A Bayesian Model for Prelaunch Sales Forecasting of Recorded Music

Jonathan Lee • Peter Boatwright • Wagner A. Kamakura

Kelley School of Business, Indiana University, SPEA/BUS 4041, 801 W. Michigan Street, Indianapolis, Indiana 46202
Graduate School of Industrial Administration, Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213
Fuqua School of Business, Duke University, Box 90120, Durham, North Carolina 27708-0120
jonalee@iupui.edu • pbhb@andrew.cmu.edu • kamakura@mail.duke.edu

---

In a situation where several hundred new music albums are released each month, producing sales forecasts in a reliable and consistent manner is a rather difficult and cumbersome task. The purpose of this study is to obtain sales forecasts for a new album before it is introduced. We develop a hierarchical Bayesian model based on a logistic diffusion process. It allows for the generalization of various adoption patterns out of discrete data and can be applied in a situation where the eventual number of adopters is unknown. Using sales of previous albums along with information known prior to the launch of a new album, the model constructs informed priors, yielding prelaunch sales forecasts, which are out-of-sample predictions. In the context of new product forecasting before introduction, the information we have is limited to the relevant background characteristics of a new album. Knowing only the general attributes of a new album, the meta-analytic approach proposed here provides an informed prior on the dynamics of duration, the effects of marketing variables, and the unknown market potential. As new data become available, weekly sales forecasts and market size (number of eventual adopters) are revised and updated. We illustrate our approach using weekly sales data of albums that appeared in *Billboard's* Top 200 albums chart from January 1994 to December 1995.
(*Forecasting; Empirical Generalization; Hierarchical Bayes Model*)

Introduction

The business of making and selling records provides a clear illustration of one of the balancing principles in business: potential profit versus potential risk. While there are considerable potential profits in the music industry, there are attendant risks at every step, risks that can be very costly to the success of the recording project (Fink 1996). For example, if mistakes in judgment are made regarding promotional planning based on an erroneous a priori sales projection, the sales of the record may never get off the ground. From the standpoint of the new product development process, sales forecasts play an important role in different stages, including design, testing, and life-cycle management (Urban and Hauser 1993).

One important problem for forecasting is the availability of relevant data. In cases where sales forecasts are made for commodities, data on past sales would be highly informative. At the other extreme, there is no past data for a unique new invention. For products such as movies, songs, and books, it is not entirely clear whether or when past data will provide relevant information for sales forecasts. What is needed in such a case is an empirical generalization, a pattern or regularity that repeats over different circumstances (Bass 1995, Barwise 1995), for a pattern across sales

0025-1909/03/4902/0179$5.00
1526-5501 electronic ISSN

datasets of a diverse set of past products can provide a starting point for forecasts of yet another seemingly unrelated item.

The Bayesian framework, which has a prior and an updating rule, naturally fits the structure of such a forecasting problem, in which we wish to incorporate various data sources for sales prediction prior to product launch, and later update those predictions as data become available. Here, the empirical generalization derived from potentially unrelated products would serve as the prior, on which one can base initial management decisions such as promotion activity and production planning. In their review of estimation techniques for diffusion models, Putsis and Srinivasan (2000) judged Hierarchical Bayes (HB) methods to have an important role in the diffusion literature, and they viewed the HB methodology as "underutilized in diffusion research." Indeed, there are few applications of HB methodology in the diffusion literature to date. Neelamegham and Chintagunta (1999) have used a hierarchical structure to forecast first-week sales of movies in international markets. Very recently, Talukdar et al. (2002) study diffusion of six products in a large number of industrialized and developing countries. Their work highlights the advantages of HB methodology in forecasting product sales, where the gains of the HB methods are greatest at the early stage of product introduction, when forecasts are often the most valuable.

In our empirical illustration, we use historical sales data from *Billboard's* Top 200 albums (January 94–December 95) to identify patterns of market penetration and to estimate eventual number of adopters (market potential). To increase the accuracy of prerelease forecasts, our model relates the diffusion patterns to planned promotional activities such as radio airplay and to background characteristics of the album/artist, including the music genre and track record of an artist. The predictions for a new album will be based on the exogenous variables and can be updated as new market information becomes available.

Our model will provide prelaunch weekly forecasts of sales of individual albums which are updated postlaunch as sales data become available. The model will also offer an estimate of the total lifetime sales of the product (market potential). For the individual album level of the analysis, we use a discrete-hazards model (Allison 1995, Kalbfleish and Prentice 1980), in which we incorporate the effect of covariates/explanatory variables on the sales process.

Discrete-hazard models are appropriate for studies that utilize aggregate data, where the exact timing of purchase is unknown. In our application, for instance, we know the sales that took place during a given week, but we do not know the timing of the sales within each week. Continuous models assume that the timing of the purchase event to be known as would be the case in a typical panel dataset.[1] We use a discrete-hazard function that is flexible enough to allow for various adoption patterns. The parameters of the hazard function are related in the second level of the hierarchy to a set of exogenous variables on product characteristics, leveraging the information from historical data to provide forecasts of new albums (Farley et al. 1995, Bayus 1993). As the marketing literature to date has relied on continuous hazards, our approach offers an example of a different model, one that is particularly relevant for the many marketing applications that rely on aggregate data.

