SUNIL GUPTA*

Consumers' purchase time decisions are important elements of their buying decision process. Stochastic models of interpurchase time, which have been used extensively in the marketing literature, are parsimonious, easy to estimate, and usually fit and predict the data well. However, there has been a striking omission of marketing variables in these models. Because empirical evidence suggests that marketing variables, such as promotions, can affect consumers' purchase time decisions, the author presents a methodology for including such variables in these stochastic models. Four commonly used models are discussed: exponential, Erlang-2 (no heterogeneity), and these two models with gamma heterogeneity. Thus one can include duration dependence, heterogeneity, and nonstationarity in the model, and also account for right-censored data. Special care is shown to be needed when covariates, such as marketing variables, vary over time. The models are analytically tractable, which makes their estimation and validation simple and fast. An illustration of the methodology is provided with scanner panel data for coffee. Inclusion of duration dependence, heterogeneity, and marketing variables is shown to improve the model's diagnostics, fit, and predictions.

Stochastic Models of Interpurchase Time With Time-Dependent Covariates

Understanding, explaining, and predicting consumers' buying behavior and how it is affected by marketing variables are key objectives of marketing researchers and practitioners. Though many studies have addressed the issue of how marketing variables affect consumers' brand choice decisions (e.g., Guadagni and Little 1983; Gupta 1988; Jones and Zufryden 1980), few have examined their effect on consumers' purchase time decisions despite a long history of stochastic models being used to represent consumers' interpurchase times (e.g. Ehrenberg 1959, 1972; Morrison and Schmittlein 1981, 1988; Zufryden 1977, 1978).

A quick survey of the literature shows that four stochastic models typically have been used in marketing: exponential and Erlang-2 at the individual-consumer level, and these two distributions with gamma heterogeneity to model aggregate-level behavior (Chatfield and Goodhardt 1973; Morrison and Schmittlein 1988; Schmittlein, Morrison, and Colombo 1987). For more than three decades, these models have been applied extensively in the marketing literature because they are parsimonious, easy to estimate, and usually fit and predict the data well. However, one limitation has consistently plagued these models—the omission of marketing variables. Because empirical evidence suggests that marketing variables, such as promotions, can accelerate purchases in time (Gupta 1988; Shoemaker 1979; Ward and Davis 1978), inclusion of such variables in the stochastic interpurchase time models would be useful for not only predictive but also diagnostic purposes. This fact has generally been recognized and discussed and, given the extensive use of the aforementioned stochastic models in marketing, the incorporation of marketing variables in these models seems long overdue. In a recent article, Morrison and Schmittlein (1988, p. 165) note, "Probably the most important area for future research with [the negative binomial distribution or] NBD-type models involves incorporating the effect of marketing variables." The purpose of this article is to present a methodology for achieving that objective.

*Sunil Gupta is Assistant Professor of Marketing, Graduate School of Business, Columbia University.
The author thanks Barbara Kahn, John Little, and Don Morrison for their helpful comments.

Journal of Marketing Research
Vol. XXVIII (February 1991), 1–15

ALTERNATIVE APPROACHES

Though not directly extending traditionally used stochastic models, some recent studies have suggested alternative approaches to capture the effect of marketing variables on consumers' purchase time decisions. The following brief discussion of these approaches along with their advantages and limitations puts the proposed approach in perspective.

Linear Regression

The first approach, exemplified by Neslin, Henderson, and Quelch (1985), uses a simple linear regression to model interpurchase time as a function of explanatory variables. Though it is easy to understand and estimate, Helsen and Schmittlein (1989) show that this approach may produce biased estimates in the presence of right-censored data and time-dependent covariates, both of which are very common in most marketing applications.

Modeling the Buy/Not Buy Decision

The second approach models the binary outcome of purchase or no purchase during the interval of interest (e.g., a week). Logit or probit models have been used to model these binary outcomes (Bucklin and Lattin 1990; Guadagni and Little 1987). Because most marketing data (e.g., scanner data) now have the exact time of purchase by a consumer, modeling the binary outcome of purchase or no purchase amounts to loss of information. However, Wheat and Morrison (1990a) show some advantages for predicting the buy/not buy event rather than the interpurchase time interval. A second difficulty in buy/not buy models is the incorporation of unobserved heterogeneity. Ignoring unobserved heterogeneity is conceptually similar to the omitted-variable problem, which may lead to biased estimates of the covariates (Heckman and Singer 1984a; Jain and Vilcassim 1990).

Partial Likelihood Approach

A third way to model duration or interpurchase time is by the proportional hazard model and the partial likelihood approach (Cox 1972, 1975; Helsen and Schmittlein 1989). This approach is appealing as it does not require the specification of the underlying purchase time distribution. Two points must be mentioned about this approach. First, it does not use all the information. For example, if a consumer buys a product after k weeks, this approach uses the covariate values of the k^{th} week only and ignores the covariates of $(k - 1)$ weeks. In other words, the information about whether or not there was a big promotion in the previous $(k - 1)$ weeks is completely ignored. Many studies have shown that in most situations the parameters obtained from the partial likelihood approach are consistent and efficient (Efron 1977; Oakes 1977). However, if there are strong time trends in covariates, the loss of information can severely affect parameter efficiency (Cox and Oakes 1984). Second, it is difficult to incorporate unobserved heterogeneity in the partial likelihood approach. As pointed out before, ignoring unobserved heterogeneity can lead to biased estimates.

Parametric Hazard Model

Related to the partial likelihood approach is the hazard model approach whereby the underlying distribution is specified—that is, the parametric hazard model approach (Flinn and Heckman 1982; Jain and Vilcassim 1990). Here, hazard rate at time t is expressed as $h(t) = h_0(t) \exp(\beta \mathbf{X}_t)$, where $h_0(t)$ is the baseline hazard rate with a prespecified distribution (e.g., exponential) and $\mathbf{X}_t$ are the covariates. Jain and Vilcassim use a Box-Cox formulation to parameterize the baseline hazard rate. Further, to include unobserved heterogeneity, an additional component $c\theta$ is introduced so that $h(t) = h_0(t) \exp(\beta \mathbf{X}_t + c\theta)$. A distribution of θ across households captures the unobserved heterogeneity.

Proposed Approach: Extension of NBD-Type Models

The proposed approach is closest in spirit to the parametric hazard model approach in that one allows the parameter of the underlying distribution to be a function of covariates. However, it differs in the way heterogeneity is incorporated in the model. One uses the approach typically used in marketing: let an individual consumer's interpurchase time be distributed exponentially or Erlang-2, and let the heterogeneity be captured by a gamma distribution over one of the parameters of these distributions. This approach has three key advantages over the parametric hazard model approach. First, it provides a natural extension of the NBD-type models employed and found useful in the marketing literature (Ehrenberg 1959; Herniter 1971; Jeuland, Bass, and Wright 1980; Morrison and Schmittlein 1981, 1988; Schmittlein, Bemmaor, and Morrison 1985; Zufryden 1978). In other words, one need not abandon the very rich and useful approach applied in marketing over the last three decades. A comment by Sabavala (1988, p. 162) stresses this point:

> To make precise conditional predictions of customer-level and market-level behavior in order to decide on marketing actions ultimately will require that the model incorporate effects caused by covariates (product and customer characteristics or time-varying market variables). . . . There may be room for research on adaptive estimation methods (time-varying r and α), or Bayesian methods, which do not appear to have been developed for NBD. . . .

A second advantage is the ease of interpreting the degree of heterogeneity in the data. Because a gamma mixing distribution is used (with parameters r and α), the value of r is a direct measure of the extent of heterogeneity in consumers' purchase rate. The higher the r value, the more homogeneous are the consumers. If the data indicate great heterogeneity in consumers' purchase rates, it may be important to segment the market before interpreting parameter estimates. Some studies have shown that heavy buyers behave differently from light buyers—for example, that heavy buyers are generally more price sensitive than light buyers (Gupta 1987).

A third advantage of the approach is that it provides simple, closed-form expressions for the likelihood function of the interpurchase time distributions, which are useful for estimation and prediction. The analytically tractable expressions make the estimation simple and fast. For example, it took less than five CPU minutes to estimate the model. Though computers are becoming more powerful, datasets are also becoming larger and hence this simpler and faster approach should have some appeal to both researchers and practitioners.

In summary, the proposed methodology incorporates time-dependent covariates in stochastic interpurchase time models, captures unobserved heterogeneity, and explicitly accounts for right-censored data. Inclusion of these covariates will help to show whether marketing variables, such as price and promotions, accelerate consumers' purchases in time. As the ensuing discussion explains, the inclusion of covariates that are time-dependent requires special care because the parameters of the distribution change from time period to time period (Cox and Oakes 1984).[1]

STOCHASTIC INTERPURCHASE TIME MODELS WITHOUT COVARIATES

Let us start with a brief review of the commonly used stochastic models of interpurchase time without including covariates. These models also serve as null models for comparison with models that include covariates.

