Forecasting Repeat Sales at CDNOW: A Case Study
Author(s): Peter S. Fader and Bruce G. S. Hardie
Source: *Interfaces*, Vol. 31, No. 3, Marketing Engineering (May - Jun., 2001), pp. S94-S107
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/25062705

Forecasting Repeat Sales at CDNOW: A Case Study

PETER S. FADER
fader@wharton.upenn.edu
The Wharton School
University of Pennsylvania
Philadelphia, Pennsylvania 19104-6371

BRUCE G. S. HARDIE
bhardie@london.edu
London Business School
Regent's Park
London NW1 4SA, United Kingdom

We conducted a modeling exercise in conjunction with the online music retailer CDNOW to develop a simple stochastic model of buyer behavior capable of forecasting medium-term aggregate CD purchasing by a cohort of new customers. We modeled weekly sales using a finite mixture of beta-geometric distributions with a separate time-varying component to capture nonstationarity in repeat buying. The resulting model can easily be implemented within a standard spreadsheet environment (for example, Microsoft Excel). It does a good job of describing the underlying sales patterns and produces an excellent medium-term forecast.

With the growth of e-commerce, many companies are facing challenges in figuring out how to make effective and efficient use of the detailed transaction information that they are rapidly accumulating. While some writers on database marketing and one-to-one marketing suggest using statistical models to gain managerial insights [Mulhern 1999; Forrester Report 1999], there are very few published examples to guide managers' efforts in this direction.

We undertook an exploratory study in conjunction with CDNOW, a leading online music retailer, to develop an easily implementable model of buyer behavior capable of forecasting medium-term aggregate CD purchasing by a cohort of CDNOW customers. Aggregate-level forecasts are critical inputs to any attempt to value a customer base, and they serve as a diagnostic to help firms gauge the effec-

0092-2102/01/3103/S094/$05.00
1526-551X electronic ISSN
This paper was refereed.

MARKETING—BUYER BEHAVIOR
FORECASTING—APPLICATIONS

tiveness of various short-term marketing programs (for example, they can provide baseline sales estimates against which the performance of promotions can be evaluated).

At the time of this study (1997 through 1998), many commentators felt that the Internet was still in its infancy and that any forecasting exercise would be futile. For example, Buchanan and Lukaszewski [1997, p. 143] made the comment that:

> At this stage of the Internet's evolution, accurate sales forecasts are as much of an oxymoron as "military intelligence."

However, it was—and still is—our view that patterns of buying behavior are fairly consistent across purchasing channels (including the Internet) and that developing a forecasting model would therefore be a fruitful exercise.

Background

CDNOW is one of the oldest and largest online retailers, having sold different forms of music (and related products) on the World Wide Web since 1994. It carries approximately 500,000 different albums—about 10 times as many as the typical bricks-and-mortar megastore—and it reports store traffic of over 200,000 visitors per day. During its first five years of operations, CDNOW attracted over 700,000 unique customers who made purchases at the Web site.

In this work, we focused on a single cohort of new customers who made their first purchase at the CDNOW Web site in the first quarter of 1997. We have data covering their initial (trial) and subsequent (repeat) purchases for the three-month period (January through March 1997) during which over 23,000 individuals bought nearly 70,000 CDs. We wanted to forecast the future (repeat) purchasing of these customers using a model calibrated with these first-quarter data. Furthermore, we wanted to develop a model that could be implemented with little difficulty.

Faced with these data, some analysts might use an existing model designed for customer-base analysis [Allenby, Leone, and Jen 1999; Colombo and Jiang 1999; Schmittlein, Morrison, and Colombo 1987; Schmittlein and Peterson 1994]. Some of these models treat underlying buyer behavior as if it were stationary. For example, Schmittlein, Morrison, and Colombo [1987] model individual-level purchasing via a Poisson counting process and overlay an exponential "death process" to capture customer attrition; they use gamma distributions to capture customer heterogeneity in these two elements. Aggregate-level predictions, as well as individual inferences, can be derived from such a model.

Patterns of buying behavior are fairly consistent across purchasing channels. Start with simple models and evolve towards more complete models.

The problem with these approaches is that they require an analyst who is used to dealing with large customer-level datasets—all of the above models are estimated using detailed customer-level data—and who has sophisticated model-building skills and access to appropriate computational software. However, such people are rare in most organizations. Furthermore, firms often view the costs (financial and psychological) of implementing and maintaining such models as outweighing the benefits that arise from doing so.

We therefore sought to develop a simple forecasting model that could be implemented easily using readily available software with which most business people would be familiar—ideally a common spreadsheet package, such as Microsoft Excel. Central to this goal was structuring the raw data in an aggregate form that would be easy for an analyst to manage, while at the same time still allowing us to develop a well-specified model of repeat purchasing.