Our model of album diffusion entails a hierarchical structure, in which album-specific hazard parameters are estimated along with parameters of a second function that links album-specific parameters to a second dataset. Lenk and Rao (1990) used a hierarchical structure to provide forecasts of durable goods as an extension of the Bass model, predicting in their conclusions that exogenous variables in the second layer of the hierarchy could greatly enhance prelaunch predictions. Neelamegham and Chintagunta (1999) have used a hierarchical structure with exogenous variables in their second layer to forecast first-week sales of movies in various international markets. In

[1] Consider for example, a medical study in which the event is onset of a particular disease in a sample of 20 rats. In such studies, the exact time of the event is often known for each individual subject. If onset occurred at some point in a long interval of time, like a week, the probability density must be integrated over the time interval in order to properly account for the uncertainty in the event timing, leading to a discrete hazard.

contrast to Neelamegham and Chintagunta's (1999) paper, which is not a hazard model, we explicitly incorporate covariates into a hazard model structure. In addition, we extend their approach to estimate weekly sales forecasts. In contrast with movies, for which the first weekend box office gross is a sufficient indicator of performance, effective planning of production and promotion of albums requires months-ahead projections of the sales evolution. The hazard specification adopted here can accommodate more flexible hazard shapes compared to, for example, the exponential model used in Sawney and Eliashberg (1996). Considering that our main goal is to obtain a priori forecasts, our hazard approach is different from the traditional ones because it is applied when the size of a risk set, the eventual number of adopters, is unknown. Therefore, we need to recognize the interaction between adoption pattern and market potential, and our model forecasts both jointly for a range of music albums before they enter the market.

What we can tell about the future is largely a result of the cumulative knowledge we have gained from past experiences. The overall performance of the proposed model depends heavily on the existence of repeated patterns in the data, which can only be ascertained after the estimation stage, when empirical generalizations are produced. Since time-invariant album/artist characteristics are used in finding a similar experience from the past for a new album, the ability to produce good initial forecasts for a new album with no sales data will depend heavily on the information content of the background characteristics as well as on the soundness of the structural framework one uses in interpreting historical information.

The purpose of this paper is to develop an approach based on empirical generalization, in which past experiences with potentially dissimilar products are utilized to produce sales forecasts and adoption patterns for new products prior to introduction. We formally implement the empirical generalization by way of a hierarchical Bayes model, which assists marketing managers by providing prelaunch forecasts so that they can design their promotional strategies and plan production and distribution of new products.

Model Development

Forecasting Model

We use a hierarchical model, where sales of individual albums are fit with a discrete-hazard function for grouped duration data, the parameters of which are related in a second level in the hierarchy to album/artist characteristics. Prelaunch forecasts of the adoption pattern and the market potential of a new album are based on the initial model parameters obtained via the second level of the hierarchical model and, as new data become available, forecasts are revised through Bayesian updating using a sampling/importance resampling algorithm.

Album-Level Model Development

In order to model the sales evolution of an individual album, we use a discrete-hazard model. As mentioned earlier, the discrete hazard is appropriate for aggregate data, where the event (here, adoption) occurs at some unobserved time within an interval. In our data, we observe weekly sales, but we do not know the day on which such sales occurred. As an analogy, one can consider modeling integer data with a gamma distribution rather than with a Poisson. Although one might argue that a continuous model may serve as a reasonable approximation to the discrete data, the only way to truly test for the accuracy of a continuous model would be to begin with disaggregate (panel) data and test for the degree of aggregation bias of that model. Given the impossibility of identifying within-period dynamics from grouped duration data (Prentice and Gloecker 1978), we chose to use a discrete-hazard specification.

In the next section, we present a diffusion model that simultaneously estimates a discrete-hazard function and the unknown market potential. We then describe our discrete-hazard function for grouped data, one that is based on the logistic density function to account for duration dependence and response to time-varying covariates.

Diffusion Model with Logistic Hazard

Two aspects of the general diffusion model (Mahajan et al. 1995, Bass 1969) are appealing in the context of prelaunch forecasting: (i) it allows various functional forms of the probability of adoption over time that lead to different diffusion processes, and (ii) it solves a right-censoring problem with unknown market potential. The latter is not trivial because a traditional survival analysis is inappropriate when the size of the risk set, the number of eventual adopters, is unknown.[2] By parameterizing the market potential and relating it to background information, we can simultaneously predict adoption behavior and market potential.

A general family of diffusion models is given as

$$\frac{dN_t}{dt} = \lambda(t) \cdot [m - N_t],$$

where dN_t/dt is the rate of diffusion at time t, N_t is the cumulative number of adopters (buyers) at time t, $\lambda(t)$ is the probability of adoption (purchase) at time t given that adoption has not yet occurred, and m is the number of eventual adopters in a population.

The estimation of album-level parameters is based upon the expected number of incremental adopters for album i in the interval $[t-1, t)$,

$$E(n_{it}) = \lambda(t)(m_i - N_i).$$

We propose Gaussian error variation around the mean, so that

$$n_{it} = \lambda(t)(m_i - N_{it}) + \varepsilon_{it}, \qquad (1)$$

where $\varepsilon_{it} \sim N(0, \sigma_i^2)$. Various functional forms for $\lambda(t)$ lead to models that imply different diffusion processes. Substituting $\lambda(t)$ with the logit hazard in Equation (1) yields a diffusion process defined a discrete logit hazard. The likelihood function for a sequence of observations of adopters of album i at time t $\{n_{it}\}$ can be derived from a sequence of conditional distributions,

$$\begin{aligned} &p(n_{i1}, n_{i2}, \ldots, n_{iT} \mid m_i, \theta_i, \beta_i) \\ &\quad = p\left(n_{iT} \mid n_{i1}, n_{i2}, \ldots, n_{i(T-1)}, m_i, \theta_i, \beta_i\right) \\ &\qquad \times p\left(n_{i(T-1)} \mid n_{i1}, n_{i2}, \ldots, n_{i(T-2)}, m_i, \theta_i, \beta_i\right) \\ &\qquad \vdots \\ &\qquad \times p(n_{i1} \mid m_i, \theta_i, \beta_i) \\ &\quad \propto \frac{1}{\sigma_i^T} \prod_{t=1}^{T} \exp\left[-\frac{1}{2\sigma_i^2}\left(n_{it} - \left[\frac{\exp(Z_i(t))}{1+\exp(Z_i(t))}\right] \times (m_i - N_{it})\right)^2\right], \end{aligned} \qquad (2)$$