The simplest model assumes a random purchasing process for an individual[2] i, leading to a Poisson distribution for the number of purchases in a fixed period of time or, equivalently, exponentially distributed interpurchase times with the density and survivor functions as

$$f_i(t) = \lambda_i \exp(-\lambda_i t) \tag{1}$$

$$S_i(t) = P_i(T > t) = \exp(-\lambda_i t). \tag{2}$$

Let n_i be the number of complete observations for consumer i. In addition, the last observation of the consumer usually will be incomplete or censored. The likelihood function for consumer i therefore can be written as

$$L_i(t|\lambda_i) = \left[\prod_{j=1}^{n_i} \lambda_i \exp(-\lambda_i t_{ij})\right] \exp(-\lambda_i t_{ic}) = \lambda_i^{n_i} \exp[-\lambda_i(t_{is} + t_{ic})], \tag{3}$$

where:

$$t_{is} = \sum_{j=1}^{n_i} t_{ij} \tag{4}$$

$=$ sum of n_i interpurchase times of consumer i, and

t_{ic} = censored time for consumer i.

If purchase rate is assumed to be the same for all consumers (i.e., $\lambda_i = \lambda \ \forall i$), then we have a homogeneous exponential distribution with the log likelihood for N consumers as

$$LL = \sum_{i=1}^{N} n_i \log \lambda - \lambda(t_{is} + t_{ic}). \tag{5}$$

This log likelihood can be maximized to estimate parameter λ.

The assumption that all consumers have the same purchase rate is clearly very restrictive and unlikely to hold in reality because some consumers buy the product more frequently than others. Hence heterogeneity in purchase rates is incorporated by allowing λ to have a gamma distribution with parameters r and α. The gamma distribution is appealing because of its flexibility and mathematical tractability. The likelihood function for a typical consumer now becomes

$$L_i(t) = \int_0^{\infty} L_i(t|\lambda_i)g(\lambda_i)d\lambda_i, \tag{6}$$

where $L_i(t|\lambda_i)$ is the conditional likelihood given in equation 3, $g(\lambda_i)$ is the density function of the gamma mixing distribution with parameters r and α, and $L_i(t)$ is the unconditional likelihood function for consumer i. The log likelihood for N consumers in the sample therefore can be written as

$$LL = \sum_{i=1}^{N} \left[r \log \alpha + \sum_{j=0}^{n_i-1} \log(r + j) - (n_i + r) \log (t_{is} + t_{ic} + \alpha) \right]. \tag{7}$$

The derivation for equation 7 is in the Appendix. Again, we can maximize the log likelihood to estimate parameters r and α. The r parameter is a direct indicator of the degree of heterogeneity in the consumers. The lower the r, the higher the heterogeneity.

If the sample consists only of cross-sectional data (instead of typical panel data that are both cross-sectional and time-series), we have only one observation per consumer. If it is a complete observation for consumer i, we substitute $n_i = 1$ and $t_{ic} = 0$ in equation 7 to obtain the log likelihood for consumer i as

$$LL_i = r \log \alpha + \log r - (r + 1) \log (t + \alpha).$$

Or the likelihood for consumer i can be written as

$$L_i = \frac{r}{\alpha}\left(\frac{\alpha}{t + \alpha}\right)^{r+1}$$

---

[1]Because of the special care necessary to incorporate time-dependent covariates, most commonly available statistical packages do not allow covariates to vary over time.

[2]The empirical analysis is done with household-level data. For the purpose of this article, the terms "individual," "consumer," and "household" are used interchangeably.

The right side of this equation is the density function of Pareto distribution (Schmittlein, Morrison, and Colombo 1987). In other words, if we have only one observation per consumer, we can use the density and survivor function of a Pareto distribution to write the likelihood function. However, in panel data we typically have multiple observations per consumer. In such a case, using the density function of Pareto distribution to write the likelihood function implies that we are capturing heterogeneity across *observations* instead of heterogeneity across *consumers*. The likelihood given in equation 7 ensures that we capture heterogeneity across consumers even when they have multiple observations.

In the preceding models, we assume that interpurchase times of a consumer follow an exponential distribution. Many researchers have questioned this assumption because of the *memoryless* property of the exponential distribution, which implies that a consumer's probability of purchasing the product in the next day or week is unaffected by the length of the time since his or her last purchase. Because this assumption is intuitively unappealing, many researchers have suggested a more *regular* purchase process than that suggested by the exponential distribution. A commonly used distribution is Erlang-2, for which many researchers have found at least partial support (Gupta 1988; Herniter 1971; Jeuland, Bass, and Wright 1980). Recently, Wheat and Morrison (1990b) examined various methods for discriminating between the exponential and Erlang-2 models at the individual level. The density and survivor functions of Erlang-2 are given by

$$f_i(t) = \lambda_i^2 t \exp(-\lambda_i t) \tag{8}$$

$$S_i(t) = (1 + \lambda_i t) \exp(-\lambda_i t). \tag{9}$$

Therefore, the likelihood function for consumer i is

$$\begin{aligned} L_i(t|\lambda_i) &= \left[\prod_{j=1}^{n_i} \lambda_i^2 t_{ij} \exp(-\lambda_i t_{ij})\right] (1 + \lambda_i t_{ic}) \exp(-\lambda_i t_{ic}) \\ &= \lambda_i^{2n_i} t_{ip} \exp[-\lambda_i(t_{is} + t_{ic})]\,(1 + \lambda_i t_{ic}), \end{aligned} \tag{10}$$

where $t_{ip} = \prod_{j=1}^{n_i} t_{ij}$ = product of interpurchase times for consumer i.

If all consumers have the same purchase rate—that is, we have a homogeneous population—the log likelihood for the sample can be written as

$$\begin{aligned} LL = \sum_{i=1}^{N} & 2n_i \log \lambda + \log t_{ip} \\ & - \lambda(t_{is} + t_{ic}) + \log(1 + \lambda t_{ic}). \end{aligned} \tag{11}$$

Again, if we allow for heterogeneity in purchase rates as per gamma distribution, we obtain the mixture Erlang-2/gamma. The log likelihood for this model is in Table 1 and its derivation is in the Appendix.

The top half of Table 1 summarizes the log likelihood for the four basic models: exponential and Erlang-2 with no heterogeneity, and those two models with gamma heterogeneity. This brief review provides background for the next section, in which we incorporate time-dependent covariates in these models.

STOCHASTIC INTERPURCHASE TIME MODELS WITH COVARIATES

Exponential Distribution

Let us begin with the simplest case, in which a consumer's interpurchase time is distributed exponentially. In the absence of covariates, the density and survivor functions are given by equations 1 and 2. To incorporate covariates, we allow the consumer's purchase rate λ_i to be a function of these covariates as

$$\lambda_i(t) = \lambda_{i0} \exp(\beta \mathbf{X}_{it}), \tag{12}$$

where $\mathbf{X}_{it}$ are time-dependent covariates, β are the response coefficients, and λ_{i0} is the base purchase rate (i.e. in the absence of covariates, $\lambda_i(t) = \lambda_{i0}$).[3] This equation gives a nonhomogeneous Poisson process—that is, times between purchases are neither independent nor identically distributed—and the process is not time homogeneous unless $\lambda_i(t) = \lambda_i$ (Lawless 1982). All consumers are assumed to have the same response parameters β. As a further simplification, we assume a homogeneous population in which all consumers have the same base purchase rate λ_0. This assumption is relaxed subsequently. It provides a benchmark to show the importance of including unobserved heterogeneity in the model.

The interpretation of the β coefficients is straightforward. A 1% change in a covariate X_l changes the purchase rate by $\beta_l X_l$ percent. An increase in purchase rate is equivalent to a decrease in the interpurchase time. If X_l is a promotion variable, we would expect β_l to be positive, indicating increasing purchase rate and hence early purchase due to promotion.

If the purchase rate is time-dependent, as in equation 12, then the purchase process is a nonhomogeneous Poisson process. For such a process, the probability of n purchases in a time period $0 - t$ has been shown (Ross 1980, p. 187) to be

$$P_n = \frac{[\theta(t)]^n}{n!} \exp[-\theta(t)], \tag{13}$$

where:

$$\theta(t) = \int_0^t \lambda(\tau)\, d\tau \tag{14}$$

(subscript i is dropped for ease of exposition).