For the cohort of customers who first purchased at the CDNOW Web site in the first quarter of 1997, we chose to work with a summary of total purchasing (Table 1). This includes the distribution of the number of units purchased for each of the 12 weeks, details of total purchasing, and the number of new customers (triers) in each week. In estimating our model, we used no information beyond these aggregate numbers. (The simplicity of this data structure is an important contribution of our work.)

While this is a very convenient summary of the customers' purchasing, it suffers from two critical shortcomings: (1) We had no explicit information on the breakdown of trial versus repeat sales in each week, and (2) we could not see the longitudinal series of purchase events at the customer-level, which made it impossible to construct a standard model of repeat

Number of CDs purchased	Week 1	2	3	4	5	6	7	8	9	10	11	12
0		1,478	3,033	4,763	6,608	8,616	10,829	12,716	14,698	16,774	18,881	20,902
1	750	852	984	1,066	1,237	1,262	1,204	1,278	1,397	1,444	1,387	1,148
2	383	387	456	484	566	649	592	606	644	659	677	663
3	191	214	270	267	293	320	302	343	365	374	355	367
4	95	120	114	161	163	196	156	195	179	187	199	182
5	55	72	68	89	96	96	80	100	95	118	94	120
6	36	40	42	40	51	54	65	45	75	71	72	54
7	18	12	27	30	36	40	39	31	41	37	30	43
8	12	15	9	21	19	21	20	24	23	29	24	32
9	9	9	8	9	21	14	21	8	14	9	12	16
10+	25	17	27	32	36	55	39	35	48	42	50	43
Total sales	3,627	3,857	4,512	5,054	5,843	6,456	5,906	6,077	6,757	6,848	6,770	6,781
Incremental triers	1,574	1,642	1,822	1,924	2,164	2,197	2,024	2,034	2,198	2,165	2,037	1,789
Cumulative triers	1,574	3,216	5,038	6,962	9,126	11,323	13,347	15,381	17,579	19,744	21,781	23,570

Table 1: This table summarizes the total purchasing for the cohort of customers who made their first-ever purchases at CDNOW in the first quarter of 1997. Each column shows the distribution of the number of CDs purchased by the group of eligible customers in that week. For example, in week 2, of the 3,216 customers who could have made a purchase, 1,478 people purchased no CDs, 852 purchased one CD, and so on.

purchasing (that is, depth-of-repeat or counting). We therefore had to develop a model of week-by-week repeat purchasing whose parameters could be estimated using the above data.

General Model Structure

Our objective was to develop a simple stochastic model of buyer behavior capable of producing a medium-term forecast of CD purchases by the cohort of new customers whose total purchasing during the first quarter of 1997 we had summarized (Table 1). In week 2, 3,216 customers could have made a purchase: 1,642 of them made their first purchases at the CDNOW Web site in week 2 and 1,574 of them first purchased at it in week 1 and could therefore have returned for additional (repeat) purchases in week 2. By definition, the 1,642 week 2 triers must have purchased at least one unit. This implies that the 1,478 people who made no purchase at the Web site in week 2 must be customers who made a trial purchase in week 1. Thus we have 1574 − 1478 = 96 week 1 triers who made repeat purchases in week 2. In other words, this observed distribution of week 2 purchasing represents a mixture of purchases by those customers whose first purchase occured in week 2 and repeat purchases by those who tried in week 1. Therefore, the probability of observing someone purchasing x units in week 2 is simply a weighted average of the probability that a week 2 trier bought x units during her initial week, and the probability that a week 1 trier bought x units on at least one repeat purchase occasion in week 2. The weights are determined by the number of triers in weeks 1 and 2:

$$P(X_2 = x) = \frac{1{,}642}{1{,}574 + 1{,}642} \times P(T_2 = x) + \frac{1{,}574}{1{,}574 + 1{,}642} \times P(R_{2|1} = x)$$

where $P(T_2 = x)$ is the probability that a randomly chosen customer making her first purchase(s) at CDNOW in week 2 buys x units, and $P(R_{2|1} = x)$ is the probability that a randomly chosen customer who first purchased in week 1 purchases x units in week 2.

Similarly, in week 3, 5,038 customers could have made a purchase that week: 3,216 of these customers made their first purchase in weeks 1 or 2, and 3,033 of these people made no (repeat) purchase in week 3. We therefore have 183 week 1 and week 2 trialists making repeat purchases in this week, but we do not observe the specific number of week 1 versus week 2 triers nor each of these groups respective distribution of units purchased. Extending the same logic from above, however, we can express the probability of observing x purchases in week 3 as a weighted average of the probability that a week 3 trier made x purchases, the probability that a week 2 trier made x repeat purchases in week 3, and the probability that a week 1 trier made x repeat purchases in week 3:

$$\begin{aligned} P(X_3 = x) &= \frac{1{,}822}{1{,}574 + 1{,}642 + 1{,}822} \times P(T_3 = x) \\ &+ \frac{1{,}642}{1{,}574 + 1{,}642 + 1{,}822} \times P(R_{3|2} = x) \\ &+ \frac{1{,}574}{1{,}574 + 1{,}642 + 1{,}822} \times P(R_{3|1} = x) \end{aligned}$$

where the weights are determined by the

number of triers in weeks 1 through 3.