where

$$\begin{aligned} Z_i(t) &= X_i\beta + h_i(t), \quad t = 1, \ldots, T_i \\ h_i(t) &= Y\theta_i, \\ Y &= [1, t/10, \ln(t)], \end{aligned}$$

and T_i is the number of observations for album i. The matrix Y contains an intercept as well as linear and log time (t). For the baseline hazard function, we knew in advance that the hazard shape would not require more than U or inverted U shape through raw data plots for most of the albums. Based on empirical comparisons at the album level, we specified the baseline hazard as

$$\lambda_0(t) = \frac{\exp(\theta_0 + \theta_1 t + \theta_2 \ln(t))}{1 + \exp(\theta_0 + \theta_1 t + \theta_2 \ln(t))}. \qquad (3)$$

The exponential term in Equation (3) is the Box-Cox continuous hazard of order one for specifying the dynamics of duration. This functional specification of the baseline hazard achieves a balance of parsimony of the model and the representation of the observed patterns of adoptions. In addition, these time variables are known a priori and do not themselves need to be forecasted.

Logit Specification of Discrete-Hazard Function

Suppose that duration of interest t is in the jth interval so that it satisfies $t_{j-1} \le t \le t_j$. We can define the time-varying index function that captures the effects of covariates and within-interval hazard for duration of interest t as

$$Z_j(t) = X_j\beta + h_j(t),$$

where X_j is a vector of time-varying covariates and $h_j(t)$ is the hazard specification for interval durations.

[2] Contrast our album sales application with a more traditional hazard application, some biomedical experiment conducted on a set of rats. In the latter, the total risk set (the number of rats in the experiment) is known and fixed. In our application, the number of potential buyers of a given album must be estimated.

The logistic hazard model for grouped duration data is then based upon the following survival function. For the jth interval, $t_{j-1} \le t \le t_j$, the logit hazard and survival functions are given as

$$\lambda_t(t, X, \beta) = \frac{\exp(Z_j(t))}{1+\exp(Z_j(t))}$$

$$S_j(t, X, \beta) = \frac{1}{1+\exp(Z_j(t))} \times \prod_{k=1}^{j-1} \{1+\exp(Z_k(t_k))\}^{-1} \quad (4)$$

and the underlying probability density for durations by definition becomes

$$f(z) = \frac{\exp(Z_j(t))}{1+\exp(Z_j(t))} \prod_{k=1}^{j-1} \{1+\exp(Z_k(t_k))\}^{-1}.$$

Therefore, given data on grouped durations, the probability model for the observed data is completely specified through either the parameterization of the survival function or of the probability density. In fact, a discrete analog of the proportional hazards model (Sueyoshi 1995, Cox and Oakes 1984), a popular model for survival analysis, is simply equivalent to estimating a pooled logit with period-specific constant terms to allow for duration effects (Sueyoshi 1995). This specification of the hazard process provides a practical framework for application of the traditional binary logit specification of duration data. The framework provides an explicit linkage between the binary specification and the underlying hazard, allowing one to easily assess the implications and assumptions of a given specification. If the h functions are assumed to be stationary and constant within equal length intervals, the pooled logit specification with a single constant term can be applied (Sueyoshi 1995, Kiefer 1988). However, if the within-interval durations are assumed to be different, the estimation of the hazard model is equivalent to estimating a pooled logit specification with a period-specific constant term. Since sales for recorded music are reported in weekly aggregates, and given the ease of estimating the logit hazard and given that it serves as a discrete analog to the proportional hazards model (the primary hazard model used in marketing), we specify our discrete hazard to be a logit.

Second Level of Model Hierarchy

We believe that underlying characteristics of the albums may affect the adoption parameters θ and β and market size m. Defining $\phi_i = [\theta_i, \log(m_i)]$ for the ith album, we propose

$$\phi_i = W_i\delta + u_i, \quad (5)$$

where $u_i \sim N(0, V)$, W_i is a set of c album characteristics of dimension $1 \times c$, θ_i is of dimension $1 \times l$, and so ϕ_i is of dimension $1 \times r$, where $r = l+1$. Stacking Equation (5) across the albums leads to

$$\phi = W\delta + u_i,$$

where W is a matrix of dimension $n \times c$, ϕ is $n \times r$, and δ is $c \times r$, and the variance of u is block diagonal with V as the blocks.

We jointly estimate the parameters of both levels of the hierarchy. A naïve strategy of estimating each album hazard model separately and then using the estimated coefficients in a second-stage model does not properly account for estimation error. If there is substantial estimation error in the album-level response coefficients, this type of two-step approach may provide misleading views about the importance of the explanatory variables. That is, the R-squared of the second-stage regression can be low due simply to estimation error rather than that the true coefficients are unrelated to these variables. In addition, a joint estimation method will usually provide gains in statistical efficiency.

To complete our specification of the model, we assume prior distributions for the model hyperparameters. For δ, we assume $\text{vec}(\delta) \sim N(\omega, \Psi)$. We set $\omega = 0$ and $\Psi = I \times 200$, where I is the identity matrix. The largest variance of the draws from the posterior of the elements of δ is 0.07, indicating that our specification of Ψ is truly diffuse. We adopt the standard inverse gamma prior on σ,

$$p(\sigma_i \mid \nu, \tau) \propto \frac{1}{\sigma_i^{\nu+1}} \exp\left(-\frac{\tau}{2\sigma_i^2}\right),$$

where we set $\nu = 2$ and $\tau = 0.7$. The expected value of this prior is close to maximum likelihood estimates of the error variances of the stores, and the long tail of the distribution allows for large values of σ. We investigated alternative specifications of this prior, finding our posterior estimates to be robust relative to variations in the hyperparameters of this prior.

