



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

A Dynamic Changepoint Model for New Product Sales Forecasting

Peter S. Fader, Bruce G. S. Hardie, Chun-Yao Huang,

To cite this article:

Peter S. Fader, Bruce G. S. Hardie, Chun-Yao Huang, (2004) A Dynamic Changepoint Model for New Product Sales Forecasting. Marketing Science 23(1):50-65. <http://dx.doi.org/10.1287/mksc.1030.0046>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

© 2004 INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

A Dynamic Changepoint Model for New Product Sales Forecasting

Peter S. Fader

Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340,
faderp@wharton.upenn.edu

Bruce G. S. Hardie

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

Chun-Yao Huang

Department of Business Administration, 135 Yuan-Tung Road, Yuan Ze University, Jung-Li 32003, Taiwan,
cyhuang@saturn.yzu.edu.tw

At the heart of a new product sales-forecasting model for consumer packaged goods is a multiple-event timing process. Even after controlling for the effects of time-varying marketing mix covariates, this timing process is not a stationary one, which means the standard interpurchase time models developed within the marketing literature are not suitable for new products.

In this paper, we develop a dynamic changepoint model that captures the underlying evolution of the buying behavior associated with the new product. This extends the basic changepoint framework, as used by a number of statisticians, by allowing the changepoint process itself to evolve over time. Additionally, this model nests a number of the standard multiple-event timing models considered in the marketing literature.

In our empirical analysis, we show that the dynamic changepoint model accurately tracks (and forecasts) the total sales curve as well as its trial and repeat components and other managerial diagnostics (e.g., percent of triers repeating).

Key words: sales forecasting; trial/repeat; new product research; duration models; nonstationarity; changepoint models

History: This paper was received October 18, 2002, and was with the authors 4 months for 2 revisions; processed by Gary L. Lilien.

1. Introduction

For many new products, it is difficult to get an accurate read on their long-term potential based on only a few initial weeks of postlaunch (or test-market) sales data. Common problems include the following types of issues: (1) Significant promotional activity can artificially skew the initial sales levels, (2) early buyers may not exhibit typical purchasing rates, and (3) repeat-purchasing patterns may be hard to sort out from the voluminous amounts of first purchase (or trial) data (Ayal 1975, Lipstein 1961, Morgan 1979). It is therefore essential for practitioners to rely on formal models of new product sales to tease apart and understand each of these underlying components to create a valid sales forecast.

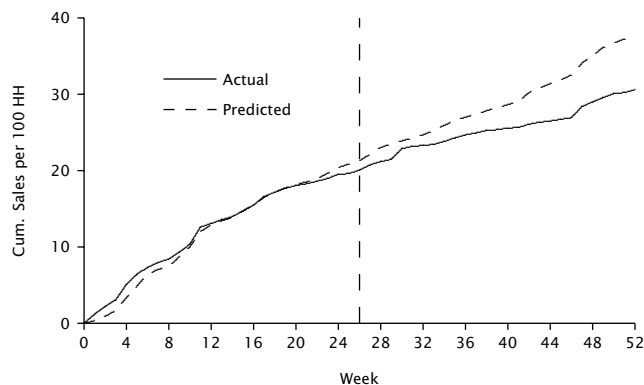
At the heart of a new product sales-forecasting model is a *multiple-event timing process*. For many behavioral processes besides new product sales forecasting, researchers need to capture a series of interpurchase cycles while accommodating customer heterogeneity. In addition, they need to filter out the influences that exogenous factors, such as promo-

tional activities, may exert within and across these multiple purchase cycles.

One of the best-known contributions in this general area is a highly regarded paper by Gupta (1991), which carefully laid out a general framework to capture the effects of time-varying explanatory variables in a multiple-event timing model. We will review some of the key technical aspects of Gupta's paper later, but for the moment we want to call attention to his main empirical result. In applying a broad set of models to scanner panel data for regular ground coffee, Gupta observed that the most appropriate specification was a model that featured an Erlang-2 interpurchase process with gamma-distributed purchase rates (to account for customer heterogeneity), while also allowing for time-varying covariates such as price and promotion.

Let us consider the case of a new juice product, with the masked name of "Kiwi Bubbles," which underwent a year-long test in two of IRI's BehaviorScan test markets prior to its national launch. Data on the purchasing of this new product were collected from 2,799 panelists. We fit Gupta's Erlang-

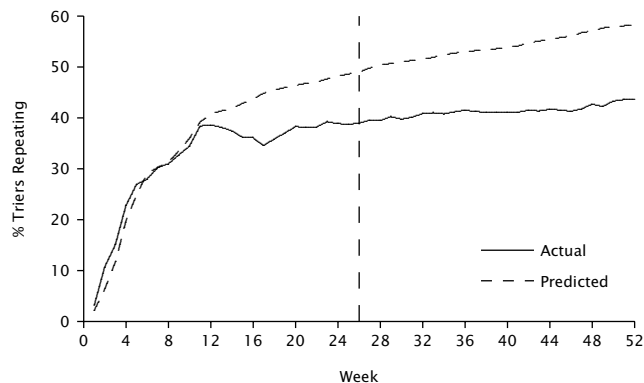
Figure 1 Sales Tracking Performance of Erlang-2/Gamma, Covariates Model



2/gamma, covariates model to the first 6 months of purchasing data and then generate a forecast of purchasing behavior for the remaining 6 months of the test market. As shown in Figure 1, this model does a fine job of tracking sales within the 26-week calibration period, but it quickly (and significantly) veers away from the actual sales pattern in a hold-out forecasting period. This substantial overprediction (the Week-52 forecast is 23% higher than the actual cumulative sales level) suggests that the interpurchase times tend to be lengthening (slowing down) as buyers gain more experience with the new product. This bias is quite typical of what we have observed for numerous new products and will motivate much of the model development that follows in the next section.

Beyond this poor forecast, the Erlang-2/gamma, covariates model is even more troubling from a diagnostic perspective. One of the commonly used managerial benchmarks for a new product is “percent triers repeating” (Clarke 1984, Rangan and Bell 1994), which provides a useful indication of how well the product has been accepted by its base of initial triers. In Figure 2, we see that the actual and predicted curves diverge quite drastically even in the middle of the calibration period. This means that the model

Figure 2 Tracking of Percent Triers Repeating by Erlang-2/Gamma, Covariates Model



is clearly not capturing the underlying purchase-to-purchase dynamics, even though the aggregate sales-tracking plot appears to behave acceptably well over the calibration period.

The main problem here is that the multiple-event timing process is not a stationary one. As customers move from one stage of purchasing to the next, there appears to be a shift occurring—in particular, early buyers are slowing down or dropping out completely. To be fair to Gupta (1991), he made no claims that his stationary models were suitable for new products and he conducted no empirical testing in this regard. However, this points out the need for such a model—one that builds upon the basic template laid out in Gupta’s paper, while allowing for the kinds of dynamics that typically occur in the new product setting.

Over the years, many researchers have dealt with some of these dynamics by developing separate models for trial and repeat behavior. The basic belief is that these two stages are behaviorally distinct, and that there is little or no “carryover” from the trial process that might improve (or otherwise affect) the performance of the repeat model. Therefore, while the trial-repeat decompositional approach has a long history in marketing (e.g., Blattberg and Golanty 1978, Eskin 1973, Fourt and Woodlock 1960, Parfitt and Collins 1968, Pringle et al. 1982), it may suffer from three potential shortcomings:

- *Separation of stages.* Does the evolution of buying patterns begin and end with the transition from the trial purchase to first repeat? Some buyers may continue their initial buying pattern through several “retrial” cycles before settling into a steady-state pattern (e.g., Aaker 1971). Other buyers may go through several evolutionary states (i.e., buying rates). In general, the repeat-buying component may be highly nonstationary, so a simple trial-repeat decomposition may be inadequate.

- *“Dependence” between stages.* As noted earlier, most traditional trial-repeat models do not allow any individual-level information about the first purchase process to come into play in the repeat model. As we transition from trial to repeat, the only piece of information we effectively retain about each person is that they made a trial purchase; we do not know when they made that purchase. However, it is known that early triers tend to be heavy category buyers (Morgan 1979, Taylor 1977); given “acceptance” of the new product, we could therefore expect them to have a higher repeat-buying rate than the later triers (who tend to be light category buyers). This failure to formulate a multiple-event timing model by first conditioning on the individual-level latent traits is known to result in biased inferences (Gupta and Morrison 1991).

- *Parsimony.* An integrated model requires fewer parameters than separate staged models, and may

therefore offer clearer, more interpretable diagnostic properties. Our most general model requires only four parameters (not including covariate effects), yet will be able to track and forecast the overall sales data extremely well. Furthermore, we will show that our model is also very effective at capturing each of the critical components (i.e., trial, first repeat, additional repeat), even though we do not develop an explicit, separate model for each of them.

A general class of models that can allow for some of these purchase dynamics is known as *change point* models. While they have not been very prevalent in the field of marketing, a number of statisticians have developed models that begin to allow for the type of ongoing evolution that may characterize new product sales patterns as shown and discussed above. These models are centered around an enduring, but stochastic, renewal process that lets the observation units (i.e., customers) change their parameters (buying rates) at random points over time.

We adopt this basic framework, and also extend it in a novel way to accommodate an important distinguishing aspect of the new product purchase process. We allow for the possibility that the change point process itself may evolve over time—hence, the name *dynamic change point* model. Specifically, our intuition and empirical observations suggest that change point occurrences (i.e., changes in buying rates) will be quite common in the early phase of a particular customer's relationship with the new product, but these changes will occur much less frequently after the buyer has gained considerable experience with the product. In fact, the change point process may at some stage terminate completely, which would leave each buyer with a permanent, steady-state (but still heterogeneous) purchasing rate. This steady-state scenario would be entirely consistent with the work of Gupta (1991), as well as that of many other researchers (Ehrenberg 1988; Jain and Vilcassim 1991, 1994; Morrison and Schmittlein 1988) who have shown the robustness and widespread nature of stationary (but heterogeneous) models for the purchasing of established products.