More generally, the distribution of purchases in week w can be modeled using a finite mixture model with known mixing weights:

$$P(X_w = x) = \frac{1}{\sum_{i=1}^{w} n_i} \times \left[n_w P(T_w = x) + \sum_{i=1}^{w-1} n_i P(R_{w|i} = x) \right] \quad (1)$$

where n_i is the number of triers in week i (that is, customers making their first purchases at the CDNOW Web site), $P(T_w = x)$ is the probability that a randomly chosen customer making her first purchase(s) at CDNOW in week w buys x units, and $P(R_{w|i} = x)$ is the probability that a randomly chosen customer who first purchased in week i buys x units in week w. We therefore need to develop submodels for $P(T_w = x)$ and $P(R_{w|i} = x)$.

Modeling Trial Purchases

Let the random variable T_w denote the number of units purchased in week w by a customer whose trial purchase occurs in week w. (By definition, T_w is a zero-truncated discrete random variable.) Our submodel for the distribution of T_w is based on the following two assumptions: (1) At the level of the individual customer, T_w is distributed according to a shifted geometric distribution with parameter q_T and probability mass function

$$P(T_w = x \mid q_T) = \begin{cases} q_T(1 - q_T)^{x-1}, & x = 1,2,\ldots; 0 < q_T < 1, \\ 0, & x = 0. \end{cases}$$

(2) q_T is distributed across the population according to a beta distribution with parameters α_T and β_T, and pdf

$$g(q_T) = \frac{1}{B(\alpha_T, \beta_T)} q_T^{\alpha_T - 1} (1 - q_T)^{\beta_T - 1},$$

$$0 < q_T < 1; \alpha_T, \beta_T > 0.$$

The intuition associated with these two assumptions is as follows. The geometric distribution corresponds to purchasing following a coin-flipping process in which the individual customer keeps buying until she tosses a head. The beta distribution is simply a means of allowing P(heads) to vary across the customer base.

It follows that the aggregate distribution of the number of units purchased by a week w trialist is given by

$$P(T_w = x) = \int_0^1 P(T_w = x \mid q_T) g(q_T) dq_T = \begin{cases} \dfrac{B(\alpha_T + 1, \beta_T + x - 1)}{B(\alpha_T + 1, \beta_T)}, & x = 1,2,\ldots \\ 0, & x = 0, \end{cases} \quad (2)$$

which we call the shifted beta-geometric distribution. Elsewhere in the marketing literature, this distribution was used by Morrison and Perry [1970] to model purchase quantity, conditional on purchase incidence. (We explore the validity of this distribution as a model of trial-week purchasing in the appendix.) The mean of this distribution is given by

$$E(T_w) = \frac{\alpha_T + \beta_T - 1}{\alpha_T - 1}. \quad (3)$$

Modeling Repeat Purchases

Let the random variable $R_{w|i}$ denote the number of (repeat) purchases made in week w by a customer who made her trial purchase in week i ($w > i$). Specifying an appropriate model for the distribution of

$R_{w|i}$ is the single most important step in this modeling effort. To do so, we start with the assertion that a new customer's purchasing at an established store (or Web site) is analogous to a consumer's purchasing a new product. We know that repeat-buying rates for new products tend to be nonstationary—at least early in a new product's life [Fader and Hardie 1999a]—with the purchase rate declining towards an equilibrium level over time. One way to capture this pattern is to assume that, for a given cohort, the number of people making zero purchases in a given week grows (at a decreasing rate), which means that the observed average number of units purchased decreases over time.

Our submodel for the distribution of $R_{w|i}$ is based on the following three assumptions:

(1) In week w, each customer is either out of the market (definitely not going to make a repeat purchase that week) or is a possible repeat buyer. The probability of a week i trialist being out of the market in week w is denoted by $\pi_{w|i}$. While such a person may be out of the market in week w, we are not assuming that she is permanently out of the market; she may consider buying again in future weeks. We assume that this probability is governed by the following time-dependent distribution:

$$\pi_{w|i} = 1 - \gamma(w - i)^{\delta}, \quad w > i.$$

When $\delta < 0$, $\pi_{w|i}$ grows at a decreasing rate as $w - i$ increases; consequently, the number of week i triers making zero purchases in week w increases over time. Likewise, δ can also be positive, allowing for the possibility that the number of repeat buyers actually increases over time. (The notion that someone is a possible repeat buyer does not ensure that she will actually purchase any units that week; it merely conveys the fact that she will *consider* purchasing with some nonzero probability.)