Note that for a homogeneous Poisson process $\lambda(t) = \lambda$, and $\theta(t) = \lambda t$, so that equation 13 reduces to the fa-

[3]We use the exponential function in equation 12 to ensure that $\lambda(t)$ is always positive. Because the hazard rate for the exponential distribution is λ, this formulation is identical to the hazard model approach.

Table 1
LOG LIKELIHOOD FOR INTERPURCHASE TIME MODELS

Model	*Log likelihood for consumer*[a]
Without covariates	
1. Exponential[b]	$n_i \log \lambda - \lambda(t_{is} + t_{ic})$
2. Erlang-2	$2n_i \log \lambda + \log t_{ip} - \lambda(t_{is} + t_{ic}) + \log(1 + \lambda t_{ic})$
3. Exp/gamma[c]	$r \log \alpha + \Sigma_{j=0}^{n_i-1} \log(r + j) - (n_i + r) \log(t_{is} + t_{ic} + \alpha)$
4. Erlang-2/gamma	$r \log \alpha + \log t_{ip} + \Sigma_{j=0}^{2n_i-1} \log(r + j) + \log\left[1 + \frac{(2n_i + r)t_{ic}}{t_{is} + t_{ic} + \alpha}\right] - (2n_i + r) \log(t_{is} + t_{ic} + \alpha)$
With covariates[d]	
5. Exponential[e]	$n_i \log \lambda_0 + \log A_{ipk} - \lambda_0(B_{isk} + B_{ick})$
6. Erlang-2	$2n_i \log \lambda_0 + \log A_{ipk} + \log B_{ipk} - \lambda_0(B_{isk} + B_{ick}) + \log(1 + \lambda_0 B_{ick})$
7. Exp/gamma	$r \log \alpha + \Sigma_{j=0}^{n_i-1} \log(r + j) + \log A_{ipk} - (n_i + r) \log(B_{isk} + B_{ick} + \alpha)$
8. Erlang-2/gamma	$r \log \alpha + \log A_{ipk} + \log B_{ipk} + \Sigma_{j=0}^{2n_i-1} \log \cdot (r + j) + \log\left[1 + \frac{(2n_i+r)B_{ick}}{B_{isk} + B_{ick} + \alpha}\right] - (2n_i + r) \log(B_{isk} + B_{ick} + \alpha)$

[a]Log likelihood for N consumers is simply the log likelihood of consumer i summed over $i = 1$ to N.
[b]λ = purchase rate, t_{ic} = censored time, t_{is} = sum of n_i purchase times for consumer i.
[c]r and α are parameters of the gamma mixing distribution.
[d]In these models, the subscript k indicates the week in which the current purchase is made, given the week of last purchase is designated as week 1, ($k \geq 1$).
[e]$A_{ipk} = \Pi_j A_{ijk} = \Pi_j \exp(\beta \mathbf{X}_{ijk})$, where $\mathbf{X}$ are covariates, B_{ijk} as defined in equation 20, $B_{isk} = \Sigma_j B_{ijk}$, and $B_{ipk} = \Pi_j B_{ijk}$. If there are no covariates, $\beta\mathbf{X} = 0$ and hence $A_{ijk} = 1$ and $B_{ijk} = t_{ij}$.

miliar Poisson expression. Using equation 13, we can easily see that the probability of zero purchases in time $0 - t$, which is also by definition the survivor function for interpurchase time, is

$$P_0 = S(t) = \exp[-\theta(t)]. \tag{15}$$

Therefore the density function for interpurchase time with time-dependent covariates is given by:

$$f(t) = -\frac{dS(t)}{dt} = \lambda(t) \exp[-\theta(t)]. \tag{16}$$

Equation 14 for $\theta(t)$ can be simplified further by recognizing that in almost all marketing applications the covariate values remain constant for a time interval—for example, a week.[4] In other words, covariate values may change from week to week, but are time-invariant within a week. Notice that $\theta(t)$ will have more than one $\lambda(t)$ component, depending on the number of weeks between two successive purchases. Without loss of generality, we can specify the week of last purchase as week 1. Then, if the current purchase is in week k, $\theta(t)$ can be specified as in the following cases.

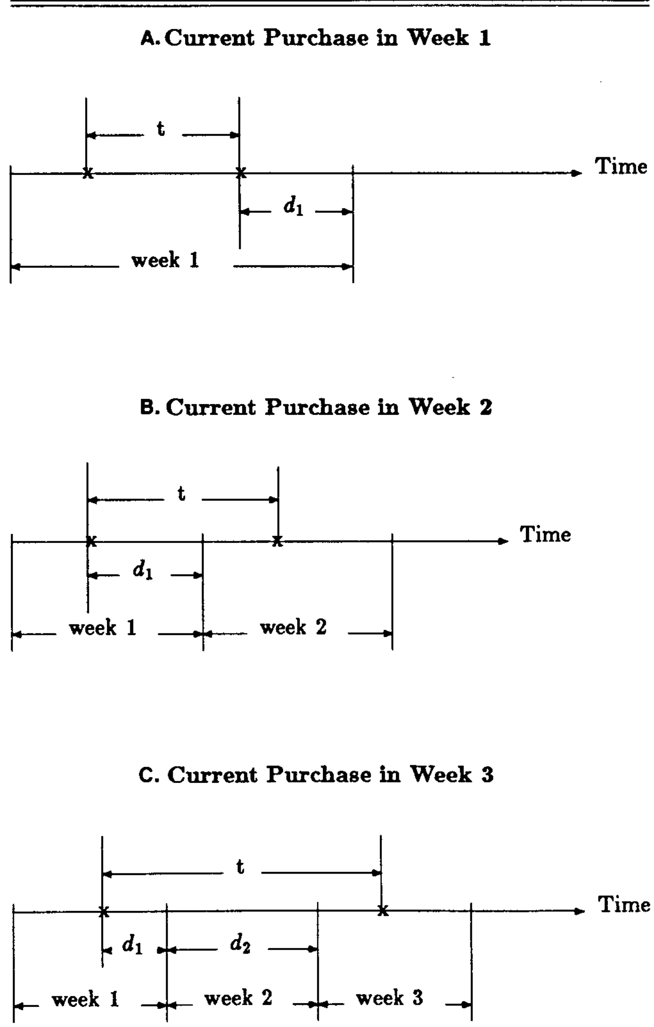
Case 1: Current purchase in week 1 (k = 1). In this case, illustrated in Figure 1A,

$$\lambda(t) = \lambda_1 = \lambda_0 \exp(\beta \mathbf{X}_1)$$

$$\theta(t) = \int_0^t \lambda(\tau) d\tau = \int_0^t \lambda_1 d\tau = \lambda_1 t.$$

Case 2: Current purchase in week 2 (k = 2). In this case, illustrated in Figure 1B,

Figure 1
CURRENT PURCHASES, CASES 1–3
(purchases represented by x's)



[4]For the purpose of this article, this time interval is assumed to be a week, a situation common in the analysis of scanner data. For example, promotional variables in the market change only on a weekly basis. However, the method can be adapted easily to any time interval. Further, these time intervals need not be of equal length.

$$\lambda(t) = \lambda_2 = \lambda_0 \exp(\beta \mathbf{X}_2)$$

$$\theta(t) = \int_0^t \lambda(\tau)d\tau = \int_0^{d_1} \lambda_1 d\tau + \int_{d_1}^t \lambda_2 d\tau$$

$$= \lambda_1 d_1 + \lambda_2(t - d_1).$$

Case 3: Current purchase in week 3 ($k = 3$). In this case, illustrated in Figure 1C,

$$\lambda(t) = \lambda_3 = \lambda_0 \exp(\beta \mathbf{X}_3)$$

$$\theta(t) = \int_0^t \lambda(\tau)d\tau = \int_0^{d_1} \lambda_1 d\tau + \int_{d_1}^{1+d_1} \lambda_2 d\tau + \int_{1+d_1}^t \lambda_3 d\tau$$

$$= \lambda_1 d_1 + \lambda_2 + \lambda_3(t - d_1 - 1).$$

From the preceding discussion, the pattern is fairly clear. Hence, for the general case in which the current purchase is in week $k (k \geq 1)$, we have

$$\lambda_k(t) = \lambda_0 \exp(\beta \mathbf{X}_k) = \lambda_0 A_k \tag{17}$$

and

$$\theta_k(t) = \lambda_1 d_1 + \sum_{w=2}^{k-1} \lambda_w + \lambda_k[t - d_1 - \gamma(k-2)] = \lambda_0 B_k, \tag{18}$$

where:

$$A_k = \exp(\beta \mathbf{X}_k), \tag{19}$$

$$B_k = A_1 d_1 + \sum_{w=2}^{k-1} A_w + A_k[t - d_1 - \gamma(k-2)], \tag{20}$$

and $\gamma = 0$ if $k = 1$; 1 otherwise. Substituting $\lambda_k(t)$ and $\theta_k(t)$ from equations 17 and 18 into equations 15 and 16, we get the survivor and density functions for the exponential distribution with time-dependent covariates. Notice that if the covariates do not vary over time, $\lambda_1 = \lambda_2 = \ldots = \lambda_k = \lambda$, so that $\lambda(t) = \lambda$ and $\theta(t) = \lambda t$. This shows why special care is required to incorporate time-dependent covariates.