In the next section we introduce the general structure of a change point process as it applies to the case of new product purchasing. We then discuss the notion of a dynamic change point process and develop the general form of our model, which includes several relevant nested versions such as the models analyzed by Gupta (1991), as well as traditional (static) change point models. This is followed by an empirical analysis in which we examine the fit and forecasting performance of the proposed model for two datasets with different repeat-purchasing characteristics. We conclude with a discussion of several issues that arise from this work and identify several areas worthy of follow-up research.

2. Model Development

Our goal is to develop a model of new product buying behavior that incorporates the effects of marketing mix variables and nonstationarity in buying rates at the individual consumer level. We obtain estimates of the model parameters using data collected in a test market environment, then generate an overall sales forecast as well as related managerial diagnostics for the new product. These data come from a consumer panel in which we track the purchasing of each panelist for the duration of the test market.

We expect to observe some nonstationarity in the underlying buying rates (after controlling for the temporary effects of marketing mix variables). The buying rate reflects the consumer's preference for the new product, and we would expect there to be some instability in preferences as the consumer gains experience with it. For example, initial enthusiasm may diminish and the consumer will return to substitute products purchased before the launch of the new product; this would be reflected by a change to a lower value of the buying-rate parameter. This change might be extreme; i.e., the consumer drops the product from future consideration after the first purchase. Alternatively, the consumer may go through several "retrial" stages before deciding whether and how the product fits in with his regular purchasing habits.

As consumers gain more experience with the product, we would expect their preferences (and therefore their underlying buying rates) to stabilize to some extent. Consequently, a desired feature of the model is that it can capture the evolution towards a stationary repeat-buying process as the product moves from being "new" to "established."

It should be noted that preference changes need not be the only driver of changes in the underlying buying rates; external "shocks" to the market, perhaps due to word-of-mouth influences or other unmeasured covariate effects, can also lead to the occurrence of change points in the purchase sequence. However, it is not our goal to identify the specific drivers of each of the change points, but simply to capture these changes to make more accurate sales forecasts.

Before we formally develop our model of new product buying behavior, let us further explore the notion of nonstationarity. Suppose we have a sequence of observations y_1, y_2, \dots, y_N ordered in time, with each observation being independent with density $f(y_n | \lambda_n)$. If the sequence of observations is viewed as being stationary, we assume that $\lambda_n = \lambda \forall n$. If $\lambda_n \neq \lambda \forall n$, we say that the sequence is nonstationary.

According to the general change point model (e.g., Barry and Hartigan 1993, Henderson and Matthews 1993, Pievatolo and Rotondi 2000, Stephens 1994), this

nonstationarity is modeled by postulating that the sequence can be partitioned into $q + 1$ contiguous sets,

$$\{y_1, \dots, y_{w_1}\}, \{y_{w_1+1}, \dots, y_{w_2}\}, \dots, \{y_{w_{q-1}+1}, \dots, y_N\}.$$

Within each block, we assume $\lambda_n = \lambda_{w_i}$ for $w_{i-1} < n \leq w_i$, with changes in the parameter values occurring at w_1, w_2, \dots, w_q , which are called changepoints.

Within the marketing literature, the logic of change-point models has been combined with traditional finite mixture methods to allow changes in segment membership over time (e.g., Böckenholt and Dillon 1997, Poulsen 1990). This set of models typically falls under the heading of *hidden Markov models* (MacDonald and Zucchini 1997), which can be viewed as a specific subclass of changepoint process models. A hidden Markov model (HMM) is a discrete-time finite-state Markov model in which s_t , the state of the system at time $t = 1, 2, \dots$, is not directly observable. Rather, we observe y_t , which is linked to s_t by the stationary density $f(y_t | s_t)$. In many cases, including the model developed in this paper, we have a continuous-time process and/or there are an infinite number of possible latent states, in which case we look beyond the standard HMM framework to more general changepoint process models.

The standard changepoint problem is one of inference: Given a single data sequence, the objective is to identify the number (q) and location (w_1, \dots, w_q) of the changepoints, along with the values of the $\lambda_{w_1}, \dots, \lambda_{w_q}$. This differs from our modeling problem in two ways. First, given our interest in forecasting, we are not interested in identifying the precise times/instances of the changepoints in the observed data sequence. Rather, we need to determine whether there is any systematic structure in the pattern of the changepoints and then use this knowledge to create a forecast of the data sequence beyond y_N . Second, we are not modeling a single aggregate data sequence; rather, we have a separate sequence of observations for each panelist, which means we must explicitly capture the cross-sectional heterogeneity in the parameters, as well as the temporal variation.

We will deal with this second point by assuming that the λ_{w_i} are distributed across the population according to some heterogeneity distribution. The first point can be accommodated within the framework of a product partition model (Barry and Hartigan 1992, Howard 1965). In particular, we assume that the partition of the sequence y_1, y_2, \dots, y_N is randomly selected according to a product partition distribution. Instead of assuming a constant probability of a changepoint occurring at each y_n , which leads to a geometric distribution for the length of the blocks within a given partition (e.g., Barry and

Hartigan 1993, Howard 1965, Yao 1984), we will allow these probabilities to follow a flexible structure that, for example, allows the probability of change in the underlying stochastic process to decrease as the consumer gains more experience with the new product. Forecasts can then be generated conditional on a given partition of future observations, and these forecasts can then be weighted and combined according to the probability of observing such a partition.

During the period $(0, t_c]$, where 0 corresponds to the launch date of the new product and t_c is the censoring point that corresponds to the end of the model calibration period, we observe a consumer making $K = J + 1$ purchases of the new product at times t_0, t_1, \dots, t_J . (By convention, $j = 0$ corresponds to the trial purchase. We refer to j as the depth-of-repeat level.) We need to specify a model for the set of $K = J + 1$ interpurchase times $t_0, t_1 - t_0, \dots, t_J - t_{J-1}$ and the censored observation $t_c - t_J$.

Let the underlying interpurchase times follow an exponential baseline distribution with marketing mix covariate effects incorporated using the standard proportional hazards framework:

$$\begin{aligned} h(t_j; \lambda_j) &= \lambda_j e^{\mathbf{x}(t_j)\boldsymbol{\beta}} \\ &\equiv \lambda_j A(t_j), \end{aligned}$$

where $\mathbf{x}(t)$ is the vector of marketing covariates at time t and $\boldsymbol{\beta}$ the effects of these covariates.

By definition, the survivor function is given by

$$\begin{aligned} S(t_j | t_{j-1}; \lambda_j) &= \exp\left[-\int_{t_{j-1}}^{t_j} h(u; \lambda_j) du\right] \\ &= \exp\left[-\lambda_j \left(\int_0^{t_j} A(u) du - \int_0^{t_{j-1}} A(u) du\right)\right]. \end{aligned}$$

Assuming the time-varying covariates remain constant within each unit of time (e.g., a week in our empirical application),

$$\begin{aligned} \int_0^t A(u) du &= \delta_{t \geq 1} \sum_{i=1}^{\text{Int}(t)} A(i) + [t - \text{Int}(t)]A(\text{Int}(t) + 1) \\ &\equiv B(t), \end{aligned}$$

where $\delta_{t \geq 1} = 1$ if $t \geq 1$, 0 otherwise. The survivor function of the with-covariates interpurchase time distribution is therefore

$$S(t_j | t_{j-1}; \lambda_j) = \exp[-\lambda_j B(t_j, t_{j-1})], \quad (1)$$

where $B(t_j, t_{j-1}) = B(t_j) - B(t_{j-1})$. Because $B(t_j, t_{j-1})$ captures the effects of the covariates between t_{j-1} and t_j , the survivor function is not just a function of the covariate values at t_j ; it is a function of the covariate values for every time period (e.g., week) since the

last transaction. It follows that the pdf of the with-covariates interpurchase time distribution is

$$f(t_j | t_{j-1}; \lambda_j) = h(t_j; \lambda_j) S(t_j | t_{j-1}; \lambda_j) \\ = \lambda_j A(\tau_j) \exp[-\lambda_j B(t_j, t_{j-1})], \quad (2)$$

where τ_j is the time period (e.g., week) in which the j th repeat purchase occurred, defined as

$$\tau_j = \begin{cases} t_j & \text{if } t_j \text{ is integer} \\ \text{Int}(t_j) + 1 & \text{otherwise.} \end{cases}$$

Following the standard discrete changepoint formulation, in which we assume that the λ_j can change immediately after the occurrence of the event of interest (i.e., the j th repeat purchase),¹ there are 2^K possible partitions of this sequence of observations. Let there be $q \leq K$ changepoints, and let $\mathbf{w} = \{w_i\}$, $i = 1, \dots, q$, be the set of changepoints, where w_i corresponds to the depth-of-repeat level (i.e., repeat-purchase number) associated with the purchase occasion that immediately precedes a change in the buying-rate parameter. (If a change occurs immediately after the trial purchase, we have $w_1 = 0$.) Let γ_j be the probability that there is a change in a consumer's underlying buying rate following his j th repeat purchase. The probability of the partition \mathbf{w} is therefore

$$P(\mathbf{w}) = \prod_{j \in \mathbf{w}} \gamma_j \prod_{j \in \mathbf{I} - \mathbf{w}} (1 - \gamma_j), \quad (3)$$

where $\mathbf{I} = \{0, 1, \dots, J\}$.