(2) For an individual who has been classified as a possible repeat buyer in week w, $R_{w|i}$ is distributed according to a geometric distribution with parameter q_R and probability mass function

$$P(R_{w|i} = x \mid q_R) = q_R(1 - q_R)^x,$$
$$x = 0,1,\ldots; \quad 0 < q_R < 1.$$

(3) q_R is distributed across the population according to a beta distribution with parameters α_R and β_R, and pdf

$$g(q_R) = \frac{1}{B(\alpha_R, \beta_R)} q_R^{\alpha_R - 1} (1 - q_R)^{\beta_R - 1},$$
$$0 < q_R < 1; \alpha_R, \beta_R > 0.$$

Qualitatively, the same type of coin-flipping story discussed earlier for the trial submodel applies here as well. However, there are two differences. First, there is no longer a truncation at zero; that is, the first coin flip determines whether a possible repeat buyer actually chooses to purchase one unit (or more). Second, the stopping probability (P(heads)) is governed by a different beta distribution than that used for the trial purchasing process.

It follows that the aggregate distribution of the number of units purchased in week w by a week i trialist ($w > i$) is given by:

$$\begin{aligned} P(R_{w|i} = x) &= \delta_{x=0}\pi_{w|i} \\ &+ (1 - \pi_{w|i}) \int_0^1 P(R_{w|i} = x \mid q_R) g(q_R)\, dq_R \\ &= \delta_{x=0}\pi_{w|i} + (1 - \pi_{w|i}) \\ &\times \frac{B(\alpha_R + 1, \beta_R + x)}{B(\alpha_R, \beta_R)} \end{aligned} \tag{4}$$

where $\delta_{x=0}$, the Kronecker delta, equals 1 if $x = 0$, 0 otherwise. We call this the time-dependent, zero-inflated beta-geometric distribution. The mean of this distribution is

$$E(R_{w|i}) = \gamma\,(w - i)^{\delta}\,\frac{\beta_R}{\alpha_R - 1}. \tag{5}$$

Parameter Estimation

Given the data (Table 1), we find the maximum likelihood estimates of the six model parameters (α_T, β_T, α_R, β_R, γ, δ) by maximizing the following log-likelihood function:

$$\begin{aligned} LL = {} & \sum_{x=1}^{9} m_{1x}\ln[P(T_1 = x)] \\ & + \left(m_{1\cdot} - \sum_{x=1}^{9} m_{1x}\right)\ln\left[1 - \sum_{x=1}^{9} P(T_1 = x)\right] \\ & + \sum_{w=2}^{12}\left\{\sum_{x=0}^{9} m_{wx}\ln[P(X_w = x)]\right. \\ & \left. + \left(m_{w\cdot} - \sum_{x=0}^{9} m_{wx}\right)\ln\left[1 - \sum_{x=0}^{9} P(X_w = x)\right]\right\} \end{aligned} \tag{6}$$

where m_{wx} is the number of people making x purchases in week w and $m_{w\cdot}$ is the total number of eligible customers in week w (the cumulative triers number from Table 1).

To evaluate the log-likelihood function, we must be able to compute $P(T_w = x)$ and $P(R_{w|i} = x)$, as given in equations (2) and (4). While it is feasible to employ these equations directly, we can achieve significant advantages in coding and estimating the model by using very simple recursive relationships that exist for both components of the model. For instance, $P(T_w = x)$ can be reexpressed as follows:

$$P(T_w = x) = \frac{P(T_w = x)}{P(T_w = x - 1)}P(T_w = x - 1).$$

Substituting (2) into the ratio $P(T_w = x)/P(T_w = x - 1)$, we find that many terms cancel out, including all the beta functions (which are quite inconvenient to evaluate in a standard spreadsheet). This leaves us with the much simpler expression:

$$P(T_w = x) = \begin{cases} 0, & x = 0, \\ \dfrac{\alpha_T}{\alpha_T + \beta_T}, & x = 1, \\ \dfrac{\beta_T + x - 2}{\alpha_T + \beta_T + x - 1}P(T_w = x - 1), & x \geq 2. \end{cases} \tag{7}$$

Similarly, we can compute the probabilities associated with the time-dependent, zero-inflated beta-geometric distribution (4) using the following forward-recursive relationship:

$$P(R_{w|i} = x) = \begin{cases} 1 - \gamma(w - i)^{\delta}\left(\dfrac{\beta_R}{\alpha_R + \beta_R}\right), & x = 0, \\ \gamma(w - i)^{\delta}\dfrac{\alpha_R\beta_R}{(\alpha_R + \beta_R)(\alpha_R + \beta_R + 1)}, & x = 1, \\ \dfrac{\beta_R + x - 1}{\alpha_R + \beta_R + x}P(R_{w|i} = x - 1), & x \geq 2, \end{cases} \tag{8}$$

Combining these simplified expressions back into (1) and then into (6) completes our description of the model as actually implemented. While the log-likelihood function (6) appears to be rather complicated—it involves the evaluation of 131 terms—each of these calculations is very simple, and we can actually construct this function in a spreadsheet quite easily using basic cut-and-paste techniques.