The likelihood function for consumer i with n_i complete observations and one censored observation now can be written (replacing appropriate subscripts) as

$$L_i(t|\lambda_0) = \left[\prod_{j=1}^{n_i} f_{ij}(t)\right] S_i(t) \tag{21}$$

$$\cdot \left[\prod_{j=1}^{n_i} \lambda_0 A_{ijk} \exp(-\lambda_0 B_{ijk})\right] \exp(-\lambda_0 B_{ick})$$

$$= \lambda_0^{n_i} A_{ipk} \exp(-\lambda_0 B_{isk}) \exp(-\lambda_0 B_{ick})$$

where:

$$A_{ipk} = \prod_j A_{ijk},$$

$$B_{isk} = \sum_j B_{ijk}, \text{ and}$$

$B_{ick} = B_k$ for consumer i's *c*ensored observation.

Thus the log likelihood for N consumers is

$$LL = \sum_{i=1}^{N} n_i \log \lambda_0 + \log A_{ipk} - \lambda_0(B_{isk} + B_{ick}). \tag{22}$$

It is interesting to note that if no covariates are included in the model (i.e., $\mathbf{X} = \mathbf{0}$), then $A_{ipk} = 1$, $B_{isk} = t_{is}$, and $B_{ick} = t_{ic}$ (see equations 19 and 20). In such a case, the log likelihood for equation 22 reduces to the log likelihood for the model without covariates (equation 5). In other words, the model without covariates is nested in the model that includes covariates.

Erlang-2 Distribution

Erlang-2 is a special case of a gamma distribution with a shape parameter $s = 2$. The survivor function of a gamma distribution with an integer shape parameter can be expressed (Meyer 1972)[5] as

$$S(t) = \sum_{m=0}^{s-1} (\lambda t)^m \exp(-\lambda t)/m!. \tag{23}$$

For Erlang-2, the shape parameter $s = 2$ and hence the survivor function (again, subscript i is dropped for simplicity) is

$$S(t) = \exp(-\lambda t) + \lambda t \exp(-\lambda t)/1! \tag{24}$$

$$= (1 + \lambda t) \exp(-\lambda t).$$

In other words, one can think of an Erlang-2 distribution as the distribution of the time between every other event in a Poisson process with the purchase rate λ. Again, if the purchase rate is a function of time-dependent covariates, $\lambda(t)$, we get a nonhomogeneous Poisson process with a survivor function as

$$S(t) = \left(1 + \int_0^t \lambda(\tau)d\tau\right) \exp\left(-\int_0^t \lambda(\tau)d\tau\right) \tag{25}$$

$$= [1 + \theta(t)] \exp[-\theta(t)].$$

The density function is therefore

$$f(t) = \lambda(t)\left(\int_0^t \lambda(\tau)d\tau\right) \exp\left(-\int_0^t \lambda(\tau)d\tau\right) \tag{26}$$

$$= \lambda(t)\theta(t) \exp[-\theta(t)],$$

where $\lambda(t)$ and $\theta(t)$ are given as before by equations 17 and 18. As in the exponential case, we can write the log likelihood function for the Erlang-2 distribution as

$$LL = \sum_{i=1}^{N} 2n_i \log \lambda_0 + \log A_{ipk} + \log B_{ipk} \tag{27}$$

$$- \lambda_0(B_{isk} + B_{ick}) + \log(1 + \lambda_0 B_{ick}),$$

where $B_{ijk} = \Pi_j B_{ijk}$ and all other terms are as defined previously.

[5]I am grateful to John Little for suggesting this approach to me.

Exponential Distribution With Gamma Heterogeneity

Both of the models just described assume no unobserved heterogeneity in the purchase rate λ. Clearly this assumption is very restrictive. As discussed before, in the models without covariates heterogeneity is easily accommodated by allowing λ to vary across the population according to a gamma distribution. When covariates are included in the model, we have a consumer's purchase rate as $\lambda(t) = \lambda_0 \exp(\beta \mathbf{X}_t)$, where λ_0 is the base purchase rate and is assumed to be time invariant. The heterogeneity in consumers' purchase rates now can be included by allowing λ_0 to vary across the population according to a gamma distribution with parameters r and α (Wagner and Taudes 1986). This assumes the heterogeneity in consumers' base purchase rate to be time invariant. In other words, the model captures the variance in consumers' purchase times by two components—one that captures the heterogeneity in consumers' base purchase rate and is time invariant, and one that assumes homogeneity in response coefficients β but captures temporal variance because it includes time-dependent covariates.

The log likelihood function for this model is in Table 1 and its derivation is in the Appendix.

Erlang-2 Distribution With Gamma Heterogeneity

If consumers' interpurchase times are distributed Erlang-2 and the base rate λ_0 is distributed gamma across people, we obtain an Erlang-2/gamma distribution with time-dependent covariates. The log-likelihood function for this model is in Table 1 and its derivation is in the Appendix. Notice that in this model we have a duration-dependent individual-level model (Erlang-2), we include heterogeneity (gamma), we incorporate nonstationarity through time-dependent covariates ($\mathbf{X}_t$), and we account for the right-censored data through the survivor function $S(t)$.

One of the appealing aspects of this approach of including covariates is the analytical tractability of the likelihood functions, which is readily seen by noting the similarity of the expressions in the top and bottom halves of Table 1. The expressions in the bottom half of the table are generalized versions of the expressions in the top half. If the covariates are excluded from the models, the expressions in the bottom half of Table 1 collapse to those in the top half. In other words, the models without covariates are nested within the models that include covariates. Therefore a likelihood ratio test can be used to see the improvement in the models' fit due to inclusion of covariates.

The log-likelihood functions as given in Table 1 are maximized by using the David-Fletcher-Powell algorithm available in the FORTRAN library GQOPT.

AN APPLICATION TO SCANNER DATA

Data and Variables

The proposed approach is illustrated with scanner panel data for regular ground coffee. After elimination of noncontinuously reporting households and very light purchasers, a random sample of 100 households was selected.[6] Data for 40 weeks were used for model calibration, and the subsequent 25 weeks were used for model validation. The covariates used in the analysis were:

—household weekly inventory,
—regular price,
—promotional price cut,
—feature and display, and
—feature or display.

Previous studies show that a consumer is likely to wait if he or she has a large product inventory at home (Blattberg, Eppen, and Lieberman 1981; Neslin, Henderson, and Quelch 1985). We therefore would expect the inventory coefficient to be negative, indicating reduction in purchase rate due to large inventory. Inventory operationalization is similar to that in previous studies—inventory in week w = inventory in week $(w - 1)$ + quantity bought in week $(w - 1)$ − weekly consumption. Average weekly purchase quantity of a household in the calibration period was used as a surrogate for weekly consumption (Gupta 1988). Further, we recognize that inventory of 30 ounces for a household that consumes 5 ounces/week is not the same as inventory of 30 ounces for a household that consumes 20 ounces/week. Hence, we convert the inventory from ounces to number of weeks it is likely to last for a particular household. In our example, the first household has inventory worth $30/5 = 6$ weeks and the second household has inventory worth $30/20 = 1.5$ weeks. Finally, to avoid any effects due to outliers, we clip the inventory if it becomes negative or goes beyond six months (Guadagni and Little 1987).

Four marketing variables are used. Because we are modeling the interpurchase time at the *product-category* level, we use these variables as the weighted average across brands and stores. The weights used are individual-household-level brand and store shares.[7] This individual-level weighting ensures that a brand loyal consumer is influenced by the price and promotion of his or

[6]To obtain stable initialization of certain variables, such as weekly consumption rate and household inventory, households that do not make at least 10 purchases in two years are eliminated. This approach is similar to those adopted by several researchers (Bass et al. 1984; Guadagni and Little 1983; Gupta 1988). Clearly this process eliminates the "tail of the distribution," making the purchase rates more homogeneous than that of the overall population. However, given that the eliminated households accounted for less than 1.5% of the sales volume, the remaining households may represent the *managerially relevant* population.