As for a prior for V, we use the hyperprior developed by Barnard et al. (2000) to allow for differential shrinkage across the various elements of δ. This prior decomposes V into a vector of standard deviations S and a correlation matrix R, $V = diag(S)Rdiag(S)$, where $diag(S)$ is a diagonal matrix. In this prior, the diagonal elements of S, s_j, are assumed to be independently distributed inverse gamma, or $s_j \sim \mathrm{IG}(\xi_j, \chi_j)$ for $j = 1, \ldots, r$. For all r elements of δ, we set $\xi_j = 1$ and $\chi_j = 1$. Following Barnard et al. (2000), we allow R to be uniformly distributed on the space of positive definite matrices.

Model Estimation

We estimated model parameters using a single-component Metropolis-Hastings chain (Gilks et al. 1996, Metropolis et al. 1953). We divide the parameter space into the components $(\phi, \delta, \sigma, S, R)$, sampling each component conditional on the remaining set. In our application, we are able to directly sample from the conditional posterior densities of δ and of σ, since our priors on these elements are conjugate. The forms of these conditional posterior densities are well known and can be found in Rossi et al. (1996). For S and for R, we follow the algorithm used by Barnard et al. (2000), again sampling from the conditional posterior density.

The proposal densities for the album-specific parameters m and θ are more complicated, due not only to nonconjugacy but also to strong correlation between m and the first element of θ (the hazard intercept). We sample from the joint space (m, θ) using a Metropolis-Hastings sampler. We use a multivariate t-distribution as the proposal density because its fatter tails help avoid the occurrence of extremely large weights in Monte Carlo summation. To facilitate movement throughout the parameter space, we centered the t-distribution at the previous draw and used a covariance matrix set equal to the current draw of V. The degrees of freedom can be used as a tuning constant; we adjusted the degrees of freedom to achieve between 40–60% acceptance rates of proposed moves in the chain (Draper 2001).

Prelaunch Forecasting

The second stage of the hierarchy provides a link between adoption patterns observed in the past and the characteristics of albums and artists. Using the distribution of the parameter estimates for the exogenous variables, the next step is to estimate the predictive density of sales and of the new album's hazard parameters. Let W_j be a vector of the new album j's characteristics, then the initial distribution of parameter estimates for ϕ_j can be obtained using Equation (5) and the draws of δ. The distribution of ϕ_j can then be used in Equation (2) to obtain the predictive density of sales.

The quality of prelaunch forecasts depends heavily on the information content of the background characteristics. If there exists a strong relationship between album-specific hazard parameters and the background characteristics in a calibration sample, we may expect the same pattern to persist in the prediction sample. If the initial predictive performance is poor, improvement in the accuracy of the forecasts depends on the efficiency of the updating scheme, as the actual sales data become available.

Bayesian Updating via Sampling/Importance Resampling (SIR)

As new data become available, forecasts for a new album can be revised at each phase of its market introduction. Although one strategy would be to refit the model each time new data become available, the same results can be more quickly obtained by reweighting existing results. Let Ω be the vector of model parameters of the new album. The joint posterior distribution of Ω can be obtained by updating the prior distribution of Ω, which is the predictive density of model parameters. For instance, consider an update after observing a single week of sales. With one week of data, the joint posterior distribution of parameters is given by

$$\pi_j\left(\Omega^{(1)}\right) = \frac{L_j\left(n_1 \mid \Omega^{(0)}\right) \cdot \pi_j\left(\Omega^{(0)}\right)}{\int L_j\left(n_1 \mid \Omega^{(0)}\right) \cdot \pi_j\left(\Omega^{(0)}\right) d\Omega}, \tag{6}$$

where $\pi_j(\Omega^{(1)})$ is the distribution of model parameters of album j after observing one week of sales, and $L_j(n_1 \mid \Omega^{(0)})$ is the likelihood of the first week of sales.

The multiple integral in Equation (6) is difficult to evaluate because is does not have a closed-form solution. This integration problem can be solved by using the sampling importance resampling (SIR) procedure (Efron 1982, Smith and Gelfand 1992). With one week of data, the SIR procedure operates as follows. First, draw M sets of parameters from $\pi_j(\Omega^{(0)})$. The next step is to reweight the M draws, where the weight for the ith draw is given as

$$w_i = \frac{L_{j,i}}{\sum_{k=1}^{M} L_k},$$

where $L_{j,i}$ is the likelihood of the ith draw for album j (Equation (2)). The reweighted draws can be used to calculate moments or other functions of parameters of the distribution $\pi_j(\Omega^{(1)})$. This procedure can be generalized to condition on more than one week of observed sales.

Empirical Illustration of Proposed Model

Description of the Data

The data analyzed here consist of weekly sales of the albums in the *Billboard* Top 200 album chart from January 1994 to December 1995, initially processed by SoundScan Inc. and supplied by National Record Mart. The data are currently collected from 14,000 retail outlets, including 40 different chains, 11 mass merchandisers, and over 600 independent retail locations. The average weekly transactions amount to 9–10 million albums, and the data are sent via modem from point-of-sale registers in the stores. The data are published in *Billboard* magazine, *Rolling Stone*, *Entertainment Weekly*, *Music Week*, and *The Wall Street Journal*. *Billboard*'s best-selling Top 200-album chart list records by title, name of artist, name of producer, and record label number. Also indicated are the number of weeks a record has been on the chart, its standing in each of the two previous weeks, the names of its writer(s), publisher, and the availability of the videos. Songs registering the greatest airplay and sales gains for the past week are indicated too. The albums included in the chart cover a wide variety of music and are classified as pop, rock, country, R&B, rap, hard rock, jazz, and movie soundtracks.