Generating a Sales Forecast

Let the random variable N_w be the total number of CDs purchased by the (eligible) cohort members in week w. Our best estimate of total sales in week w is given by

$$E(N_w) = \begin{cases} n_w E(T_w) + \sum_{i=1}^{w-1} n_i E(R_{w\mid i}), & w \leq 12, \\ \sum_{i=1}^{12} n_i E(R_{w\mid i}), & w > 12, \end{cases} \quad (9)$$

where $E(T_w)$ and $E(R_{w\mid i})$ are calculated using (3) and (5), respectively, and n_i is the number of people who made their trial purchase in week i. Given the maximum likelihood estimates of the model parameters, we use this equation to generate a week-by-week forecast of the total purchasing by the cohort.

Summary of Model Development

The objective of our modeling effort was to develop a simple model for forecasting medium-term CD purchasing by a cohort of customers who made their first purchases at CDNOW during the first 12 weeks of 1997. We used equation (9) to create such forecasts. Central to this are expressions for the number of CDs purchased on a trial occasion and the number of CDs purchased in week w, given trial in week i; expressions for these are given in (2) and (4), respectively. However, the nature of reported data is such that we do not observe these separate components of sales—we observe only the overall distribution of total purchases by all customers in a given week. Using the finite mixture model presented in (1), we can estimate the parameters of the submodels for trial and repeat purchasing using the aggregated data and then use them to create the sales forecast.

Readers familiar with the marketing literature on stochastic models of buyer behavior may wonder why we used the beta-geometric distribution as the underlying counting distribution, instead of the more common NBD (negative binomial distribution) that is widely used within marketing [Morrison and Schmittlein 1988]. The primary reason is one of communication. Successful implementation requires management's acceptance of the underlying model. The chance of managers accepting a model increases with their ability to understand the workings of the model, at the very least at an intuitive level. The logic of the beta-geometric distribution is much easier to communicate to a managerial audience, using the coin-flipping analogy, than the rationale for the NBD model (Poisson purchasing at the individual level with gamma heterogeneity). Two secondary reasons for using the beta-geometric distribution as opposed to the NBD are (1) the ease of handling the zero-truncated distribution for trial purchasing, and (2) a marginally better fit associated with the beta-geometric distribution.

Applying the Model

Given the data set (Table 1), we implemented the above model entirely within the Microsoft Excel spreadsheet package. (A copy of the spreadsheet, along with a note on how to build it, can be found at ⟨http://brucehardie.com/pmnotes.html⟩.) We obtained parameter estimates using the Excel add-in known as Solver to maximize the log-likelihood function given in (6). The estimation procedure is extremely fast (requiring only a few seconds on a standard PC) and highly robust (always converging very close to the global optimum). The maximum value of the log-likelihood function is $LL = -112{,}923.9$, which occurs when the model parameters take on the following values:

α_T	β_T	α_R	β_R	γ	δ
6.901	7.185	5.024	5.595	0.122	-0.291

Using (3), we found that the mean number of CDs purchased during a new customer's trial week is 2.22 units. For any customer who is a possible repeat buyer in a given week, the expected number of CDs purchased computed using (5) is 1.39 units. Being classified as a possible repeat buyer by no means implies that the person actually makes a purchase—there is a 47-percent chance that such a person makes no purchase at all. The fact that $\hat{\delta} < 0$ implies that $\pi_{w|i}$ increases over time (at a decreasing rate). Consequently, the number of buyers who will definitely not make a repeat purchase in a given week grows as we move further from their trial week, and therefore the observed number of repeat purchases will decline with time.

The predicted distribution of weekly purchases, obtained using (1), produces a good fit to the observed data (Table 1), as judged visually (Figure 1) and by a chi-square goodness-of-fit test ($\chi^2_{113} = 129.21$, $p = 0.141$).

The estimates of total sales, computed using (9), track observed total sales (Table 1) very well. The model also allows us to decompose these total sales figures into separate trial and repeat components. Comparing these estimates to the actual trial and repeat numbers determined using the original (disaggregate) data set

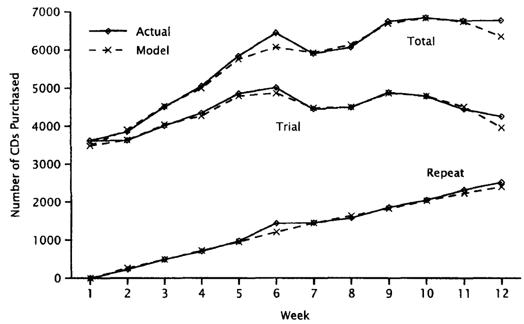

Figure 1: The estimates of total weekly sales generated by our model are compared to the actual sales data for the first 12 weeks of 1997. Also, model-based estimates of the trial and repeat components of weekly sales are compared to the actual trial and repeat sales data. The ability of these estimates to track total sales and its components provides support for the validity of our model.

shows that the model recovers these underlying components (which were not separately identified in Table 1) very well. This provides further support for the validity of our model and a high degree of confidence that we can extrapolate beyond the 12-week calibration period.