[7]If the last purchase is in week 1 and the current purchase is in week k, we could use the actual price and promotion values for the brands bought in those two weeks. However, it is not quite clear which brand's price and promotion values should be used during the intermediate $(k - 2)$ weeks. This information is needed, as shown in equations 18 and 20. The marketing variables therefore are weighted. Different weighting schemes were used, such as equal weights, market share weights, and weight of 1 to favorite brand and store. All weighting schemes gave approximately the same results.

her favorite brand only. Regular price and promotional price discounts are in cents/ounce. As feature and display are very highly correlated in this dataset, two variables are created—feature and display (1 if both feature and display are present, 0 otherwise) and feature or display (1 if either feature or display is present but not both, 0 otherwise). All promotional variables (i.e., feature and display, feature or display, and price cut) are expected to increase consumers' purchase rates. Hence we would expect all these variables to have a positive coefficient. In contrast, a high regular price is likely to reduce the propensity to buy and hence it is likely to have a negative coefficient.

Analysis and Results

The effect of covariates on consumers' interpurchase time is discussed first, and then the different models are compared on the basis of performance.

Effect of covariates

The results for the models are reported in Table 2. All models with covariates have the expected sign for the coefficients of household inventory and feature and/or display. The negative coefficient of inventory suggests that high inventory is likely to make consumers wait longer and hence reduce their purchase rates. In contrast, feature and/or display increase consumers' purchase rates and make them buy early. To get a better understanding of the relative effects of these variables on consumers' purchase rates, purchase rate elasticities were calculated. As indicated before, purchase rate elasticity with respect to a covariate X_l is equal to $\beta_l X_l$. Point elasticities were calculated at the mean value of the covariates. Results are reported in Table 3 and suggest that:

1. Household inventory has a much greater effect on consumers' purchase rates than feature and/or display. Purchase rate elasticities due to inventory are two to three times larger than the corresponding elasticities due to feature and display.
2. Purchase rate elasticity due to inventory is higher for exponential and exp/gamma models than for Erlang-2 and Erlang-2/gamma models. Why is this so? The intuitive reason is simple. A significant negative sign for inventory indicates that higher inventory makes consumers wait longer. In other words, a purchase in the previous period raises inventory in the current period, thereby reducing the probability of purchase. This *dead period effect* (i.e., low purchase probability soon after the last purchase) is built into the Erlang-2 distribution (its mode is not at zero), but not the exponential distribution (its mode is at zero). Therefore, in the exponential model, the inventory variable is capturing all the *dead period effect* whereas in the Erlang-2 model that effect is captured by both the inventory variable and the model *per se*. Consequently, purchase rate elasticity due to inventory is lower for Erlang-2 models than for exponential models.
3. In the models without heterogeneity, purchase rate elasticity due to feature and display is more than twice the elasticity due to feature or display, suggesting a synergistic effect of feature and display. However, this synergistic effect disappears when heterogeneity is added in the models. This finding shows the importance of including heterogeneity in the models.

Though the effect of feature and/or display is statistically significant, price cut and regular price have no

Table 2

CALIBRATION RESULTS[a]

	Exponential		*Erlang-2*		*Exp/gamma*		*Erlang-2/gamma*	
	Without covariates	*With covariates*	*Without covariates*	*With covariates*	*Without covariates*	*With covariates*	*Without covariates*	*With covariates*
λ	.376	.421	.768	.786	—	—	—	—
	(43.10)	(23.65)	(85.07)	(38.02)				
r		—		—	4.955	5.692	4.270	4.938
					(4.89)	(4.83)	(5.83)	(5.91)
α		—		—	13.258	13.184	5.602	6.500
					(4.59)	(4.53)	(5.36)	(9.54)
Feature and display		1.648		1.499		1.601		1.374
		(10.19)		(12.06)		(10.44)		(11.10)
Feature or display		.561		.412		.780		.682
		(2.93)		(3.18)		(3.98)		(4.31)
Inventory		−.068		−.040		−.075		−.037
		(−11.88)		(−13.97)		(−10.13)		(−7.14)
Price cut		.010		.058		−.033		−.025
		(.16)		(1.38)		(−.53)		(−.50)
Regular price		−.034		.005		−.089		−.014
		(−1.08)		(.25)		(−1.51)		(−.54)
−LL (n=1526)	2819.85	2695.93	2847.62	2725.83	2751.71	2639.11	2640.40	2556.68
ρ^2	—	.044	—	.043	—	.041	—	.032
AIC	−2820.85	−2701.93	−2848.62	−2731.83	−2753.71	−2646.11	−2642.40	−2563.68

[a]t-values in parentheses.

Table 3
PURCHASE RATE (λ) ELASTICITY[a]

Covariates	*Exponential*	*Erlang-2*	*Exp/gamma*	*Erlang-2/gamma*
Feature and display	.137	.124	.133	.114
Feature or display	.049	.036	.068	.059
Inventory	−.408	−.240	−.450	−.222

[a]Numbers in the table indicate percentage change in the purchase rate (λ) for 1% change in the covariates. These are point elasticities at the mean value of the covariates.

significant effect. As noted previously (Gupta 1988, p. 348), "This finding suggests that a consumer who is not planning to buy coffee in a given week may not check the prices or price discounts on coffee brands unless his or her attention is attracted to them through feature and/or display." Guadagni and Little (1987) also found that in comparison with price, household coffee inventory and category attractiveness (which is primarily influenced by promotion) have greater influence on consumers' probability of purchasing the category.

All the models with covariates were reestimated after exclusion of the price cut and regular price variables. No major changes in the parameters for the remaining variables were found. Further, a likelihood ratio test indicated that the models' fit did not drop significantly.

Model comparison

The preceding discussion on models shows that the models differ on three dimensions:

1. The model does/does not include covariates.
2. The model does/does not include heterogeneity in purchase rates.
3. The individual-level model is/is not duration dependent (Erlang-2 vs. exponential).

Hence, there are eight different models (see Table 1). To gain some insights into the performance of these models, all of them were estimated on the same data and compared on the basis of their fit to the data. The models with and without covariates are nested models and hence can be compared by likelihood ratio test. The nonnested models (e.g., exp/gamma vs. Erlang-2/gamma) are compared by using Akaike's information criterion, AIC (Rust and Schmittlein 1985).

Models without covariates versus models with covariates. Comparing the four models (exponential, Erlang-2, exp/gamma, and Erlang-2/gamma) with and without covariates leads to two key observations. First, the inclusion of covariates significantly improves the models' fit. In all cases the likelihood ratio test suggests that the inclusion of three covariates—inventory, feature, and/or display—is desirable.

Second, though the improvement is statistically significant, the overall variance explained by the inclusion of these variables is still rather small (ρ^2 is about 5%).[8] This finding is consistent with those of previous studies showing that marketing mix variables do not explain a large portion of the variance in consumers' interpurchase times (Bucklin and Lattin 1990; Guadagni and Little 1987; Gupta 1988).

Models without heterogeneity versus models with heterogeneity. We can assess the impact of including heterogeneity in the model by comparing exponential with exp/gamma models and Erlang-2 with Erlang-2/gamma models. First, we note that the inclusion of heterogeneity improves the model fit, as is evident from the higher AIC value of these models.[9] Second, we note that the effect of heterogeneity on parameter estimates is not very large. This finding is not surprising if we note that the degree of heterogeneity in the sample is low as indicated by the high value of the r parameter. In accordance with other research, households were selected that made at least 10 purchases in two years. Though the eliminated households account for less than 1.5% of the sales volume, making the remaining sample more relevant to managers, this elimination of very light buyers makes the sample relatively homogeneous as reflected by the high value of r. Because of the low heterogeneity in the sample, the parameters do not vary dramatically when heterogeneity is included in the models. In general, inclusion of heterogeneity is important to ensure that the parameter estimates are not biased.

Models without duration dependence versus models with duration dependence. If an individual consumer's interpurchase time is assumed to be distributed exponentially, we implicitly assume that time since last purchase has no impact on the consumer's current purchase time decision (i.e., there is no duration dependence). An Erlang-2 model, in contrast, is duration dependent. Comparing these models after including heterogeneity (i.e.,

[8]$\rho^2 = 1 - L(\mathbf{X})/L(\mathbf{0})$, where $L(\mathbf{0})$ is the log likelihood of the null or restricted model and $L(\mathbf{X})$ is the log likelihood of the model with covariates (Ben-Akiva and Lerman 1985; Hauser 1978).