Data Preprocessing

There are more than 1,200 albums that appeared at least once in the chart. We reduced the set by removing left-censored and right-censored albums. We also deleted Christmas albums and movie soundtracks due to their atypical sales patterns. Both categories combined account for only 8% of the total sales volume of recorded music. Finally, we deleted albums with 10 or fewer data points (number of weeks in the chart) to provide maximal information content in the set of albums for the computation time. After applying these criteria, we ended up with 295 albums, from which we randomly chose 50 for a holdout sample, leaving the remaining 245 as the calibration set.

Seasonally adjusted data are used for the estimation. Seasonal adjustment uses the weekly index based on the total sales for 200 albums in each week; our adjusted data are proportional to the actual data divided by the weekly index. Sales of recorded music are highly seasonal. For example, during the Christmas week there are more than 25 million transactions, which account for 30–35% of annual gross revenues (Fink 1996). Though a significant portion of the increased sales during that period comes from those albums not included in the sample (Christmas, soundtrack, etc.), the analogous index calculated with the 245 albums in the calibration set shows a similar tendency. Sales volume of the albums in the calibration set ranges from 300,000 to 3,500,000 units. Table 1 provides the general information on the 245 albums used for the calibration set.

Album Characteristics

For the set of exogenous variables, we used album/artist characteristics. These variables were (i) music category of an album, (ii) gender of the artist, (iii) total number of albums released by an artist, (iv) number of gold albums, (v) number of platinum albums, (vi) AMG ratings for the album, and (vii) promotional effort. First, music category and gender are the basic classification variables expected to influence both adoption patterns and market potential. Second, the track record of an artist, such as the number of gold or platinum albums, could have some impact on initial demand and market potential. Third, album review is based on the ratings by the All-Music Guide

Table 1 Description of Albums of the Calibration Sample

	Frequency of artist type		
Music type	Male	Female	Group
Country	26	13	9
POP	18	16	13
RAP	18	4	17
Alternative	4	6	21
Rock	12	3	15
R&B	10	7	12
Hard rock	6	0	15

Number of weeks in the chart (T)	Frequency
$11 \leq T \leq 12$	24
$13 \leq T < 24$	99
$24 \leq T < 36$	51
$36 \leq T < 48$	36
$48 \leq T$	35

Descriptive statistics for continuous variables

	Mean	St. dev.	Min	Max
AMG	3.55	0.93	1.0	5.0
Promotion	7.07	4.46	0.53	27.87

(AMG). AMG's music ratings and review are done by a coordinated system of freelance writers and music experts. For gold plus albums, the average AMG rating was 4.25, compared to 3.60 for all other albums. Finally, airplay for the first six weeks is used as a proxy for promotional effort of the record company. Record companies consider radio airplay to be the most direct way of exposing a record to the buying public (Blake 1992, Fink 1996). The main tools are special promotional copies of the record, called "promo" records, which are placed in the hands of broadcasters and programming consultants. Record promoters of a major record company must persuade broadcasters to schedule their company's newest releases for playing on the air. Since many stations broadcast a weekly "playlist" of only 30 or fewer current hits, and because there are hundreds of new releases each month, airplay serves as a measure of how much promotional effort was made.

We used stepwise regression to aid selection from the set of exogenous variables. We first estimated market potential m and hazard parameters θ in a hierarchical model with no album characteristics as covariates ($W_j = 1$). Using the posterior means of $\log(m)$ and the hazard parameters as dependent variables, we used stepwise regressions to select candidate variables to include in the matrix of covariates, W.

Based on the results of the regression, we restricted some of the elements of δ to equal zero. For the purpose of estimation of the elements of δ, we calculated the distribution of the remaining elements of δ conditional on the subset being equal to zero, using this conditional distribution to obtain draws of the remaining elements of δ. For a list of the elements of δ that were not set to zero, please refer to Table 3 (which is discussed in detail below).

Results of Model Estimation

Although the main thrust of the paper is prediction of sales in a holdout sample of albums, there are interesting aspects of the fitted parameters that explain performance of different kinds of albums. In this section, we discuss the fit of the model to the calibration sample and interpret the coefficients estimates.

Because our proposed hierarchical Bayes hazard model provides album-specific parameters, and due to the flexibility of the functional form of the hazard, the model fits are quite close. Figure 1 shows the sales and fitted sales of 12 albums. Over the whole set of 245 albums, the mean absolute percentage error (MAPE) was 18%. Because we have a long time series for each album, album-specific parameters are primarily influenced by the sales data of each album, and consequently, album characteristics, which affect the mixing distribution across all albums, do not greatly increase model fit. The MAPE of our model without album characteristics was also 18%. We must note, however, that the purpose of these album characteristics is not to explain past sales, but to aid prelaunch sales forecasts for new albums, where a timeseries is not yet available.

Distribution of Market Potential Estimates

For each album, we plot the posterior mean in the first panel of Figure 2. The distribution of posterior means is highly skewed. While the market potential (unit sales) of some albums is extremely high, the majority (84%) of the market potentials are less than 1,000,000 ($\log(\text{market}) < 4.6$, since market potential is