The repeat sales numbers appear to grow in a linear manner over time as a steady number of new triers enters the market in each week. Because the number of new triers drops to zero after week 12—we are forecasting sales only for the cohort of customers who made their first-ever purchases at CDNOW in the first 12 weeks of 1997—we might expect this curve to stop growing, perhaps remaining close to its final level of approximately 2,500 units per week for the entire cohort. (This would be equivalent to each repeater buying on average one CD every 10 weeks.) We call this the linear-projection forecast. The linear-projection forecast is consistent with and perhaps better than other back-

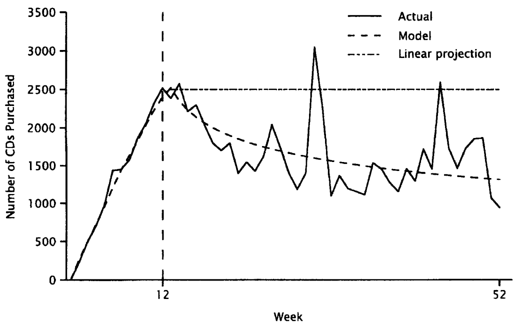

Figure 2: The weekly repeat-sales forecasts generated by our model and the linear projection are compared to the actual sales data for the cohort through the end of 1997. In contrast to the linear projection, our model provides a far more realistic picture of repeat purchasing for the 40 weeks beyond the 12-week calibration period.

of-the-envelope forecasts one might make based on the first-quarter sales data.

For this cohort of 23,570 people who first purchased in quarter 1 of 1997, we extracted records of their total purchasing for the 40 weeks beyond the calibration period (that is, April through December 1997). We evaluated the forecasting performance of the model against these actual purchasing numbers. Given the parameter estimates, we computed the aggregate sales forecast using (9). Our model provided a much more realistic picture of future repeat purchasing than the linear projection (Figure 2). In sharp contrast to the linear projection, our model predicts a decreasing level of repeat purchasing by this cohort as it ages, and the number of possible repeat buyers in a given week shrinks over time.

The first observed major deviation from our forecast corresponds to a midyear promotion CDNOW ran, while the second spike corresponds to the Christmas season. The model's projections seem to serve as an accurate and potentially valuable benchmark for understanding what expected sales levels would have been in the absence of these special events.

The performance of our model is further demonstrated by the same data in cumulative form (Figure 3). At the end of the year (week 52), the forecast index (relative to actual) for our model is 98.7 percent, while the index associated with the linear projection is 140.7 percent. The underprediction associated with our model should come as no surprise because the actual sales numbers contain promotional and seasonal events not captured within our model.

Reexamining Figure 2, we may be tempted to assert that post-trial sales did at first decline but then showed a slow increase; perhaps the sales from weeks 13 to 52 could be represented by a U-shaped curve. If this were the case, the model would underpredict future sales, as the expected sales curve will continue to decline (albeit at a decreasing rate). We obtained the cohort's purchasing data for the first

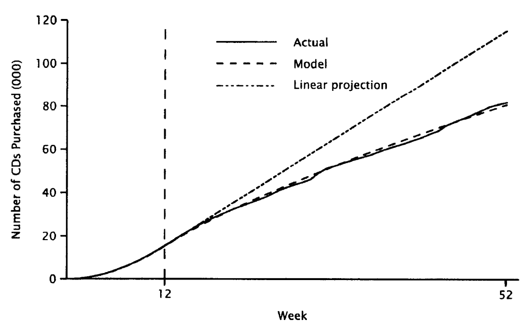

Figure 3: The performance of our model is clearly demonstrated when the weekly sales data from Figure 2 are presented in cumulative form. Our model underpredicts the cohort's first year purchasing by less than two percent. In contrast, the linear projection overpredicts the cohort's first-year purchasing by more than 40 percent.

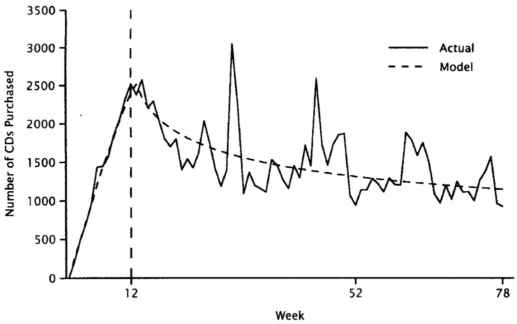

Figure 4: The purchasing data for the cohort through the end of the first half of 1998 show that actual purchasing continued to follow the underlying trend predicted by our model.

six months of 1998 and extended our forecast to include that time period. Our model continues to capture the underlying trend in repeat-purchase behavior, albeit with obvious deviations because of promotional activities (Figure 4). The ability of a simple six-parameter model, calibrated on 12 weeks of data, to forecast the underlying trend of purchasing 66 weeks into the future is quite remarkable.