[9]AIC = log(maximum likelihood) − (number of estimated parameters). Larger values of AIC indicate more preferred models (Rust and Schmittlein 1985).

comparing exp/gamma and Erlang-2/gamma models) leads to two observations. First, the duration-dependent Erlang-2/gamma model fits the data much better than the exp/gamma model (Table 2). Second, the r parameter for the Erlang-2/gamma model is less than that for the exp/gamma. In other words, the Erlang-2/gamma model suggests greater heterogeneity in consumers' purchase rates than that indicated by the exp/gamma model. How can we explain this result? The intuitive explanation is simple. Conceptually, one can visualize the total variance in consumers' purchase times as consisting of two components: within-consumer variance and between-consumer variance (Morrison and Schmittlein 1988). An Erlang-2/gamma model posits an Erlang-2 distribution at the individual consumer level, which implies purchases more "regular" than exponential. In other words, Erlang-2 suggests smaller within-consumer variance than that suggested by the exponential. Therefore, given the same total variance in purchase times, an Erlang-2/gamma distribution in general will show larger between-consumer variance than the exp/gamma distribution. A large between-consumer variance implies more heterogeneity and hence smaller r value. Therefore an Erlang-2/gamma model in general will have a smaller r value than the exp/gamma model. Results of Table 2 support this argument.

Model validation. A further comparison of the models is done by validating them on a holdout sample. Schmittlein, Bemmaor, and Morrison (1985, p. 264) stress the importance of model validation: "It is well known that simpler models . . . can outperform more complicated models even if the more complicated model is 'true'. So . . . the researcher . . . will want . . . to use some type of cross validation as an aid in model selection." The models were validated on a 25-week holdout period. The interpurchase time, weekly value of the covariates in the validation period, and model parameters as estimated in the calibration period were used to evaluate the log likelihood for each model during the validation period. The results (Table 4) are consistent with the calibration results. Specifically, (1) the inclusion of covariates improves the model and this improvement is moderate though statistically significant, (2) inclusion of heterogeneity improves the model, and (3) the Erlang-2/gamma model with covariates performs the best.

A final comparison of the models is between the actual and model-predicted number of purchases in each of the 25 weeks of the holdout period. Actual values of the covariates were used in conjunction with parameter estimates of a model to calculate a consumer's purchase probability in a week. The predicted number of purchases in a week therefore is given by individual purchase probabilities summed across the sample of 100 households. For any model, the purchase probability for a consumer is estimated as

$$
\begin{aligned}
p_k &= \text{probability of buying in week } k, \\
&\quad \text{given the week of last purchase} \\
&\quad \text{is designated as 1,} \\
&\quad \text{(i.e., interpurchase time is } k \text{ weeks)} \\
&= P(T \le k | T > k - 1) \\
&= 1 - P(T > k | T > k - 1) \\
&= 1 - \frac{P(T > k)}{P(T > k - 1)} \\
&= 1 - \frac{S_k}{S_{k-1}},
\end{aligned}
$$

where S_k is the survivor function for a model. For models that include heterogeneity, we first simulate a consumer's base purchase rate (λ_0) as per the r and α parameters of the gamma distribution. We next use the exponential or the Erlang-2 survivor function along with the parameter estimates as given in Table 2 to predict the purchase probability of a consumer for a given week. Then we perform a Monté Carlo simulation in which the purchase

Table 4
VALIDATION RESULTS

Model	*−LL (n = 853)*	*Number of parameters*	*AIC*[a]	χ^{2}[b]	p[c]
1. Exponential, no covariates	1553.24	1	−1554.24	—	—
2. Exponential, covariates	1535.82	4	−1539.82	34.84	.0001
3. Erlang-2, no covariates	1575.26	1	−1576.26	—	—
4. Erlang-2, covariates	1551.50	4	−1555.50	47.50	.0001
5. Exp/gamma, no covariates	1501.32	2	−1503.32	—	—
6. Exp/gamma, covariates	1486.24	5	−1491.24	30.16	.0001
7. Erlang-2/gamma, no covariates	1411.90	2	−1413.90	—	—
8. Erlang-2/gamma, covariates	1392.42	5	−1397.42	38.96	.0001

[a]AIC = log likelihood − number of estimated parameters. Larger values indicate preferred model.

[b]−2* difference in the log likelihood between the model with covariates and model without covariates is distributed χ^2. This likelihood ratio test is used to test nested models (e.g., exponential with covariates vs. exponential without covariates).

[c]For example, p value for $\chi^2 = 34.84$ with d.f. $= 4 - 1 = 3$ additional parameters is .0001.

Table 5
VALIDATION RESULTS: ACTUAL VERSUS PREDICTED NUMBER OF PURCHASES[a]

Week	Actual number of purchases	Exponential: Without covariates	Exponential: With covariates	Erlang-2: Without covariates	Erlang-2: With covariates	Exp/gamma: Without covariates	Exp/gamma: With covariates	Erlang-2/gamma: Without covariates	Erlang-2/gamma: With covariates
1	49	31	43	32	44	30	42	33	41
2	42	31	45	32	45	30	44	30	37
3	29	31	42	31	39	29	42	29	36
4	21	31	41	33	38	29	41	29	34
5	43	31	44	31	42	30	43	31	40
6	27	31	35	32	34	29	34	29	32
7	23	31	39	33	37	29	38	29	35
8	27	31	37	32	37	29	36	30	32
9	45	31	43	32	42	30	42	28	40
10	28	31	41	33	39	30	40	28	36
11	30	31	40	31	37	30	39	31	33
12	32	31	36	34	36	30	36	30	33
13	20	31	39	31	37	29	39	30	35
14	48	31	38	32	37	30	37	30	35
15	41	31	38	32	38	30	37	30	33
16	22	31	38	31	35	29	37	29	32
17	51	31	42	33	41	29	41	30	37
18	41	31	37	33	34	30	37	30	33
19	31	31	37	33	35	30	36	29	33
20	30	31	38	33	34	30	38	30	34
21	39	31	44	33	42	30	43	30	39
22	38	31	46	32	43	29	45	30	41
23	32	31	37	34	34	28	36	29	31
24	30	31	38	31	37	28	37	27	34
25	40	31	37	31	34	28	36	29	32
Theil's U	—	.1416	.1303	.1369	.1166	.1483	.1274	.1408	.1130

[a]Predicted number of purchases is based on predicted purchase probabilities of consumers, summed across the sample and rounded to the nearest decimal.

probability is used to simulate whether or not this consumer buys coffee in this week. If no coffee purchase is predicted, we estimate the purchase probability for the next week and repeat the process. If a coffee purchase is predicted, we assume that the consumer buys a quantity equal to his or her *average* product quantity purchase, as estimated in the calibration period. This quantity then is used to update the consumer's product inventory, which is one of the covariates used in future predictions.

The predicted number of product purchases in each of the 25 weeks of the validation period is compared with the actual number of purchases in that week. The predictive quality of a model is assessed by using Theil's inequality coefficient, which is calculated as

$$U = \frac{\sqrt{\sum_{w=1}^{25} (n_w - \hat{n}_w)^2/25}}{\sqrt{\sum_{w=1}^{25} (n_w)^2/25} + \sqrt{\sum_{w=1}^{25} (\hat{n}_w)^2/25}},$$

where n_w and $\hat{n}_w$ are actual and model-predicted number of purchases in week w. U ranges from 0 to 1, where smaller values mean better predictions.

The prediction results, reported in Table 5, support the results of Table 4 by showing that a model's predictions improve (as indicated by Theil's U) when covariates and heterogeneity are included. Again, the Erlang-2/gamma model with covariates performs best, though its predictions are only marginally better than those of the competing models. Relatively small explanatory power of the covariates as demonstrated by low ρ^2 and relatively low heterogeneity in the sample as indicated by high r are in part responsible for this result.

CONCLUSIONS

Stochastic models have been used extensively in marketing for more than three decades. In spite of their parsimony and good fit to the data, they have been criticized for not including explanatory variables. In this article a methodology is proposed by which the commonly used stochastic models can be extended easily to incorporate time-dependent covariates, such as marketing mix variables. The methodology provides simple, closed-form expressions for the likelihood function of these models, making estimation and prediction simple and fast. Further, the heterogeneity parameter (r) gives useful insights about the variations in consumers' purchase rates.

If the data show a large degree of heterogeneity, and frequent buyers are believed to respond to marketing variables differently than infrequent buyers, it may be useful to consider segment-level analysis. We have seen why special care is needed to handle covariates that vary over time. The methodology is illustrated by an application to scanner panel data for ground coffee. The inclusion of covariates helps improve the model fit, diagnostics, and predictions. For the data used in the application, the Erlang-2/gamma model performs best.