Figure 1 Model Fits for 12 Example Albums

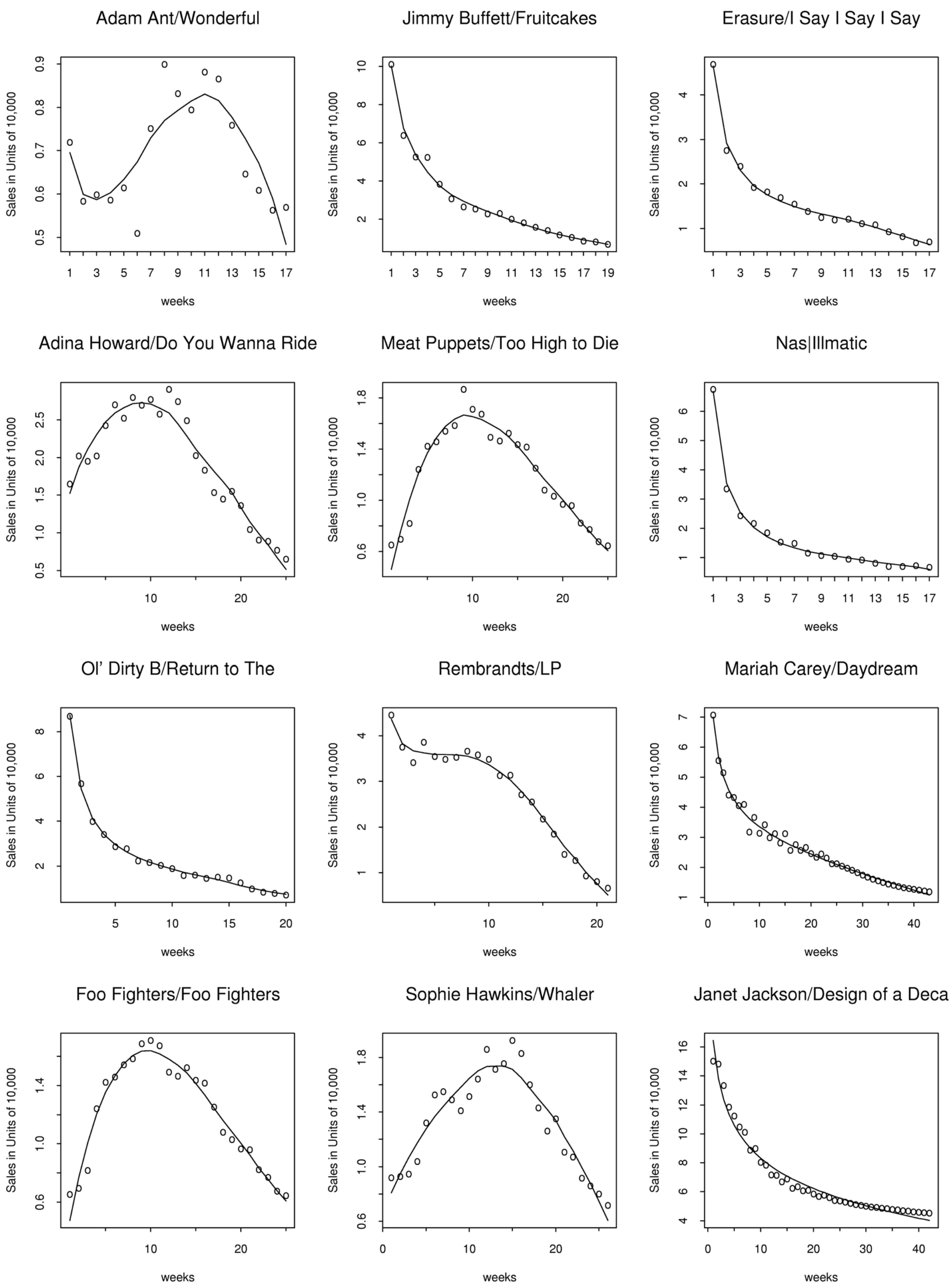
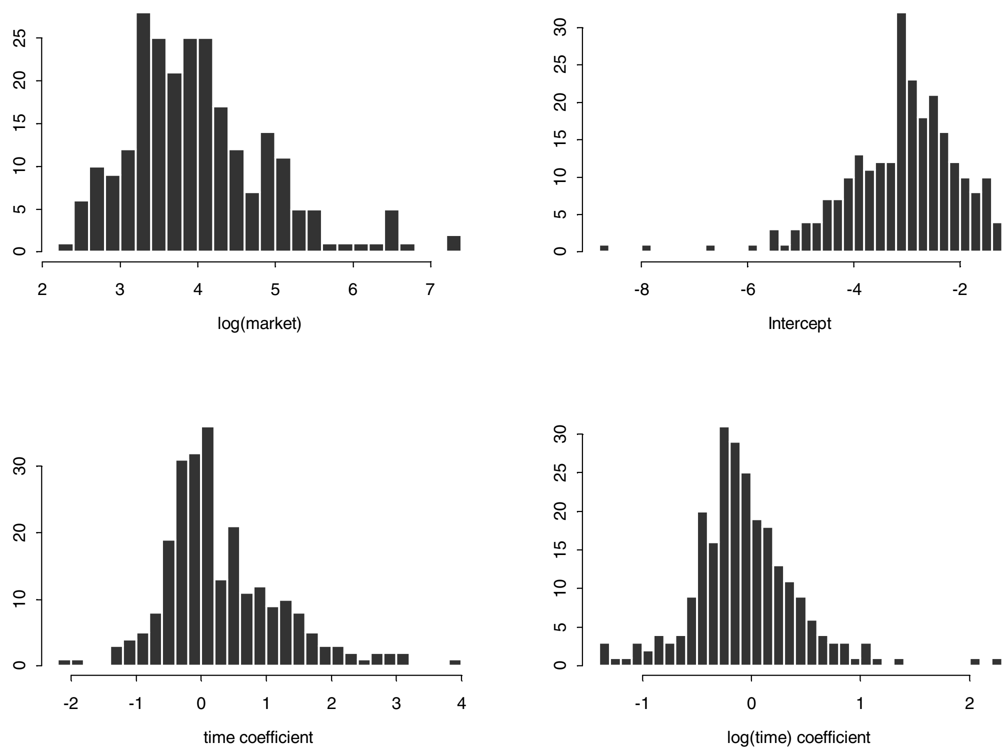


Figure 2 **Histogram of Posterior Means of Model Parameters (245 Albums)**



measured in units of 10,000). High correlation of the parameter space inflates the variance of the marginal distributions of parameters. Even so, the variance of the distributions of the market potentials is, from a managerial perspective, reasonably low. For instance, the 12 panels of Figure 3 show the marginal distributions of the market potential for the same albums that were pictured in Figure 1. These distributions reveal that the uncertainty on market potential is within a managerially reasonable range. For instance, with probability 0.95, the market potential of "Wonderful" is between 124,291 and 128,979 albums sold. The high degree of precision for the market potential for such albums is understandable in light of the sales diffusion pattern of these albums, shown in Figure 1; we are using the entire product life cycle to estimate market potential. Since we have the entire life cycle for most of our albums, our posterior estimates of market potential exhibit low variance.