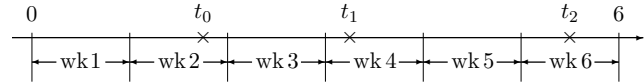
The standard assumption of product partition models is that $\gamma_j = \gamma \forall j$. However, in the case of a new product, we expect that the probability of a consumer revising his preferences following a purchase (with the corresponding change in the underlying buying rate) will *decrease* as he gains more experience with the new product (i.e., makes additional purchases). To accommodate this, we assume that

$$\gamma_j = 1 - \psi(1 - e^{-\theta(j+1)}), \quad j = 0, 1, 2, \dots, \quad (4)$$

where $\eta, \psi \in [0, 1]$ and $\theta \geq 0$. This parametric expression is motivated by equivalent expressions used in other papers (Eskin 1973, Fader and Hardie 2001, Kalwani and Silk 1980) to capture the evolution of repeat-buying patterns over time.

The relationship between γ_j and j is illustrated in Figure 3 for three sets of values of ψ and θ . We note that as j increases, γ_j tends to $1 - \psi$. Therefore, if $\psi = 1$, the probability of a change tends to zero as a consumer moves to higher depth-of-repeat levels; in

other words, the model evolves to a stationary process that would be consistent with the stabilization



of consumer preferences. On the other hand, if $\psi < 1$, $\gamma_j > 0 \forall j$, which means that individual consumer preferences will not stabilize; in other words, there is long-term nonstationarity in the marketplace. This general specification for γ_j results in what we call the *dynamic changepoint* model. If $\theta \rightarrow \infty$, then γ_j is independent of j and equals $1 - \psi \forall j$, and we call this the *static changepoint* model.

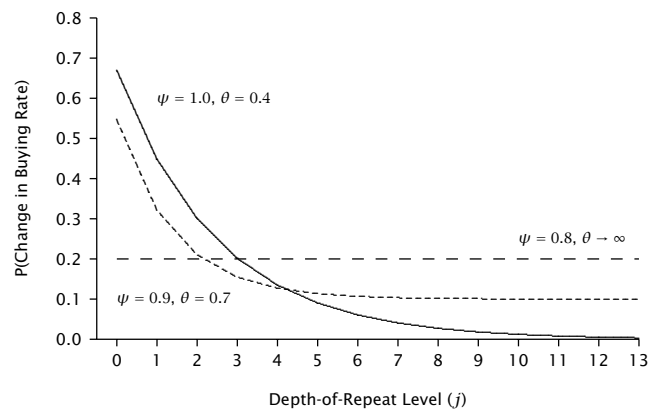
We incorporate heterogeneity in buying rates by allowing them to be distributed across the population according to a gamma distribution with shape parameter r and scale parameter α :

$$g(\lambda) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}.$$

When a change in the underlying buying rate occurs, the consumer draws a new value of his buying rate, independent of his previous one, from the same gamma distribution of purchase rates described above. This principle of independent sampling from a given mixing distribution was first raised in Howard's "dynamic inference" model (Howard 1965) and was used by Sabavala and Morrison (1981) in their model of media exposure. More recently, it has been used by Barry and Hartigan (1992) and Yao (1984), among others, in Bayesian analyses of the changepoint problem.

To illustrate and convey the intuition of the proposed model, let us consider the following scenario of a consumer who makes three purchases of the new product in the first 6 weeks of it being on the market: trial at t_0 , first repeat at t_1 , and second repeat at t_2 . Given t_0, t_1, t_2 , we do not know whether the consumer ever revised his preferences and, if he did, how many times and at which points in time. For this consumer,

Figure 3 Probability of Change in Buying Rate by Depth of Repeat



¹ One behavioral "story" consistent with this is to assume that consumption immediately follows purchase and that preference revisions would immediately follow consumption.

Table 1 Possible Partitions of Three Purchases

Partition	Changepoint Occurs After			No. of Changepoints
	Trial	1st Repeat	2nd Repeat	
(i)				0
(ii)	✓			1
(iii)		✓		1
(iv)			✓	1
(v)	✓	✓		2
(vi)	✓		✓	2
(vii)		✓	✓	2
(viii)	✓	✓	✓	3

the number of changepoints could have ranged from zero to three. The set of eight possible partitions of this purchase sequence is given in Table 1.

Let us first assume that the consumer never revised his preferences in $(0, 6]$ (i.e., the first partition). Given the constant underlying buying rate λ (i.e., before we consider cross-sectional heterogeneity), the conditional likelihood function for this consumer is the product of the density and survivor functions; that is,

$$L(\lambda, \beta) = f(t_0 | 0; \lambda) f(t_1 | t_0; \lambda) f(t_2 | t_1; \lambda) S(6 | t_2; \lambda) \\ = \lambda^3 A(2) A(4) A(6) \exp[-\lambda B(6, 0)].$$

The corresponding unconditional likelihood function is obtained by accounting for heterogeneity in the underlying buying rate λ :

$$L(r, \alpha, \beta) = \int_0^\infty L(\lambda, \beta) \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)} d\lambda \\ = \frac{\Gamma(r+3) A(2) A(4) A(6)}{\Gamma(r)} \\ \cdot \left(\frac{1}{\alpha + B(6, 0)} \right)^3 \left(\frac{\alpha}{\alpha + B(6, 0)} \right)^r. \quad (5)$$

This likelihood function is identical to the exp/gamma, covariates model from Gupta (1991).

Alternatively, suppose that the consumer revised his preferences following his second (first repeat) purchase (i.e., partition (iii) in Table 1, above). This partition is defined by $\mathbf{w} = \{1\}$. Let the purchasing rate λ_a reflect the consumer's initial preference for the new product, and λ_b reflect the consumer's revised preference following his first repeat purchase. Conditional on λ_a and λ_b , the likelihood function for this consumer is therefore

$$L(\lambda_a, \lambda_b, \beta) \\ = f(t_0 | 0; \lambda_a) f(t_1 | t_0; \lambda_a) f(t_2 | t_1; \lambda_b) S(6 | t_2; \lambda_b) \\ = \lambda_a^2 A(2) A(4) \exp[-\lambda_a B(t_1, 0)] \lambda_b A(6) \exp[-\lambda_b B(6, t_1)].$$

We note that the change in buying rate results in a new value being drawn from the same underlying

gamma distribution. The unconditional likelihood function is obtained by accounting for heterogeneity in λ_a and λ_b :

$$L(r, \alpha, \beta) \\ = \int_0^\infty \int_0^\infty L(\lambda_a, \lambda_b, \beta) \frac{\alpha^r \lambda_a^{r-1} e^{-\alpha\lambda_a}}{\Gamma(r)} \frac{\alpha^r \lambda_b^{r-1} e^{-\alpha\lambda_b}}{\Gamma(r)} d\lambda_a d\lambda_b \\ = \frac{\Gamma(r+2) A(2) A(4)}{\Gamma(r)} \left(\frac{1}{\alpha + B(t_1, 0)} \right)^2 \left(\frac{\alpha}{\alpha + B(t_1, 0)} \right)^r \\ \cdot \frac{\Gamma(r+1) A(6)}{\Gamma(r)} \left(\frac{1}{\alpha + B(6, t_1)} \right) \left(\frac{\alpha}{\alpha + B(6, t_1)} \right)^r. \quad (6)$$

Given the partition probability distribution defined by (3) and (4), the probability of observing the first partition with corresponding likelihood function (5) is $(1 - \gamma_0)(1 - \gamma_1)(1 - \gamma_2)$, while the probability of observing the third partition with likelihood function (6) is $(1 - \gamma_0)\gamma_1(1 - \gamma_2)$. The overall likelihood function for this consumer is computed by taking the weighted average of the likelihood function associated with each of the possible partitions, where the weights are the partition probabilities.

More generally, let $\mathbf{T}_h = \{t_0, t_1, \dots, t_J\}$ be the set of times at which consumer h , ($h = 1, \dots, H$), makes his K purchases of the new product in the period $(0, t_c]$. For convenience, we suppress the household subscript for t_0, \dots, t_J .

The exact nature of the likelihood function for consumer h depends on whether $K = 0$ or $K > 0$.

(i) If no purchase of the new product is observed (i.e., $K = 0$), this is because the consumer has not yet had the opportunity or need to make a trial purchase. Therefore, the likelihood function for a consumer making no purchases is

$$L(\mathbf{T}_h) = \left(\frac{\alpha}{\alpha + B(t_c, 0)} \right)^r, \quad (7)$$

which is simply the with-covariates survivor function (i.e., the probability that no purchase occurred in $(0, t_c]$) mixed with the gamma distribution.

(ii) When $K > 0$, the possibility of there being changepoints in the sequence of interpurchase times emerges. As we cannot tell exactly when (or if) changes in the buying rates take place, we first formulate the likelihood function conditional on a given partition \mathbf{w} with $n \leq K$ changepoints.

(a) For the case of no changepoints ($n = 0$), we have

$$L(\mathbf{T}_h | \mathbf{w}) = \left\{ \prod_{j=0}^J A(\tau_j) \right\} \frac{\Gamma(r+J+1)}{\Gamma(r)} \\ \cdot \left(\frac{1}{\alpha + B(t_c, 0)} \right)^{J+1} \left(\frac{\alpha}{\alpha + B(t_c, 0)} \right)^r. \quad (8)$$

(This is Gupta's 1991 exp/gamma, covariates model.)