Looking Ahead

The medium-term sales forecast our model provides is quite precise; managers can use this tool to determine the overall value of a customer base or to estimate the incremental sales associated with their firm's marketing activities.

In thinking about extensions to the modeling framework we developed, an obvious next step would be to apply this model to other cohorts (for example, customers making their first purchases at CDNOW in the second quarter of 1997) and to examine the stability of the associated model parameters. Furthermore, it would be useful to develop a model for the arrival of new customers to the Web site. Coupling these two models would enable us to forecast a site's *overall* sales (as opposed to those for a given cohort).

The model we developed does not take full advantage of the richness of the individual-level transaction data that the customer purchase histories give us. A second focus for future research would be to model more formally the separate components of purchasing and the dynamics of buyer behavior, using the disaggregate panel data. Our model focuses on units purchased per week, which arise from the combination of two separate processes—the number of transactions an individual makes in a given week and the number of units purchased during each transaction. A more sophisticated model would explicitly recognize this decomposition of total purchasing. Second, the model treats the data as a series of cross sections. We do not model the longitudinal series of purchase events at the level of the individual customer. Consequently, we are unable to make customer-level predictions of future behavior or to profile individual customers—activities that are vital to many database marketing efforts. In future work, transactions could be modeled using an individual-level counting process (for example, number of transactions across unit time intervals) or an intertransaction timing model.

Finally, we should consider the $\pi_{w|i}$ term, the probability of a week-i trialist being out of the market in week w. While this captures the nonstationarity in buying, as evident through the decline in repeat purchasing, it cannot tell us whether this decline is caused by a slowdown among active repeat customers, customers dropping out of the market, or both. A model that captures nonstationarity at the individual level, such as Fader and Hardie's [1999a] NSEG model, would provide such insights.

While such extensions would lead to more "correct" models of buying behavior, they would have costs. The resulting models would be quite complex and would have to be calibrated using customer-level data, as opposed to the summary of the data used by our model (Table 1). Furthermore, the analyst would need fairly sophisticated modeling skills and access to appropriate computational software. These two factors, combined with the difficulty of explaining the models to managers, would impede their implementation, especially compared to a simple model, such as that we developed. It would be useful to compare the aggregate forecasting performance of such models to that of our far simpler model, which captures the sales patterns but cannot properly diagnose the causes of the observed behavior. Based on a similar comparison undertaken for new-product sales forecasting models [Fader and Hardie 1999b], we expect that the aggregate forecasting performance of our model would be on par with that of any more complex model.

Furthermore, in introducing marketing models to an organization, it is also good to start with simple models and evolve towards more complete (and complex) models as the key personnel become more comfortable with making use of marketing models and more willing to commit the resources that the more complex models require [Urban and Karash 1971].

Acknowledgments

We thank the editors and reviewers for an encouraging set of comments on an earlier draft of this paper.

APPENDIX

Validating the Trial Purchasing Model

The overall fit of the model and its ability to decompose trial and repeat sales from the aggregate data suggest that the assumptions underlying our model are reasonable. However, this conclusion is based on an analysis that uses a mixture of beta-geometric distributions, which may not be valid. In this appendix, we explore the validity of the beta-geometric distribution itself.

In Table 1, the column corresponding to week 1 presents trial-week-only purchases by a group of 1,574 customers. Using these data, we can examine the assumptions underlying the beta-geometric model or, more correctly, the shifted beta-geometric model (to acknowledge the truncation at zero) associated with trial purchases, as given in (2).

In modeling trial, we may first be tempted to fit a homogeneous shifted-geometric distribution to the trial-purchases data; that is, to assume all customers have the same value of the latent trial-purchasing-propensity parameter, q_T. Fitting the shifted-geometric distribution to the week-1 purchase data, the maximum likelihood estimate of q_T is 0.444. On the basis of the chi-square goodness-of-fit test, the fit of this model to the actual purchase data for week 1 is poor ($\chi^2_8 = 51.8$, $p < 0.001$).

A potential cause for the poor fit of this model is that it ignores differences in people's propensity to make multiple purchases (q_T). Under the assumption of beta heterogeneity, we have the shifted beta-geometric distribution of trial-week purchase quantities. Fitting (2) to the week-1 purchase data, the maximum likelihood estimates of α_T and β_T are 5.912 and 6.283, respectively. The fit of this model to the actual purchase data for week 1 is now excellent, as judged visually (Figure 5) and