Also note that the posterior distributions of some albums are asymmetric. The market potential for Janet Jackson's hit (the last panel of Figure 3) might be much higher than the mode, while it is less likely to be far lower than the mode. These asymmetries can allow managers to recognize which albums might end up with exceedingly high or low sales.

Distribution of the Baseline Hazard Parameters (θ)
The elements of the θ vector determine the shape of the sales curve. Although the posterior means for the linear and log-time coefficients are relatively homogeneous, the distribution of posterior means exhibits thick tails relative to Gaussian tails (Panels 3 and 4 of Figure 2), indicating that some albums are quite different from the rest. Summary statistics of posterior means of θ across albums are given in Table 2. Note in particular that even though the marginal distributions of log and linear time are centered near zero, the majority of the elements of θ are significant (see the last line of Table 2).

As one would expect, the linear and log-time parameters are negatively correlated. Although this correlation increases the variance on the marginal

Figure 3 Marginal Distributions of Market Potential Parameters for 12 Example Albums

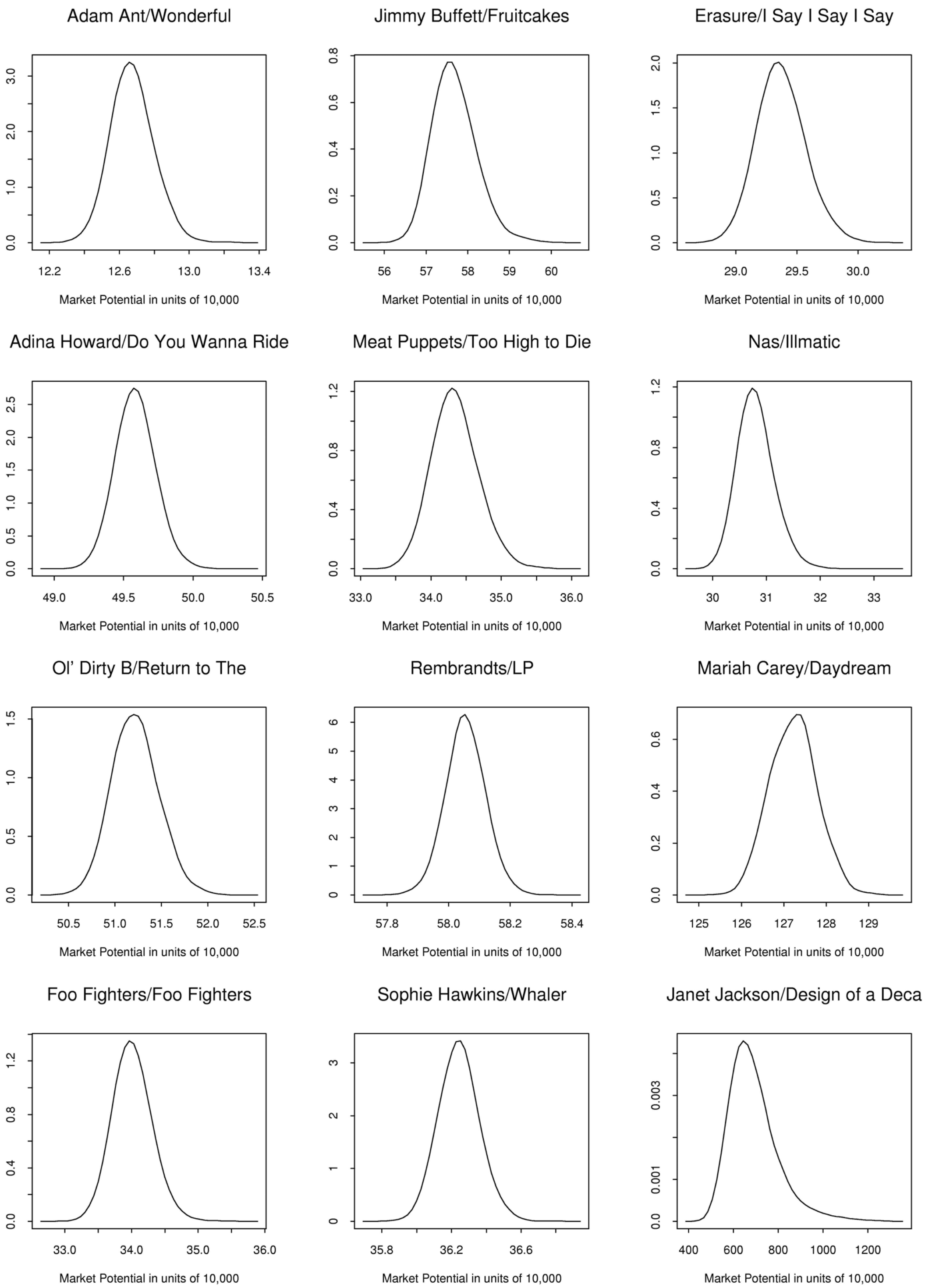


Table 2 Summary Statistics of Posterior Means of θ

	Intercept	Time	log(time)	log(market)
Mean	−3.1141	0.3094	−0.0661	4.0009
Std. dev.	1.0790	0.8998	0.4906	0.9100
Median	−3.0172	0.0910	−0.1154	3.8565
Number of significant* parameters	245/245	157/245	194/245	245/245

Note. *significant indicates that the central 95% of the mass of the posterior distribution does not contain zero.