(b) For $n > 0$ changepoints, we have

$$L(\mathbf{T}_h | \mathbf{w}) = \left\{ \prod_{j=0}^J A(\tau_j) \right\} \prod_{i=1}^n \left\{ \frac{\Gamma(r + w_i - w_{i-1})}{\Gamma(r)} \cdot \left(\frac{1}{\alpha + B(t_{w_i}, t_{w_{i-1}})} \right)^{w_i - w_{i-1}} \left(\frac{\alpha}{\alpha + B(t_{w_i}, t_{w_{i-1}})} \right)^r \right\} \cdot \left\{ \frac{\Gamma(r + J - w_n)}{\Gamma(r)} \left(\frac{1}{\alpha + B(t_c, t_{w_n})} \right)^{J - w_n} \cdot \left(\frac{\alpha}{\alpha + B(t_c, t_{w_n})} \right)^r \right\}. \quad (9)$$

Recalling the partition probability distribution defined by (3) and (4), the likelihood function for consumer h with $K > 0$ is simply the weighted average of the partition-specific likelihoods:

$$L(\mathbf{T}_h) = \sum_s L(\mathbf{T}_h | \mathbf{w}_s) P(\mathbf{w}_s), \quad (10)$$

where the summation is over the possible partitions indexed by $s = 1, 2, \dots, 2^K$. (For $K = 0$, the likelihood function is given by (7).) The overall sample log-likelihood function is given by

$$LL = \sum_{h=1}^H \ln[L(\mathbf{T}_h)]. \quad (11)$$

Equations (3), (4), and (7)–(11) define the model as fitted to a given dataset. Maximum likelihood estimates of the model parameters $(r, \alpha, \psi, \theta, \beta)$ are obtained by maximizing the log-likelihood function given in (11). Standard numerical optimization methods are employed, using the MATLAB programming language, to obtain the parameter estimates. When we observe a large number of repeat purchases for the new product, the evaluation of the model log-likelihood function can become computationally complex given the need to evaluate each of the 2^K possible partitions of the purchase sequence. A discussion of this issue and the examination of a recommended approach for minimizing computation time can be found in the appendix.

2.1. Properties of the Model

In its most general form, the model requires the estimation of $4 + p$ parameters, where p is the number of marketing covariates. It is a very flexible model that can capture many patterns of buying behavior. Examples of such buying phenomena include:

- “Traditional” stationary buying behavior. If $\gamma_j = 0 \forall j$ (i.e., $\theta \rightarrow \infty$ and $\psi = 1$), we have the exp/gamma, covariates model considered by Gupta (1991). With the additional constraint that $\beta = 0$, we have

the two-parameter exponential-gamma model of stationary repeat-buying behavior that is the timing counterpart of the NBD counting model (Gupta and Morrison 1991). The estimates of r and α would equal to those obtained by fitting the NBD model to the data.

- The transition from a “new” to “established” product. If $\psi = 1$ and θ is finite, then $\gamma_j \rightarrow 0$ as j increases; that is, the probability of seeing a change in the underlying buying rate tends to zero as a consumer moves to higher depth-of-repeat levels. This means that the initial nonstationary buying process evolves to a stationary process as the product becomes more established (i.e., when most buyers have made a large number of repeat purchases). Therefore, the model is consistent with the notion of nonstationary buying behavior during the early stages of a new product’s life and stationary buying behavior—as characterized by the NBD model—once it has become established in the marketplace.

- Long-term nonstationarity in repeat buying. When $\psi < 1$, the probability of there being a change in the underlying buying rate will *always* be nonzero, which means that the repeat buying process is always nonstationary. If $\theta \rightarrow \infty$, γ_j is a constant $1 - \psi$. For finite θ , $\gamma_j \rightarrow 1 - \psi$ as j increases; that is, the probability of a change in the underlying buying rate tends to the constant $1 - \psi$ as a consumer moves to higher depth-of-repeat levels. Such a model can easily capture the “leakage” of repeat buyers phenomena observed by East and Hammond (1996). If the underlying gamma distribution has an effective mode at zero ($r \leq 1$), an ongoing low level of changes will see some consumers drawing a very small value of λ (which yields an almost zero probability of making a repeat purchase in the foreseeable future) on a given change, thereby “dropping out” of the market for the product of interest. Other researchers (e.g., Schmittlein et al. 1987) have proposed NBD-based models that include a “death” process. However our model is far more flexible, allowing for other forms of nonstationarity (e.g., “speeding up” and “slowing down” of latent purchase rates) beyond a simple “death” process.

2.2. Generating Sales Forecasts

To evaluate the tracking performance of the proposed model, or to use the model for forecasting sales beyond the model calibration period, it is necessary to generate sales numbers (i.e., counts) from this timing model. We are interested in a number of sales-related measures for the new product: (i) the cumulative trial sales by time t , $T(t)$; (ii) the cumulative repeat sales by time t , $R(t)$; (iii) the total sales by time t , $S(t)$, which is by definition equal to $T(t) + R(t)$; and (iv) the depth-of-repeat components of repeat sales. Defining $R_j(t)$ as the number of consumers who have made

at least j repeat purchases of the new product by time t , we have $R(t) = \sum_{j=1}^{\infty} R_j(t)$. Our goal is to generate these numbers over the time interval $(0, t_f]$, where t_f denotes the end of the forecast period.

While we have a simple closed-form expression for expected cumulative trial sales,

$$E[T(t)] = H \cdot \left[1 - \left(\frac{\alpha}{\alpha + B(t, 0)} \right)^r \right],$$

it is not possible to write out a closed-form expression for $R(t)$, and consequently $S(t)$. We therefore propose a simulation-based approach to computing the sales numbers.

Let the nonzero elements of the vector \mathbf{N}_h denote the times at which consumer h made his trial, first repeat, etc. purchases (if any at all). For a given individual, we simulate the elements of \mathbf{N}_h in the following manner. We start by drawing a value of λ from the gamma distribution. Using this value of λ and the actual values of the covariates, we simulate an interpurchase time off the exponential-with-covariates interpurchase time distribution. This gives us the consumer's simulated value of t_0 , the time of his trial purchase. If $t_0 > t_f$, the consumer is deemed to have made zero purchases of the new product by time t_f and the procedure moves on to the next consumer. If $t_0 \leq t_f$, we record the time of this purchase ($N_h(0) = t_0$) and then draw a uniform random number to determine whether or not this purchase occasion corresponds to a changepoint: With probability $1 - \gamma_0$ the consumer retains his value of λ and with probability γ_0 he obtains a new value from the underlying gamma distribution. Another exponential-with-covariates interpurchase time is then simulated and added to t_0 to give us the consumer's simulated value of t_1 , the time of his first repeat purchase. If $t_1 > t_f$, the consumer is deemed to have made only a trial purchase by time t_f and the procedure moves on to the next consumer. If $t_1 \leq t_f$, we record the time of this first repeat purchase ($N_h(1) = t_1$), and the whole process continues for this consumer until $t_j > t_f$, at which time the procedure moves to the next consumer.

Once we have simulated \mathbf{N}_h for all individuals, we can compute total sales and its components in the following manner:

$$T(t) = \sum_{h=1}^H I\{0 < N_h(0) \leq t\},$$

$$R_j(t) = \sum_{h=1}^H I\{0 < N_h(j) \leq t\},$$

$$R(t) = \sum_{j=1}^{\infty} R_j(t),$$

$$S(t) = T(t) + R(t),$$

where $I\{\cdot\}$ is an indicator function that equals one if the logical condition is true, and zero otherwise.

We repeat this simulation, say 100 times, and take the average of the run-specific $S(t)$, $T(t)$, etc. This simulation-based approach is used in the following empirical analysis.

2.3. Extension to the Basic Model

As with numerous other stochastic models of buyer behavior, our model is based on the assumption that individual consumer interpurchase times can be characterized by the exponential distribution. Two potentially troubling characteristics of this distribution are that it is memoryless (i.e., there is no influence of time since the last purchase) and that the mode of the distribution is at zero (which means that the next purchase is most likely to occur immediately after the last one). Consequently, a number of researchers have proposed that the Erlang-2 distribution be used to model interpurchase times, as it allows for a more regular purchase process (Chatfield and Goodhardt 1973, Herniter 1971, Jeuland et al. 1980). We therefore consider the case of Erlang-2-distributed interpurchase times as an extension to the basic model.

Using Gupta's (1991) approach to incorporating the effects of time-varying covariates into the Erlang-2 distribution, the survivor function and pdf of the with-covariates interpurchase time distribution are given by

$$S(t_j | t_{j-1}; \lambda_j) = \exp[-\lambda_j B(t_j, t_{j-1})][1 + \lambda_j B(t_j, t_{j-1})],$$

$$f(t_j | t_{j-1}; \lambda_j) = \lambda_j^2 A(\tau_j) B(t_j, t_{j-1}) \exp[-\lambda_j B(t_j, t_{j-1})].$$

This results in new expressions for the partition-specific likelihood functions:

- For a consumer making no purchases in the calibration period $(0, t_c]$:

$$L(\mathbf{T}_h) = \left(\frac{\alpha}{\alpha + B(t_c, 0)} \right)^r \left[1 + \frac{r B(t_c, 0)}{\alpha + B(t_c, 0)} \right]. \quad (12)$$

- For $K > 0$ purchases with no changepoints ($n = 0$), we have

$$L(\mathbf{T}_h | \mathbf{w}) = \left\{ \prod_{j=0}^J A(\tau_j) B(t_j, t_{j-1}) \right\} \cdot \left\{ \frac{\Gamma(r + 2(J + 1))}{\Gamma(r)} \left(\frac{\alpha}{\alpha + B(t_c, 0)} \right)^r \cdot \left(\frac{1}{\alpha + B(t_c, 0)} \right)^{2(J+1)} \cdot \left[1 + \frac{(r + 2(J + 1)) B(t_c, t_J)}{\alpha + B(t_c, 0)} \right] \right\}. \quad (13)$$

(This is Gupta's 1991 Erlang-2/gamma, covariates model.)