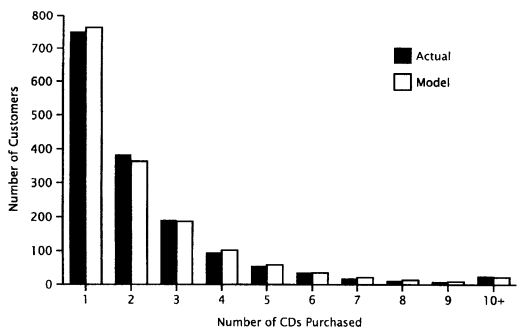

Figure 5: The validity of the shifted beta-geometric distribution, which is used to model trial-week purchase quantity, is examined by fitting it to the week 1 purchase data. The week 1 data represent "clean," trial-only purchases. The excellent fit of the model to the actual distribution of the number of CDs purchased provides evidence for the appropriateness of this submodel.

on the basis of a chi-square goodness-of-fit test ($\chi^2_7 = 3.3$, $p = 0.86$). Using (3), we see that $E(T_1) = 2.28$. This implies that the expected number of total purchases in week 1, $E(N_1)$, is 3,587, which is within 1.1 percent of the actual number of units purchased (Table 1), thus providing further

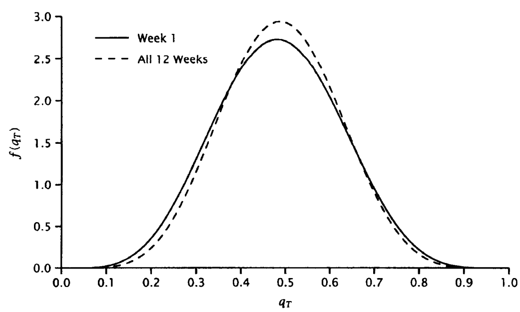

Figure 6: The parameter q_T captures the propensity of a customer to make multiple purchases on her first purchase occasion at CDNOW. The underlying distribution of q_T across the population as estimated using all 12 weeks worth of data is compared to that estimated using just the week 1 data. We observe that the differences between the two estimated distributions of q_T are negligible, which provides support for the assumptions underlying our model.

evidence of the validity of the beta-geometric distribution.

We can compare the underlying distribution of q_T estimated using the week-1 data ($\hat{\alpha}_T = 5.912$ and $\hat{\beta}_T = 6.283$) with that derived using all 12 weeks worth of data ($\hat{\alpha}_T = 6.901$ and $\hat{\beta}_T = 7.185$). The 12-week estimate is derived using the mixture of trial and repeat purchasing distributions, that is, equation (1); it is not based on "clean," trial-only data. The mean of the distribution of q_T derived using all of the data is slightly higher than that associated with the week-1-only data (0.490 vs. 0.485), and the corresponding variance is slightly lower (0.017 vs. 0.019). However, the differences are negligible (Figure 6). This similarity provides more support for the assumptions underlying our basic model.

References

Allenby, Greg M.; Leone, Robert P.; and Jen, Lichung 1999, "A dynamic model of purchase timing with application to direct marketing," *Journal of the American Statistical Association*, Vol. 94, No. 446, pp. 365–374.

Buchanan, Robert W., Jr. and Lukaszewski, Charles 1997, *Measuring the Impact of Your Web Site*, John Wiley and Sons, New York.

Colombo, Richard and Jiang, Weina 1999, "A stochastic RFM model," *Journal of Interactive Marketing*, Vol. 13, No. 3, pp. 2–12.

Fader, Peter S. and Hardie, Bruce G. S. 1999a, "Modeling the evolution of repeat buying," Working paper, Marketing Department, The Wharton School, University of Pennsylvania.

Fader, Peter S. and Hardie, Bruce G. S. 1999b, "Investigating the properties of the Eskin/Kalwani & Silk model of repeat buying for new products," in *Marketing and Competition in the Information Age*, Proceedings of the 28th EMAC Conference, eds. Lutz Hildebrandt, Dirk Annacker, and Daniel Klapper, Humboldt University, Berlin.

Forrester Report 1999, *Measuring Web Success*, November, Forrester Research, Inc., Cambridge, Massachusetts.

Morrison, Donald G. and Perry, Arnon 1970, "Some data based models for analyzing sales fluctuations," *Decision Sciences*, Vol. 1, Nos. 3

and 4, pp. 258–274.
Morrison, Donald G. and Schmittlein, David C. 1988, "Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort?" *Journal of Business and Economic Statistics*, Vol. 6, No. 2, pp. 145–159.
Mulhern, Francis J. 1999, "Customer profitability analysis: Measurement, concentration, and research directions," *Journal of Interactive Marketing*, Vol. 13, No. 1, pp. 25–40.
Schmittlein, David C.; Morrison, Donald G.; and Colombo, Richard 1987, "Counting your customers: Who are they and what will they do next?" *Management Science*, Vol. 33, No. 1, pp. 1–24.
Schmittlein, David C. and Peterson, Robert A. 1994, "Customer base analysis: An industrial purchase process application," *Marketing Science*, Vol. 13, No. 1, pp. 41–67.
Urban, Glen L. and Karash, Richard 1971, "Evolutionary model building," *Journal of Marketing Research*, Vol. 8, No. 1, pp. 62–66.