Future research in this area can take two directions: methodological and substantive. The methodological extensions may include more general formulations of the purchase time distribution and the heterogeneity. Good examples of this approach are found in the work of Robbins (1977), Morrison and Schmittlein (1981), and Jain and Vilcassim (1990), who show how arbitrary heterogeneity in consumers' purchase rates can be included. However, as Morrison and Schmittlein (1988) note, these generalizations come at certain costs. For example, unless the sample size is very large, the statistical estimates are not very robust. Further, using these generalized approaches makes the estimation and specially the predictions very cumbersome and time consuming. Even these "generalized" approaches do not account for all possible means of modeling the purchase time phenomenon. For example, Kahn and Morrison (1989) show that a geometric distribution with mean interpurchase time of two shopping trips will appear to be an Erlang-2 distribution.

Substantively, it may be useful to extend this work to include shopping trip behavior. For example, in a recent study Kahn and Schmittlein (1989) found differences in interpurchase time and intershopping times. Because purchase in a product category is contingent on whether or not a shopping trip is made by the consumer, inclusion of such information may enrich our understanding of consumers' buying behavior. Future research also could examine not only purchase acceleration due to promotions, but also purchase delays. In other words, if consumers expect promotions in the near future, their expectations of future promotions may cause them to postpone their purchases. All these studies should add to our understanding of consumer purchase dynamics and how it is affected by marketing variables.

APPENDIX
LIKELIHOOD FUNCTIONS FOR EXPONENTIAL AND ERLANG-2 DISTRIBUTIONS WITH GAMMA HETEROGENEITY

Exp/Gamma Model

We start with a model without covariates. Here consumer i has n_i complete observations, which are distributed exponentially with parameter λ_i. The likelihood function for this consumer (see equation 3) is

$$\text{(A1)} \qquad L_i(t|\lambda_i) = \lambda_i^{n_i} \exp[-\lambda_i(t_{is} + t_{ic})],$$

where:

$$t_{is} = \sum_{j=1}^{n_i} t_{ij}$$

= sum of n_i interpurchase times of consumer i, and

t_{ic} = censored time for consumer i.

If we allow heterogeneity in consumers' purchase rates according to a gamma distribution $g(\lambda_i)$, the unconditional likelihood for consumer i is

$$\text{(A2)} \qquad L_i(t) = \int_0^\infty L_i(t|\lambda_i) g(\lambda_i)\, d\lambda_i$$

where:

$$\text{(A3)} \qquad g(\lambda_i) = \frac{\alpha}{\Gamma(r)} (\alpha\lambda_i)^{r-1} \exp(-\alpha\lambda_i).$$

Using equations A1 through A3, we get

$$\text{(A4)} \quad L_i(t) = \int_0^\infty \lambda_i^{n_i} \exp[-\lambda_i(t_{is} + t_{ic})] \cdot \frac{\alpha}{\Gamma(r)} (\alpha\lambda_i)^{r-1} \exp(-\alpha\lambda_i)\, d\lambda_i$$
$$= \frac{\alpha^r}{\Gamma(r)} \int \lambda_i^{n_i+r-1} \exp[-\lambda_i(t_{is} + t_{ic} + \alpha)]\, d\lambda_i$$
$$= \frac{\alpha^r}{\Gamma(r)} \frac{\Gamma(n_i + r)}{(t_{is} + t_{ic} + \alpha)^{n_i+r}}$$
$$= \frac{\alpha^r}{(t_{is} + t_{ic} + \alpha)^{n_i+r}} \prod_{j=0}^{n_i-1} (r + j).$$

Therefore, log likelihood for N consumers is

$$\text{(A5)} \qquad LL = \sum_{i=1}^{N} \left[r \log \alpha + \sum_{j=0}^{n_i-1} \log(r + j) - (n_i + r) \log (t_{is} + t_{ic} + \alpha) \right].$$

If covariates are included in the model, the conditional likelihood for consumer i (see equation 21) is

$$\text{(A6)} \qquad L_i(t|\lambda_0) = \lambda_0^{n_i} A_{ipk} \exp[-\lambda_0(B_{isk} + B_{ick})],$$

where A_{ipk}, B_{isk}, B_{ick} are as defined in equation 21. If no covariates are included, $A_{ipk} = 1$, $B_{isk} = t_{is}$, $B_{ick} = t_{ic}$, so that equation A6 is identical to equation A1.

Again, if heterogeneity is included as per gamma distribution, the unconditional likelihood for consumer i is

$$\text{(A7)} \quad L_i(t) = \int_0^\infty L_i(t|\lambda_0) g(\lambda_0)\, d\lambda_0$$
$$= \int \lambda_0^{n_i} A_{ipk} \exp[-\lambda_0(B_{isk} + B_{ick})] \cdot \frac{\alpha}{\Gamma(r)} (\alpha\lambda_0)^{r-1} \exp(-\alpha\lambda_0)\, d\lambda_0$$

$$= \frac{\alpha^r A_{ipk}}{\Gamma(r)} \int \lambda_0^{n_i+r-1} \exp[-\lambda_i(B_{isk} + B_{ick} + \alpha)]\, d\lambda_0$$

$$= \frac{\alpha^r A_{ipk}}{\Gamma(r)} \frac{\Gamma(n_i + r)}{(B_{isk} + B_{ick} + \alpha)^{n_i+r}}$$

$$= \frac{\alpha^r A_{ipk}}{(B_{isk} + B_{ick} + \alpha)^{n_i+r}} \prod_{j=0}^{n_i-1} (r + j).$$

Notice that the terms A and B include household-specific covariates $\mathbf{X}$. The integration allows us to go from $L_i(t|\lambda_0, \mathbf{X})$ to $L_i(t|\mathbf{X})$. In other words, $L_i(t|\lambda_0,\mathbf{X})$ is the likelihood function with both unobservable and observable heterogeneity whereas $L_i(t|\mathbf{X})$ is the likelihood function with only observable heterogeneity (e.g., household product inventory). Heckman and Singer (1984b) indicate that this method is appropriate if one makes two assumptions: (1) λ_0 is distributed independently of $\mathbf{X}$ and (2) there are no functional restrictions connecting the conditional distribution of T given λ_0 and $\mathbf{X}$ and the marginal distributions of λ_0 and $\mathbf{X}$.

The log-likelihood function for N consumers now can be written as

$$\text{(A8)} \quad LL = \sum_{i=1}^{N} \left[r \log \alpha + \log A_{ipk} + \sum_{j=0}^{n_i-1} \log(r + j) - (n_i + r) \log (B_{isk} + B_{ick} + \alpha) \right].$$

Erlang-2/Gamma Model

Again, we start with a model without covariates. In this case the distribution of interpurchase time at the individual-consumer level is assumed to be Erlang-2. The likelihood function for consumer i (see equation 10) is

$$\text{(A9)} \quad L_i(t|\lambda_i) = \lambda_i^{2n_i} t_{ip} \exp[-\lambda_i(t_{is} + t_{ic})]\, (1 + \lambda_i t_{ic}).$$

Allowing heterogeneity in purchase rates, we get the unconditional likelihood for consumer i as

$$\text{(A10)} \quad L_i(t) = \int_0^\infty L_i(t|\lambda_i) g(\lambda_i)\, d\lambda_i$$

$$= \int \lambda_i^{2n_i} t_{ip} \exp[-\lambda_i(t_{is} + t_{ic})]\, (1 + \lambda_i t_{ic}) \cdot \frac{\alpha}{\Gamma(r)} (\alpha\lambda_i)^{r-1} \exp(-\alpha\lambda_i)\, d\lambda_i$$

$$= \frac{\alpha^r t_{ip}}{\Gamma(r)} \left[\int \lambda_i^{2n_i+r-1} \exp(-\lambda_i(t_{is} + t_{ic} + \alpha))\, d\lambda_i + t_{ic} \int \lambda_i^{2n_i+r} \exp(-\lambda_i(t_{is} + t_{ic} + \alpha))\, d\lambda_i \right]$$

$$= \frac{\alpha^r t_{ip}}{\Gamma(r)} \left[\frac{\Gamma(2n_i + r)}{(t_{is} + t_{ic} + \alpha)^{2n_i+r}} + \frac{\Gamma(2n_i + r + 1) t_{ic}}{(t_{is} + t_{ic} + \alpha)^{2n_i+r+1}} \right]$$

$$= \frac{\alpha^r t_{ip}}{(t_{is} + t_{ic} + \alpha)^{2n_i+r}} \left[1 + \frac{(2n_i + r) t_{ic}}{t_{is} + t_{ic} + \alpha} \right] \prod_{j=0}^{2n_i-1} (r + j).$$

Therefore the log likelihood for N consumers is

$$\text{(A11)} \quad LL = \sum_{i=1}^{N} \left[r \log \alpha + \log t_{ip} + \sum_{j=0}^{2n_i-1} \log(r + j) + \log\left(1 + \frac{(2n_i + r) t_{ic}}{t_{is} + t_{ic} + \alpha} \right) - (2n_i + r) \log (t_{is} + t_{ic} + \alpha) \right].$$