- For $n > 0$ changepoints, we have

$$L(\mathbf{T}_h | \mathbf{w}) = \left\{ \prod_{j=0}^J A(\tau_j) B(t_j, t_{j-1}) \right\} \cdot \prod_{i=1}^n \left\{ \frac{\Gamma(r+2(w_i - w_{i-1}))}{\Gamma(r)} \left(\frac{\alpha}{\alpha + B(t_{w_i}, t_{w_{i-1}})} \right)^r \cdot \left(\frac{1}{\alpha + B(t_{w_i}, t_{w_{i-1}})} \right)^{2(w_i - w_{i-1})} \right\} \cdot \left\{ \frac{\Gamma(r+2(J - w_n))}{\Gamma(r)} \left(\frac{\alpha}{\alpha + B(t_c, t_{w_n})} \right)^r \cdot \left(\frac{1}{\alpha + B(t_c, t_{w_n})} \right)^{2(J - w_n)} \cdot \left[1 + \frac{(r+2(J - w_n))B(t_c, t_j)}{\alpha + B(t_c, t_{w_n})} \right] \right\}. \quad (14)$$

Equations (12)–(14) replace Equations (7)–(9), respectively. Consequently, Equations (3), (4), and (10)–(14) define the model as fitted to a given dataset when we assume the underlying interpurchase times follow the Erlang-2 distribution.

We explore the value of this extension in the following empirical analysis.

3. Empirical Analysis

We now examine the performance of the dynamic changepoint model (and its nested variants), using test market data for two new products that were tested using IRI's BehaviorScan service prior to national launch. The first new product is "Kiwi Bubbles," a masked name for a shelf-stable juice drink, aimed primarily at children, which is sold as a multipack of several single-serve containers. The second new product is "Four Seasons Biscuits," a masked name for a snack product.

Focusing first on the Kiwi Bubbles dataset, we begin by discussing the fit of the dynamic changepoint model and its variants in the calibration period. We then examine the out-of-sample forecasting performance of these models. This is followed by an exploration of the impact of the length of the model calibration period on model forecasting. To further illustrate and emphasize the performance of the dynamic changepoint model, we report the key results obtained from performing a similar analysis using the Four Seasons Biscuits dataset.

3.1. Model Fit

Prior to its national launch, the Kiwi Bubbles product underwent a year-long test conducted in two of IRI's BehaviorScan test markets. We use panel data drawn from 2,799 panelists in the two markets. (As is typically the case with new product launch data, we have

detailed purchasing information and marketing mix covariates only for the test brand.)

Using data for the 267 panelists who tried the new product by the end of Week 26, we wish to forecast the purchasing behavior of the *whole panel* (i.e., 2,799 panelists) to the end of the year (Week 52). That is, we fit the model specification to the first 6 months of purchasing data and generate sales forecasts for the whole year. We have information on the marketing activity over the 52 weeks the new product was in the test market; this comprises a standard market-level scanner data measure of promotional activity (i.e., %ACV with any merchandising) and a measure of coupon activity. (The coupon variable is an aggregate measure of coupon activity, generated by IRI for modeling purposes, that reflects the face value and circulation of the coupon along with standard decays in the redemption rate.)

As noted above, the dynamic changepoint model developed in this paper nests a number of simpler models. First, we can vary the nature of the changepoint process. Setting $\theta \rightarrow \infty$ results in the static changepoint model in which the probability of there being a change in the underlying buying rate is not allowed to evolve as consumers gain experience with the product. Additionally, setting ψ to 1.0 "turns off" the changepoint process and assumes that the underlying buying rates are stationary. This corresponds to the with-covariates models developed in Gupta (1991). Second, we can remove the covariate effects, which is equivalent to setting β to zero. Finally, we can replace the assumption of the underlying exponential distribution with the Erlang-2 distribution. We therefore have 12 different models' specifications to consider. We fit these 12 models to the first 6 months of purchasing data for the Kiwi Bubbles new product. The results are summarized in Table 2.

Looking at the model log-likelihoods, we immediately observe that the fit of the exponential model specifications strictly dominates their Erlang-2 counterparts. While this is contrary to the findings of Gupta (1991), this result is consistent with recent work on the modeling of trial purchasing for new grocery products, which finds strong support for the exponential distribution (Fader et al. 2003, Hardie et al. 1998).

Within the set of six exponential models, we observe that the inclusion of covariates results in a significant improvement in calibration-period model fit; however, the improvement in fit is less when we allow for more nonstationarity in the underlying buying rates: $\Delta LL = 79.4$ for the no-changepoint specification (Model 1 versus Model 4), $\Delta LL = 47.9$ for the static changepoint specification (Model 2 versus Model 5), and $\Delta LL = 45.4$ for the dynamic changepoint specification (Model 3 versus Model 6). For both the with- and without-covariates cases, the

Table 2 Model Estimation Results for the Kiwi Bubbles Dataset

Model	Baseline	Covariates	Changepoint	LL	# Parameters	r	α	ψ	θ	Coupon	Promotion
1	Exponential	N	None	−3,812.40	2	0.079	71.375	(1)	(∞)		
2	Exponential	N	Static	−3,779.19	3	0.049	26.797	0.750	(∞)		
3	Exponential	N	Dynamic	−3,771.98	4	0.047	24.057	0.851	1.144		
4	Exponential	Y	None	−3,733.00	4	0.076	138.239	(1)	(∞)	5.182	0.014
5	Exponential	Y	Static	−3,731.28	5	0.066	97.661	0.912	(∞)	5.059	0.012
6	Exponential	Y	Dynamic	−3,726.56	6	0.061	80.228	0.966	1.367	5.204	0.012
7	Erlang-2	N	None	−3,973.44	2	0.095	33.094	(1)	(∞)		
8	Erlang-2	N	Static	−3,797.52	3	0.044	7.243	0.637	(∞)		
9	Erlang-2	N	Dynamic	−3,783.96	4	0.044	6.937	0.812	0.836		
10	Erlang-2	Y	None	−3,824.23	4	0.091	74.370	(1)	(∞)	2.401*	0.019
11	Erlang-2	Y	Static	−3,760.54	5	0.051	17.086	0.707	(∞)	3.180**	0.010
12	Erlang-2	Y	Dynamic	−3,747.86	6	0.050	15.872	0.867	0.910	3.428**	0.010

Note. All parameter estimates are significantly different from 0 at $p < 0.001$, except those marked with * ($p < 0.05$) and ** ($p < 0.01$).

dynamic changepoint specification provides a significant improvement in fit over its no-change point counterpart. Compared to the Gupta (1991) exp/gamma, covariates specification (Model 4), our dynamic changepoint specification (Model 6) represents a significant improvement in calibration-period model fit ($p = 0.0016$ using the likelihood ratio test).

We also note that the dynamic changepoint model provides a significant improvement in fit over the static changepoint model, both in the case of the with-covariates specification ($p = 0.002$ using the likelihood ratio test) and the without-covariates specification ($p < 0.001$ using the likelihood ratio test).

Substituting the estimates of ψ and θ in (3) yields the implied probabilities of a change in the underlying buying rate as a function of depth-of-repeat level (i.e., experience with the product). These are plotted in Figure 4. For the dynamic changepoint specification (Model 6), we see that there is a 28% chance that a consumer will change his purchase rate following his trial purchase; this is entirely consistent with the widely held view that trial purchases are different from repeat purchases and that trial rates are a poor predictor of repeat buying. There is a 10% chance that

he will revise his buying rate following his first repeat purchase. This implies that there is a 65% chance that a consumer will not have revised his buying rate after two purchases of the new product. By the time three repeat purchases have been made, the change probability has effectively reached its asymptotic value of $1 - \psi = 3.5\%$.

3.2. Forecasting Performance

While the dynamic changepoint specification (Model 6) offers a statistically significant improvement in fit over both the static changepoint specification (Model 5) and the stationary specification (Model 4), the improvement in model fit is relatively small given the amount of data. However, it is dangerous to assess a forecasting model in terms of calibration-period fit, as the absence of a positive link between a model's calibration (in-sample) fit and out-of-sample forecasting performance is well known (Armstrong 2001).

We assess the forecasting performance of each model specification using several standard measures. First, we index the year-end (cumulative) sales forecast against actual Week 52 cumulative total sales (Wk52 Index); indices greater than 100 represent an overforecast. We also consider MAPE (mean absolute percentage error) computed over the 26-week forecast horizon. We compute MAPE for total sales (MAPE-Tot) as well as for the trial (MAPE-TR), first repeat (MAPE-FR), and additional repeat (MAPE-AR) components of total sales. (We focus on this decomposition because managers are interested in the breakdown of total sales into these components—see, for example, Clarke 1984.) These results are summarized in Table 3.

The dominance of the dynamic changepoint specification (Model 6) is now very clear, driven largely by its excellent performance in forecasting both of the repeat-sales components. Its forecasting performance is illustrated in Figure 5; observe that the model-based

Figure 4 Estimated Probability of Change in Buying Rate by Depth of Repeat

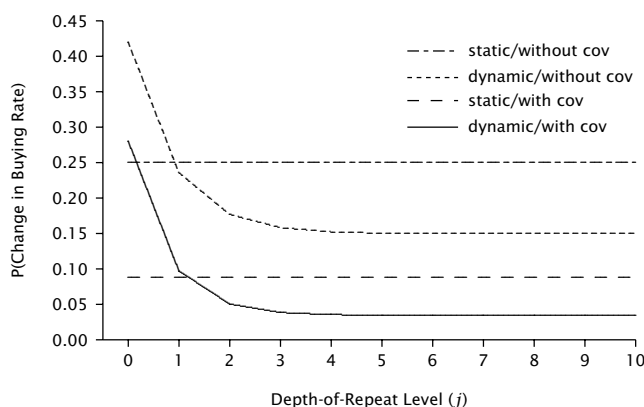


Table 3 Model Forecasting Performance Results for the Kiwi Bubbles Dataset

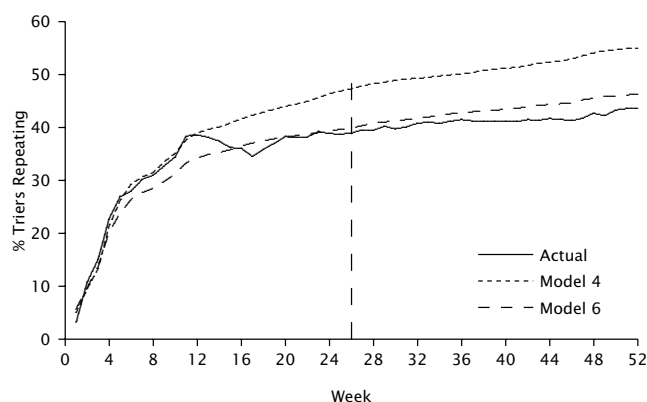
Model	Wk52 Index	MAPE-Tot	MAPE-TR	MAPE-FR	MAPE-AR
1	130.7	17.0	6.2	34.4	23.2
2	96.0	2.7	1.9	23.6	16.9
3	93.2	5.6	2.0	3.1	13.9
4	112.7	5.4	2.1	22.2	7.1
5	104.9	2.5	2.0	21.7	5.0
6	104.7	2.6	2.0	4.1	4.3
7	148.5	28.5	9.5	44.8	41.7
8	86.8	8.9	1.9	19.2	32.6
9	86.5	10.0	2.2	5.8	25.2
10	123.3	13.6	2.5	32.3	18.6
11	91.6	4.7	1.9	22.0	21.3
12	91.8	6.5	1.9	5.1	13.8

predictions provide an accurate tracking of both the total sales curve and its trial and repeat components. At year's end, this is only a 5% error in the overall sales forecast.

