y - X\beta)^T (y - X\beta)_h \; n - /2$$

Let b be the solution of the normal equations:

$$X^1 X b = X^1 y$$

If h is known, we proceed as follows:

Let the prior of β be normal $\approx N(\underline{b}^1, (h n^1)^{-1})$

(where n^1 is a positive definite and symmetric. Note that n below is $X^1 X$.)

Multiplying the kernal of the prior with the kernal of the likelihood gives:

$$T = (y - X\beta)^T (y \div X\beta) + (\beta - b^1)^T n^1 (\beta - b^1) = T_1 + T_2$$

After some algebra:

$$T_1 = (\beta - b^{11})^T n^{11} (\beta - b^{11})$$

and

$$T_2 = b^{1T} h^1 b^1 + y^T y - b^{11T} n^{11} b^{11}$$

where

$$n^{11} = n + n^1 = X^T X + n^1$$

and

$$b^{11} = (n^{11})^{-1} [n^b + n^1 b^1]$$

Now β only appears in T_1 (clearly in normal form) so the kernal of the posterior is normal $\simeq N(b^{11}, (hn^{11})^{-1})$.

When the precision (h) is not known, the analysis is similar, but with a normal-gamma (Studentized) prior and posterior densities.

Here we need to let:

ρ = rank n $\quad\quad \nu = n - \rho$

ρ^1 = rank n^1 and $\mu = \frac{1}{\nu}(y - X\beta)^T (y - X\beta)$

ρ^{11} = rank n^{11}

The prior joint density is:

$f(\beta, h / b^1, \nu^1, n^1, \mu^1)$ ~

$\sim e^{-\frac{1}{2}h(\beta-b^1)^T n^1 (\beta-b^1)} h^{\frac{1}{2}} e^{-\frac{1}{2}h\nu^1\mu^1} h^{\frac{1}{2}\nu^1 - 1}$

And the posterior is:

$f(\beta, h / b^{11}, \nu^{11}, h^{11}, \mu^{11})$

where

$n^{11} = n^1 + n, \; b^{11} = (n^{11})^{-1} (n^1 b^1 + nb)$

$\nu^{11} = \nu^1 + \nu + \rho$ and

$\mu^{11} = \frac{1}{\nu^{11}} [(\nu^1\mu^1 + b^{1T}n^1 b^1) + (\nu\mu + b^T nb) - b^{11T}n^{11}b^{11}]$

Note here that the marginal densities of B and h, respectively, are:

$f(\beta / b, \; h / \nu, \mu) \simeq [\nu + (\beta-b)^T (n/\nu) (\beta-b)]^{-\frac{1}{2}(\nu+r)}$
  (student)

and

$f(h / \nu, \mu) \quad \sim e^{-\frac{1}{2}h\nu\mu} h^{\frac{1}{2}\nu - 1}$
  (gamma)

References

1. Arndt, J. Word of Mouth Advertising. New York: Advertising Research Foundation, 1967.

2. Bauer, R. A. and L. H. Wertzel. "Doctor's Choice: The Physician and his Sources of Information," Journal of Marketing Research. Vol. 3, No. 1 (February 1966).

3. Bass, Frank M. "A New Product Growth Model for Consumer Durables," Management Science. Fol. 15, No. 5 (January 1969).

4. Blattberg, Robert and John Golanty. "TRACKER: An Early Test Market Forecasting and Diagnostic Model for New Product Planning," Journal of Marketing Research. Vol. 15, No. 2 (May 1978).

5. Coleman, James, E. Katz and H. Menzel. "The Diffusion of an Innovation Among Physicians," Sociometry. XX (1957), 253-270.

6. Dodson, J. A. and E. Muller. "Models of New Product Diffusion through Advertising and Word of Mouth," Management Science. Vol. 15, No. 4 (November 1978).

7. Eskin, Gerald J. "Dynamic Forecasts of New Product Demand Using a Depth of Repeat Model," Journal of Marketing Research. Vol. 10 (May 1973).

8. Fourt, Louis A. and Joseph W. Woodlock. "Early Prediction of Market Success for New Grocery Products," Journal of Marketing (October 1960).

9. Kalwani, Manohar U. and Alvin J. Silk. "Structure of Repeat Buying for New Packaged Goods," MIT Sloan School of Management Working Paper 1032-78 (October 1978).

10. Little, J. D. C. "Models and Managers: The Concept of a Decision Calculus," Management Science. Vol. 16, No. 8 (April 1970).

11. Little, John D. C. "BRANDAID: A Marketing Mix Model, Part 2: Implementation, Calibration and Case Study," Operations Research. Vol. 23, No. 4 (July-August 1975).

12. Mahajan, V. and R. A. Peterson. "Innovation Diffusion in a Dynamic Potential Adopter Population," Management Science. Vol. 15, No. 4 (November 1978).

13. Montgomery, D. B., A. J. Silk, and C. E. Zaragoza. "A Multiple Product Sales Force Allocation Model," Management Science. Vol. 18, No. 4, Part III (December 1971).

14. Parfitt, J. H. and B. J. K. Collins. "The Use of Consumer Panels for Brand-Share Prediction," Journal of Marketing Research. (May 1968).

15. Raiffa, Howard and Robert Schlaiffer. <u>Applied Statistical Decision Theory</u>. Harvard University Division of Research, Boston, MA, 1961.

16. Silk, A. J. and H. Davis. "Small Group Theory," in R. Ferber, ed., <u>Handbook of Marketing Research</u>. New York: McGraw-Hill, 1974.

17. Silk, A. J. and G. L. Urban. "Pre-Test Market Evaluation of New Packaged Goods: A Model and Measurement Methodology," <u>Journal of Marketing Research</u>. Vol. 15, No. 2 (May 1978).

1097 -80

1098

←1089 -79

1099-80

1091-79

1100-80

1093-79

1094-79

1095

1096-79