If covariates are included in the model, the conditional likelihood for consumer i can be written (see equation 27) as

$$\text{(A12)} \quad L_i(t|\lambda_0) = \lambda_0^{2n_i} A_{ipk} B_{ipk} \exp[-\lambda_0(B_{isk} + B_{ick})] \cdot (1 + \lambda_0 B_{ick}).$$

Again, allowing gamma heterogeneity, we get the unconditional likelihood for consumer i as

$$L_i(t) = \int_0^\infty L_i(t|\lambda_0) g(\lambda_0)\, d\lambda_0.$$

Using a procedure similar to equation A10, we can easily show that

$$\text{(A13)} \quad L_i(t) = \frac{\alpha^r A_{ipk} B_{ipk}}{(B_{isk} + B_{ick} + \alpha)^{2n_i+r}} \left[1 + \frac{(2n_i + r) B_{ick}}{B_{isk} + B_{ick} + \alpha} \right] \cdot \prod_{j=0}^{2n_i-1} (r + j).$$

Hence the log likelihood for N consumers is

$$\text{(A14)} \quad LL = \sum_{i=1}^{N} \left[r \log \alpha + \log A_{ipk} + \log B_{ipk} + \sum_{j=0}^{2n_i-1} \log(r + j) + \log\left(1 + \frac{(2n_i + r) B_{ick}}{B_{isk} + B_{ick} + \alpha} \right) - (2n_i + r) \log (B_{isk} + B_{ick} + \alpha) \right].$$

The results are summarized in Table 1.

REFERENCES

Bass, Frank M., Moshe M. Givon, Manohar Kalwani, David Reibstein, and Gordon P. Wright (1984), "An Investigation Into the Order of the Brand Choice Process," *Marketing Science*, 3 (Fall), 267–87.

Ben-Akiva, Moshe and Steven R. Lerman (1985), *Discrete Choice Analysis*. Cambridge, MA: MIT Press.

Blattberg, Robert C., Gary D. Eppen, and Joshua Lieberman (1981), "A Theoretical and Empirical Evaluation of Price

Deals for Consumer Nondurables," *Journal of Marketing*, 45 (Winter), 116–29.

Bucklin, Randolph E. and James M. Lattin (1990), "A Two-State Model of Purchase Incidence and Brand Choice," *Marketing Science*, forthcoming.

Chatfield, Christopher and Gerald J. Goodhardt (1973), "A Consumer Purchasing Model With Erlang Inter-Purchase Times," *Journal of the American Statistical Association*, 68 (December), 828–35.

Cox, D. R. (1972), "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, B, 74, 187–220.

——— (1975), "Partial Likelihood," *Biometrika*, 62 (2), 269–76.

——— and D. O. Oakes (1984), *Analysis of Survival Data*. New York: Chapman and Hall.

Efron, B. (1977), "The Efficiency of Cox's Likelihood Function for Censored Data," *Journal of the American Statistical Association*, 72 (September), 557–65.

Ehrenberg, Andrew S. C. (1959), "The Pattern of Consumer Purchases," *Applied Statistics*, 8 (March), 26–41.

——— (1972), *Repeat Buying: Theory and Applications*. Amsterdam: North-Holland Publishing Company.

Flinn, C. and James J. Heckman (1982), "Models for the Analysis of Labor Force Dynamics," in *Advances in Econometrics*, Vol. 1, G. Rhodes and R. Basmann, eds. Greenwich, CT: JAI Press, Inc., 35–95.

Guadagni, Peter and John D. C. Little (1983), "A Logit Model of Brand Choice Calibrated on Scanner Data," *Marketing Science*, 2 (Summer), 203–38.

——— and ——— (1987), "When and What to Buy: A Nested Logit Model of Coffee Purchases," Working Paper No. 1919–87, Massachussets Institute of Technology.

Gupta, Sunil (1987), "An Integrated Model of Interpurchase Time, Brand Choice, and Purchase Quantity," unpublished dissertation, Graduate School of Business, Columbia University.

——— (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (November), 342–56.

Hauser, John R. (1978), "Testing the Accuracy, Usefulness and Significance of Probabilistic Choice Models: An Information Theoretic Approach," *Operations Research*, 26 (May), 406–21.

Heckman, James J. and Bud Singer (1984a), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52 (March), 271–320.

——— and ——— (1984b), "Econometric Duration Analysis," *Journal of Econometrics*, 24 (1), 63–132.

Helsen, Kristiaan and David C. Schmittlein (1989), "Analyzing Duration Times in Marketing Research," working paper, Wharton School, University of Pennsylvania.

Herniter, Jerome (1971), "A Probabilistic Market Model of Purchase Timing and Brand Selection," *Management Science*, 18 (December), Part II, 102–13.

Jain, Dipak C. and Naufel J. Vilcassim (1990), "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach," *Marketing Science*, forthcoming.

Jeuland, Abel P., Frank M. Bass, and Gordon P. Wright (1980), "A Multibrand Stochastic Model Compounding Heterogeneous Erlang Timing and Multinomial Choice Process," *Operations Research*, 28 (March–April), 255–77.

Jones, Morgan and Fred Zufryden (1980), "Adding Explanatory Variables to a Consumer Purchase Behavior Model: An Exploratory Study," *Journal of Marketing Research*, 17 (August), 323–34.

Kahn, Barbara E. and Donald G. Morrison (1989), "A Note on 'Random' Purchasing: Additional Insights From Dunn, Reader and Wrigley," *Applied Statistics*, 38 (1), 111–14.

——— and David C. Schmittlein (1989), "Shopping Trip Behavior: An Empirical Investigation," *Marketing Letters*, 1 (December), 55–70.

Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*. New York: John Wiley & Sons, Inc.

Meyer, Paul L. (1972), *Introductory Probability and Statistical Applications*. Reading, MA: Addison-Wesley Publishing Company.

Morrison, Donald G. and David C. Schmittlein (1981), "Predicting Future Random Events Based on Past Performance," *Management Science*, 27 (September), 1006–23.

——— and ——— (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?" *Journal of Business and Economic Statistics*, 6 (April), 145–66.

Neslin, Scott A., Caroline Henderson, and John Quelch (1985), "Consumer Promotions and the Acceleration of Product Purchases," *Marketing Science*, 4 (Spring), 147–65.

Oakes, D. O. (1977), "The Asymptotic Information in Censored Survival Data," *Biometrika*, 64, 441–8.

Robbins, H. (1977), "Prediction and Estimation for the Compound Poisson Distribution," *Proceedings of the Academy of Science*, 74, 2670–1.

Ross, Sheldon M. (1980), *Introduction to Probability Models*, 2nd ed. New York: Academic Press, Inc.

Rust, Ronald T. and David C. Schmittlein (1985), "A Bayesian Cross-Validated Likelihood Method for Comparing Alternative Specifications of Quantitative Models," *Marketing Science*, 4 (Winter), 20-40.

Sabavala, Darius J. (1988), "Comment on Generalizing the NBD Model," *Journal of Business and Economic Statistics*, 6 (2), 161–2.

Schmittlein, David C., A. C. Bemmaor, and Donald G. Morrison (1985), "Why Does the NBD Model Work? Robustness in Representing Product Purchases, Brand Purchases and Imperfectly Recorded Purchases," *Marketing Science*, 4 (Summer), 255–66.

———, Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science*, 33 (January), 1–24.

Shoemaker, Robert W. (1979), "An Analysis of Consumer Reactions to Product Promotions," in *American Marketing Association Proceedings*. Chicago: American Marketing Association, 244–8.

Wagner, Udo and Alfred Taudes (1986), "A Multivariate Polya Model of Brand Choice and Purchase Incidence," *Marketing Science*, 5 (Summer), 219–44.

Ward, Ronald W. and James E. Davis (1978), "A Pooled Cross-Section Time Series Model of Coupon Promotions," *American Journal of Agricultural Economics*, 60 (August), 393–401.

Wheat, Rita D. and Donald G. Morrison (1990a), "Assessing Purchase Timing Models: Whether or Not is Preferable to When," *Marketing Science*, forthcoming.

——— and ——— (1990b), "Measuring Purchase Regularity," *Journal of Marketing Research*, 27 (February), 87–93.
Zufryden, Fred S. (1977), "A Composite Heterogeneous Model of Brand Choice and Purchase Timing Behavior," *Management Science*, 24 (October), 121–36.
——— (1978), "An Empirical Evaluation of a Composite Heterogeneous Model of Brand Choice and Purchase Timing Behavior," *Management Science*, 25 (March), 761–73.

Reprint No. JMR281100