Even though the level of additional repeat sales is relatively low at the end of the calibration period, it is evident that additional repeat will quickly bypass the other sales components, and will comprise the lion's share of total sales in the period following Week 52. The ability of our model to accurately track and forecast this key component is, perhaps, the strongest indicator of its validity and usefulness.

Two other widely monitored measures of new product performance are "percent triers repeating" and "repeats per repeater" (Clarke 1984, Rangan and Bell 1994). At any point in time t , percent triers repeating is computed as $R_1(t)/T(t)$, while repeats per repeater is computed as $R(t)/R_1(t)$. In Figures 6 and 7 we compare the actual development of these two measures with the predictions derived from the dynamic change-point specification (Model 6) and the basic exponential-gamma, covariates specification (Model 4).

Comparing the tracks for Models 4 and 6 further illustrates performance of our dynamic change-point

Figure 6 Tracking Percent Triers Repeating

specification. To reinforce this, it is useful to contrast Model 6 with the performance of the Erlang-2/gamma, covariates specification (Model 10), both in terms of overall sales (Figure 1 versus Figure 5) and percent triers repeating (Figure 2 versus Figure 6). It is clearly important that a model incorporates a mechanism for accommodating changes in the underlying buying rates, reflecting, among other things, preference changes as the consumer gains more experience with the new product.

3.3. Exploring Sensitivity to Calibration-Period Length

In the analysis presented above, we have used a 6-month calibration period, as this is the point at which these forecasts are typically made in practice. However, given the costs of testing (and the opportunity costs of waiting too long), a manager conducting a test market often wants to know: "How many weeks are needed for an accurate forecast?" To explore the impact of the length of the model calibration period on forecasting performance, we recalibrated the dynamic change-point process model (i.e., Model 6) using two shorter calibration periods: (i) the first 12 weeks of purchasing data, and (ii) the first

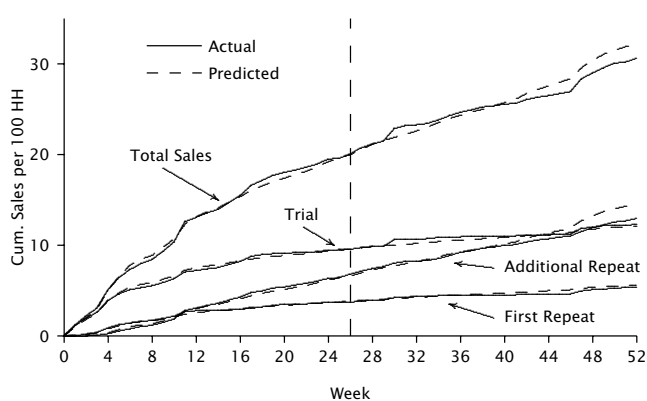
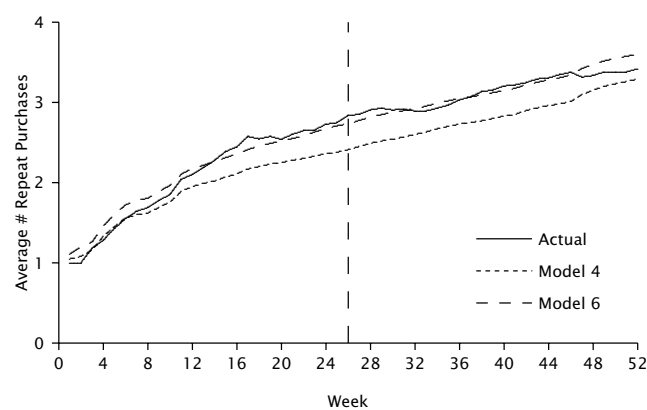
Figure 5 Predicted Sales**Figure 7** Tracking Repeats per Repeater

Table 4 Model Estimation Results by Length of Calibration Period

Length (wk)	LL (calib)	LL (26 wk)	r	α	ψ	θ	Coupon	Promotion
12	−2,387.34	−3,734.05	0.067	86.573	1.000	2.353	5.779	0.009
20	−3,343.71	−3,729.17	0.071	97.682	1.000	1.747	4.965	0.011
26	−3,726.56	−3,726.56	0.061	80.228	0.966	1.367	5.204	0.012

20 weeks of purchasing data. The model estimation results are reported in Table 4, along with the results associated with a 26-week calibration period (taken from Table 2). To facilitate the comparison across calibration period lengths, we applied the 12- and 20-week parameter estimates to the first 26 weeks of purchasing data and computed the associated value of the log-likelihood function, LL (26 wk).

We observe that even between 20 and 26 weeks, there is movement in the values of the parameter estimates and this is reflected in the almost 3-point difference in the associated value of the 26-week log-likelihood function. From a managerial perspective, the impact of calibration-period length (i.e., length of the test market) is of greater interest. The forecasting performance measures associated with the three calibration periods are therefore reported in Table 5.

We observe that while shortening the calibration period does not adversely impact our ability to track and forecast trial purchasing, it has a noticeable negative impact on the repeat-purchasing aspect of the model which leads to error in the total sales forecast. For the case of Kiwi Bubbles we can say that using a model calibration period shorter than the standard 26 weeks has a substantial impact on forecast quality, but we are not in a position to generalize.

Just as Fader et al. (2003) have explored the sensitivity of trial forecast accuracy to the length of model calibration period across a number of datasets from different product categories, there is a need to examine the impact of calibration period length on our forecasts of the repeat-purchase component of new product sales. Clearly, any conclusions about calibration-period length will be dependent, in part, on the average interpurchase time, which will vary from category to category. Researchers examining this issue in the future should therefore try to select datasets from categories that vary systematically on such a measure.

Table 5 Model Forecasting Performance Results by Length of Calibration Period

Length (wk)	Wk52 Index	MAPE-Tot	MAPE-TR	MAPE-FR	MAPE-AR
12	137.5	18.7	2.1	20.6	39.8
20	127.3	14.8	3.3	15.5	27.5
26	104.7	2.6	2.0	4.1	4.3

3.4. Additional Analysis of Model Performance

To further illustrate and emphasize the properties of the dynamic changepoint model, we briefly present the empirical results obtained from a similar analysis performed using the Four Seasons Biscuits dataset. Prior to its national launch, this snack product underwent a year-long test conducted in two of IRI's BehaviorScan markets, comprising a total of 3,953 panelists. However, the available data on repeat purchasing are much sparser than for Kiwi Bubbles, and thus this dataset represents a much tougher test for the dynamic changepoint model. First, the new product reached a relatively small number of triers by the end of Week 26 (148, compared to 267 for Kiwi Bubbles); second, the level of repeat purchasing per trier is much lower (1.06 repeats/trier over the 52 weeks, versus 1.79 repeats/trier for Kiwi Bubbles). The combination of these two factors leads to a dataset with only 77 repeat purchases in the calibration period (compared to 295 for Kiwi Bubbles). The marketing covariates provided with the dataset include the coupon and merchandising variables used in the Kiwi Bubbles analysis, along with a measure of TV advertising (GRPs transformed using a standard exponentially smoothed "stock" variable, e.g., Broadbent 1984 to account for carryover effects), and a "Week 1" dummy variable suggested by the firm commissioning the test to accommodate unusual marketing activity during the launch week.

Limiting ourselves to the exponential model specifications, we fit each of the six models to the first 26 weeks of purchasing data; the results are summarized in Table 6.

Once again, we observe the dominance of the dynamic changepoint specification (Model 6). In terms of forecast accuracy, the Week 52 index is 94.5. In terms of forecast period MAPE (computed over Weeks 27–52), the values for trial, first repeat, and additional repeat are 7.2, 9.8, and 15.5, respectively. Some of the errors in these sales components actually cancel out, yielding a total sales MAPE of only 2.9. In contrast to the case of Kiwi Bubbles, there is greater error in the repeat-sales component; this should come as no great surprise given the smaller number of repeat purchases per trier in the model calibration period ($77/148 = 0.52$ for Four Seasons Biscuits versus $295/267 = 1.10$ for Kiwi Bubbles).

It is interesting to compare the nature of the dynamic changepoint process between these two

Table 6 Model Estimation Results for the Four Seasons Biscuits Dataset

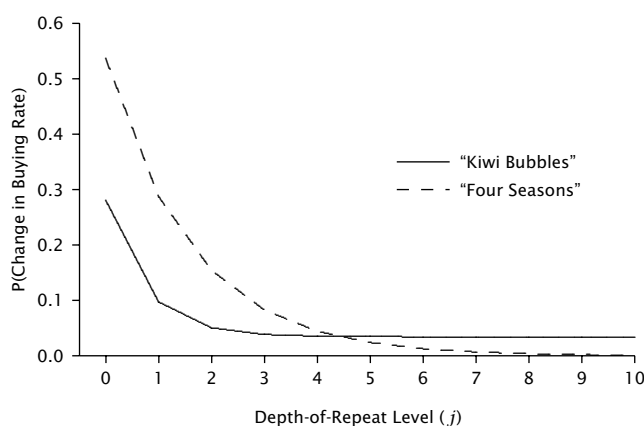
Model	Covariates	Changepoint	LL	# Parameters	r	α	ψ	θ	Any Merch	Dummy	TV	Coupon
1	N	None	-1,864.89	2	0.049	160.900	(1)	(∞)				
2	N	Static	-1,846.47	3	0.018	25.770	0.527	(∞)				
3	N	Dynamic	-1,839.73	4	0.017	23.576	1.000	0.494				
4	Y	None	-1,827.41	6	0.049	329.611	(1)	(∞)	0.014	0.913**	0.004 ^{ns}	4.668
5	Y	Static	-1,825.80	7	0.027	104.648	0.669	(∞)	0.013	0.433 ^{ns}	0.002 ^{ns}	2.910 [†]
6	Y	Dynamic	-1,819.98	8	0.023	68.693	1.000	0.623	0.013	0.231 ^{ns}	0.002 ^{ns}	2.352 ^{ns}

Note. All parameter estimates are significantly different from 0 at $p < 0.001$, except those marked with ^{ns} (not significant), [†] ($p < 0.10$), and ** ($p < 0.01$).

datasets. Substituting the two sets of estimates of ψ and θ in (3) yields the implied probabilities of a change in the underlying buying rate as a function of depth-of-repeat level, which are plotted in Figure 8. We observe a greater level of nonstationarity in the purchasing of Four Seasons Biscuits. There is a 54% chance that a consumer will change his purchase rate following his trial purchase and a 29% chance that he will revise his buying rate following his first repeat purchase. This implies that there is only a 33% chance that a consumer will not have revised his buying rate after two purchases of the new product (compared to 65% for Kiwi Bubbles). However, the parameter estimates suggest that the underlying purchasing evolves toward a stationary process ($\hat{\psi} = 1.0$), with a minuscule chance of a changepoint occurring anytime after the sixth repeat purchase.

4. Conclusions

While certain “hot” topics come and go in the field of marketing research, there has always been interest (shared by academics and practitioners alike) in the issue of forecasting new product sales. One characteristic of new product purchasing is the existence of nonstationarity in the buying process, even after we have controlled for the effects of time-varying marketing-mix covariates.

Figure 8 Estimated Probability of Change in Buying Rate by Depth of Repeat

Recognizing that existing approaches for handling these dynamics suffer from several shortcomings, we have proposed a dynamic changepoint model that captures the underlying evolution of the buying behavior associated with the new product. This model extends the basic changepoint framework, as used by a number of statisticians, by allowing the changepoint process itself to evolve over time. Additionally, this model nests a number of the standard multiple-event timing models that have previously been considered in the marketing literature. In our empirical analysis, we showed that the dynamic changepoint model accurately tracks (and forecasts) the total sales curve as well as its trial and repeat components and other managerial diagnostics (e.g., percent triers repeating).

Beyond the context discussed so far in the paper (i.e., two new grocery products), it is worth discussing other relevant applications/extensions for the general type of methodology presented here. First, it is important to emphasize that the behavioral “story” behind our model is by no means limited to the grocery products setting. A similar pattern will likely emerge for other types of products and services (although the specific parameters that characterize the various components of the model will likely vary from one context to another). Likewise, the model might apply nicely to new consumers who are first encountering an existing product/service. For instance, as people first gain broadband access to the Internet and learn about various Web sites, their behavior over time should conform to the basic set of assumptions outlined here; this would be a very promising area for future investigation.

As we run the model over multiple products/services, it will be instructive to look for “meta-patterns” in the resulting model parameters. Our empirical analysis revealed one particular type of nonstationary behavior, but it would be useful to catalogue different forms of nonstationarity (and covariate effects) and begin to associate them with product characteristics or other external measures. Many firms (e.g., ACNielsen BASES) attempt to database hundreds or thousands of products using simple sales summaries to enable early forecasts for future launches. Such a process can be greatly enhanced by

using the parameters from a complete (and behaviorally plausible) model rather than relying strictly on summary statistics (such as repeats per repeater and the other measures we discussed earlier). As our field continues to make rapid advances with hierarchical Bayes methods, this task should become a workable possibility, even for practitioners, in the near future.

Reflecting on the assumptions underlying this model, the least appealing one is that when a change in the underlying buying rate occurs, the consumer's new buying rate is independent of his previous one. A more plausible mechanism for modeling the evolution in the buying rate would be to assume that $\lambda_{w_{i+1}}$ equals λ_{w_i} modified by a random component (e.g., Chernoff and Zacks 1964, Moe and Fader 2003). However, such a mechanism is considerably more complex from a computational perspective and we leave the examination of this as an area of future research. It is worth noting that if this random shock was linked to the specific store environment in which each purchase of the new product occurred, we could accommodate Anderson and Simester's (2003) observation that deeper promotions increase repeat rates among first-time customers.

One potentially important limitation of the model developed in this paper is that it does not consider distribution build. A unique characteristic of controlled test market services such as BehaviorScan is the fact that they ensure complete retail distribution for the product of interest. In using this model in other test market settings, it would be desirable to modify the underlying formulation to accommodate the fact that retail distribution should be treated as a bottleneck that can prevent the trial and repeat purchasing of the product. (It is therefore inappropriate to simply treat distribution in the same way as other marketing covariates such as feature and display activity.) Having made such a modification, it would be natural to consider using the model for tracking a new product's rollout across geographic regions. In such an application, we should also account for the endogeneity of distribution (e.g., Bronnenberg and Mela 2003, Elberse and Eliashberg 2003).

Another potentially important limitation of our model is the absence of any competitive effects. It is interesting to think about how new product entry can affect—and be affected by—existing market structures. (See Bronnenberg et al. 2000 for a recent review of this literature.) Also, we may desire additional diagnostic measures that provide insight into the source of volume for the new product (e.g., Chatterjee and Ramaswamy 1996). However, beyond these past approaches—mostly post hoc econometric models that were not intended for forecasting purposes—that other researchers have employed, we are intrigued by an extension of our product-specific

stochastic model to one that can simultaneously deal with the sales patterns and marketing activities of competitors.

Beyond the simple case of expressing the measures of marketing activity relative to competition, one approach would be to build on a competing hazards framework. Alternatively, a dynamic discrete choice model could be used to model product choice given incidence, with a nonstationary timing process being used to model category purchasing. So, while we view our integrated model as offering a reasonably accurate and managerially useful picture of the trial-repeat process for a given new product, we see it as just one step towards the creation of a "Holy Grail" model that builds in competition and other category-level phenomena to be able to anticipate the complete set of market dynamics that surround a new product launch. However, we must realize that when the forecasting of new product sales is the modeling objective, the competitive dimension may increase the range/magnitude of potential errors, since we need to condition the sales forecasts on forecasts of competitive actions (e.g., Alsem et al. 1989, Danaher 1994). (The quality of these forecasts depends on the quality of the manager's judgements concerning future competitive activity.) The ultimate performance of the model speaks for itself: The ability of our "noncompetitive" model to forecast sales (and related managerial diagnostics) so accurately suggests that adding competitive effects may do little to improve pure sales forecasting accuracy.

Acknowledgments

The second author acknowledges the support of the London Business School Centre for Marketing.

Appendix: Issues in Parameter Estimation

At first glance, the task of maximizing the likelihood function given in (11) appears to be a straightforward numerical optimization problem. However, looking closely at (10), we note that each evaluation of the likelihood function involves the evaluation of 2^K partition-specific likelihood functions for each consumer, where K is the number of purchases made by the consumer. For relatively long model calibration periods (e.g., 52 weeks) and/or categories with high average purchase frequencies, it is not uncommon to observe consumers with 15 or more repeat purchases. For a consumer making 15 purchases, the likelihood functions associated with each of the $2^{15} = 32,786$ separate changepoint patterns have to be evaluated. For average-sized panels, each evaluation of the overall likelihood function (11) therefore involves millions of computations, and the overall computation time associated with maximizing the likelihood function may become impractical.

One approach to minimizing computation time is to limit the number of possible changepoints for each consumer to M . If the number of purchases made by a consumer (K) is less than M , all the possible changepoint sets for that number of purchases are evaluated. However, for a consumer

making $K > M$ purchases, the number of changepoint sets with M or fewer changes in the latent purchase rate is

$$\sum_{i=0}^M \binom{K}{i}.$$

It is important to note that we are not placing a limit on *when* the changepoints occur (i.e., their location); rather, we are limiting how many in total can occur. For example, if $M = 2$, the changepoints could occur immediately after trial and first repeat, trial and second repeat, first repeat and fifth repeat, and so on; or only once, after trial, or third repeat, and so on; or there could be no changepoints at all.

When this limit is placed on the number of changepoints, the summation in (10) is no longer over all possible partitions of the observed purchase sequence but over all partitions with $q \leq M$ changepoints. Consequently, the $P(\mathbf{w}_s)$ terms for each of the allowable partitions must be rescaled so that they sum to one. More specifically, let Ω be the set of partitions (\mathbf{w}_s) such that $|\mathbf{w}_s| \leq M$. Equation (10) becomes

$$L(\mathbf{T}_h) = \frac{\sum_{s \in \Omega} L(\mathbf{T}_h | \mathbf{w}_s) P(\mathbf{w}_s)}{\sum_{s \in \Omega} P(\mathbf{w}_s)}.$$

Let us say $M = 4$. For a consumer making 15 purchases in the calibration period, each evaluation of the likelihood function would see us evaluating 1,941, instead of 32,786, separate partition-specific likelihood functions. The computational savings are even more apparent for those consumers with a larger number of purchases. For a consumer making 20 purchases in the calibration period, setting M to 4 would reduce the number of partition-specific likelihood functions evaluated from $2^{20} = 1,048,576$ to 6,196—a 99.4% reduction in the number of evaluations.

Our experience to date shows that for relatively small values of M , this truncation in the number of possible changepoints has an inconsequential effect on the parameter estimates. To illustrate, we reestimated Model 6 (the dynamic changepoint model with an exponential baseline) on the Kiwi Bubbles dataset, varying the maximum number of (latent) changepoints in the given purchase sequence (M) from 1 to 4. For each set of parameter estimates, we compute the value of the log-likelihood function with no constraints placed on the maximum number of renewals. These results are reported in Table A, along with the results for the case where no limit is placed on the number of possible changepoints and the case where no changepoints are allowed (i.e., Model 4); the parameters for these two cases are taken from Table 2.

We observe that limiting the number of changepoint opportunities for each consumer to four yields effectively the same parameter estimates as those obtained when no

limit is placed on the number of possible changepoints, with no difference in the resulting value of the log-likelihood function. (At the margin, we could probably get away with limiting the maximum number of changepoints to three.) It is interesting to note, however, that at least two changepoints seem to be required, indicating the need to accommodate the underlying nonstationarity in buying rates for a new product.

References

- Aaker, David A. 1971. The new-trier stochastic model of brand choice. *Management Sci.* 17(April) B435–B450.
- Alsem, Karel J., Peter S. H. Leeflang, Jan C. Reuyl. 1989. The forecasting accuracy of market share models using predicted values of competitive marketing behavior. *Internat. J. Res. Marketing* 6 183–198.
- Anderson, Eric, Duncan Simester. 2003. Long-run effects of promotion depth on new versus established customers: Three field studies. *Marketing Sci.* 23(1) 4–20.
- Armstrong, J. Scott. 2001. Evaluating forecasting methods. J. Scott Armstrong, ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer Academic Publishers, Norwell, MA, 443–472.
- Ayal, Igal. 1975. Simple models for monitoring new product performance. *Decision Sci.* 6(April) 221–236.
- Barry, Daniel, J. A. Hartigan. 1992. Product partition models for change point problems. *Ann. Statist.* 20(March) 260–279.
- Barry, Daniel, J. A. Hartigan. 1993. A Bayesian analysis for change point problems. *J. Amer. Statist. Assoc.* 88(March) 309–319.
- Blattberg, Robert, John Golanty. 1978. TRACKER: An early test-market forecasting and diagnostic model for new-product planning. *J. Marketing Res.* 15(May) 192–202.
- Böckenholt, Ulf, William R. Dillon. 1997. Some new method for an old problem: Modeling preference changes and competitive market structures in pretest market data. *J. Marketing Res.* 34(February) 130–142.
- Broadbent, Simon. 1984. Modelling with adstock. *J. Market Res. Soc.* 26(4) 295–312.
- Bronnenberg, Bart J., Carl F. Mela. 2003. Market roll-out and retail adoption for new brands. Working paper.
- Bronnenberg, Bart J., Vijay Mahajan, Wilfried R. Vanhonacker. 2000. The emergence of market structure in new repeat-purchase categories: The interplay of market share and retailer distribution. *J. Marketing Res.* 37(February) 16–31.
- Chatfield, C., G. J. Goodhardt. 1973. A consumer purchasing model with Erlang inter-purchase times. *J. Amer. Statist. Assoc.* 68(December) 828–835.
- Chatterjee, Rabikar, Venkatram Ramaswamy. 1996. An extended mover-stayer model for diagnosing the dynamics of trial and repeat for a new brand. *Appl. Stochastic Models Data Anal.* 12(September) 165–178.
- Chernoff, H., S. Zacks. 1964. Estimating the current measure of a normal distribution which is subject to changes in time. *Ann. Math. Statist.* 35(September) 999–1018.
- Clarke, Darral G. 1984. G. D. Searle & Co.: Equal low-calorie sweetener (B). Harvard Business School Case 9-585-011, Harvard University, Cambridge, MA.
- Danaher, Peter J. 1994. Comparing naive with econometric market share models when competitors' actions are forecast. *Internat. J. Forecasting* 10 287–294.
- East, Robert, Kathy Hammond. 1996. The erosion of repeat-purchase loyalty. *Marketing Lett.* 7(March) 163–171.
- Ehrenberg, A. S. C. 1988. *Repeat Buying*, 2nd ed. Charles Griffin, London, U.K.

Table A Calibration Results

M	LL	r	α	ψ	θ	Coupon	Promotion
0	−3,733.00	0.076	138.239	(1)	(∞)	5.182	0.014
1	−3,730.47	0.056	67.021	0.843	2.086	4.990	0.010
2	−3,728.04	0.057	65.722	0.893	1.450	5.098	0.010
3	−3,726.61	0.060	77.202	0.953	1.383	5.181	0.011
4	−3,726.56	0.061	79.952	0.964	1.369	5.202	0.011
No limit	−3,726.56	0.061	80.228	0.966	1.367	5.204	0.012

- Elberse, Anita, Jehoshua Eliashberg. 2003. Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Sci.* **22**(4) 329–354.
- Eskin, Gerald J. 1973. Dynamic forecasts of new product demand using a depth of repeat model. *J. Marketing Res.* **10**(May) 115–129.
- Fader, Peter S., Bruce G. S. Hardie. 2001. Forecasting repeat sales at CDNOW: A case study. *Interfaces* **31**(May–June) S94–S107.
- Fader, Peter S., Bruce G. S. Hardie, Robert Zeithammer. 2003. Forecasting new product trial in a controlled test market environment. *J. Forecasting*. **22**(August) 391–410.
- Fourt, Louis A., Joseph W. Woodlock. 1960. Early prediction of market success for new grocery products. *J. Marketing* **25**(October) 31–38.
- Gupta, Sunil. 1991. Stochastic models of interpurchase time with time-dependent covariates. *J. Marketing Res.* **28**(February) 1–15.
- Gupta, Sunil, Donald G. Morrison. 1991. Estimating heterogeneity in consumers' purchase rates. *Marketing Sci.* **10**(3) 264–269.
- Hardie, Bruce G. S., Peter S. Fader, Michael Wisniewski. 1998. An empirical comparison of new product trial forecasting models. *J. Forecasting* **17**(June–July) 209–229.
- Henderson, R., J. N. S. Matthews. 1993. An investigation of change-points in the annual number of cases of haemolytic uraemic syndrome. *Appl. Statist.* **42**(3) 461–471.
- Herniter, Jerome. 1971. A probabilistic market model of purchase timing and brand selection. *Management Sci.* **18**(12) P102–P113.
- Howard, Ronald A. 1965. Dynamic inference. *Oper. Res.* **13**(September–October) 712–733.
- Jain, Dipak C., Naufel J. Vilcassim. 1991. Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Sci.* **10**(1) 1–23.
- Jain, Dipak C., Naufel J. Vilcassim. 1994. Estimating household purchase rates for consumer nondurable goods. *Appl. Stochastic Models Data Anal.* **10**(March) 15–26.
- Jeuland, Abel P., Frank M. Bass, Gordon P. Wright. 1980. A multi-brand stochastic model compounding heterogeneous Erlang timing and multinomial choice processes. *Oper. Res.* **28**(2) 255–277.
- Kalwani, Manohar, Alvin J. Silk. 1980. Structure of repeat buying for new packaged goods. *J. Marketing Res.* **17**(August) 316–322.
- Lipstein, Benjamin. 1961. Tests for test marketing. *Harvard Bus. Rev.* **39**(March–April) 74–77.
- MacDonald, Ian L., Walter Zucchini. 1997. *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman & Hall, London, U.K.
- Moe, Wendy W., Peter S. Fader. 2003. Capturing evolving visit behavior in clickstream data. *J. Interactive Marketing* **18**(Winter) 5–19.
- Morgan, Fred W., Jr. 1979. Are early triers heavy users? *J. Bus.* **52**(July) 429–434.
- Morrison, Donald G., David C. Schmittlein. 1988. Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *J. Bus. Econom. Statist.* **6**(April) 145–159.
- Parfitt, J. H., B. J. K. Collins. 1968. Use of consumer panels for brand-share prediction. *J. Marketing Res.* **5**(May) 131–145.
- Pievatolo, Antonio, Renata Rotondi. 2000. Analysing the interevent time distribution to identify seismicity phases: A Bayesian nonparametric approach to the multiple-changepoint problem. *Appl. Statist.* **49**(4) 543–562.
- Poulsen, Carsten Stig. 1990. Mixed Markov and latent Markov modelling applied to brand choice behaviour. *Internat. J. Res. Marketing* **7**(1) 5–19.
- Pringle, Lewis G., R. Dale Wilson, Edward I. Brody. 1982. NEWS: A decision-oriented model for new product analysis and forecasting. *Marketing Sci.* **1**(1) 1–29.
- Rangan, V. Kasturi, Marie Bell. 1994. Nestlé refrigerated foods: Contadina pasta & pizza (A). Harvard Business School, Case 9-595-035, Harvard University, Cambridge, MA.
- Sabavala, Darius J., Donald G. Morrison. 1981. A nonstationary model of binary choice applied to media exposure. *Management Sci.* **27**(June) 637–657.
- Schmittlein, David C., Donald G. Morrison, Richard Colombo. 1987. Counting your customers: Who they are and what will they do next? *Management Sci.* **33**(January) 1–24.
- Stephens, D. A. 1994. Bayesian retrospective multiple-changepoint identification. *Appl. Statist.* **43**(1) 159–178.
- Taylor, James W. 1977. A striking characteristic of innovators. *J. Marketing Res.* **14**(February) 104–107.
- Yao, Yi-Ching. 1984. Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *Ann. Statist.* **12**(December) 1434–1447.