Table 3 Estimates of δ

Explanatory	θ and log(m)	Posterior mean	Posterior std. dev.	$P(\delta > 0)$
Intercept	Intercept	−2.0128	0.2709	<0.0001
Intercept	time	0.5027	0.0958	>0.9999
Intercept	log(time)	−0.2172	0.1015	0.0160
Intercept	log(m)	2.7826	0.2372	>0.9999
# of Albums	Intercept	0.0310	0.0095	>0.9999
# of Gold	log(m)	−0.0828	0.0314	0.0040
# of Platinum	time	−0.0430	0.0172	0.0078
# of Platinum	log(m)	0.0768	0.0228	0.9998
AMG rating	Intercept	−0.3067	0.0689	<0.0001
AMG rating	log(time)	0.0510	0.0264	0.9727
AMG rating	log(m)	0.2175	0.0575	>0.9999
Promotion	time	−0.0210	0.0100	0.0195
Promotion	log(m)	0.0502	0.0119	>0.9999
Country	Intercept	−0.4045	0.1124	<0.0001
Rock	time	0.2470	0.1094	0.9890
R&B	Intercept	−0.6176	0.1965	0.0010
R&B	log(time)	0.1774	0.0756	0.9908
R&B	log(m)	0.3013	0.1660	0.9640
Male	log(time)	−0.1330	0.0353	0.0002

densities of each of the time parameters, such correlation does not decrease the predictive ability of the model, just as multicollinearity does not decrease the prediction capabilities of linear regression models.

Distribution of the Parameters for Album Characteristics (δ)

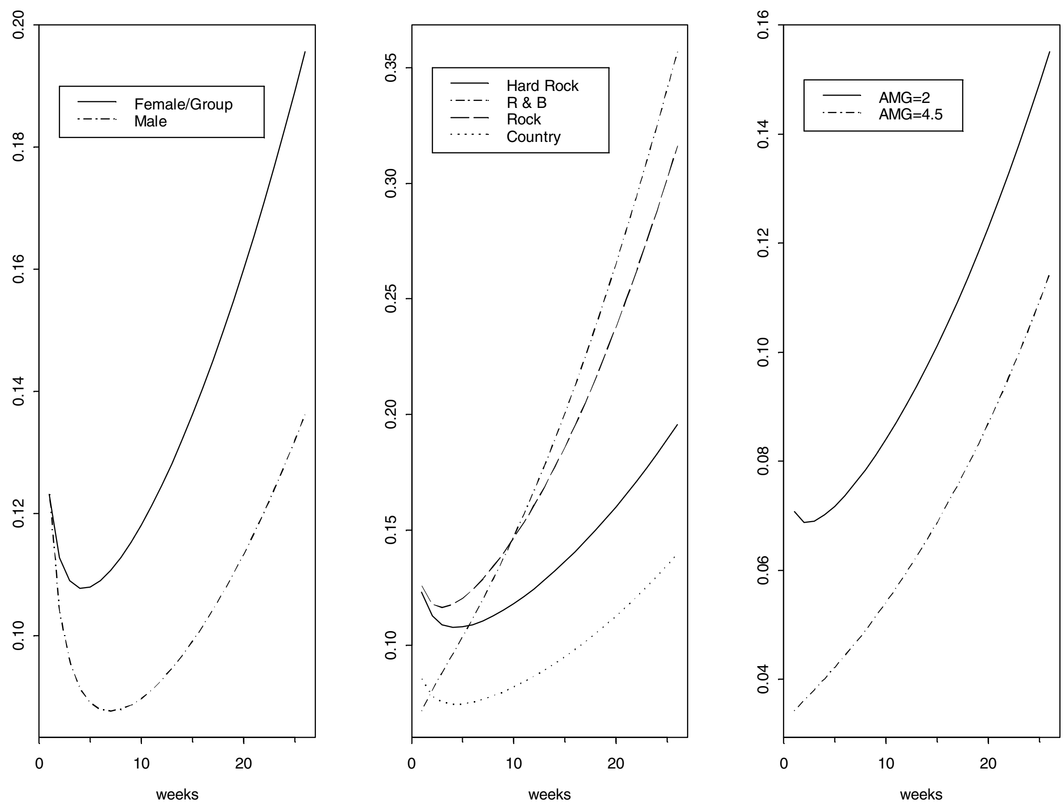
The results regarding δ indicate that variation in the hazard function parameters and in market potential can be attributed in part to variation in the explanatory set of variables. Thus, the explanatory information set can be used to provide an informative prior distribution over θ of a new album before the launch of that album. In addition, δ serves to explain differences between the albums, both with respect to market potential and with respect to the adoption patterns (i.e., hazard function) across albums.

Table 3 gives the posterior means of the elements of δ along with the mass of the positive region of the marginal distribution of each element. The interpretation of this final column of Table 3 is straightforward, for it gives the probability that the parameter in question is greater than zero. If this probability is very high or very low, then the probability that this parameter is close to zero is small. In classical terms, a high or low probability indicates that the parameter is significantly different from zero. All of these parameters are significant, as we conducted variable selection (described above) with the goal of producing a model with excellent forecasting ability.

These results can be interpreted in a manner similar to a multivariate regression. In other words, each of the hazard parameters (and market potential) is a linear function of the explanatory factors, including an intercept term, added to an error term that may be correlated with the error term of other parameter estimates. For instance, the results show that, all else equal, the log of market potential is higher on average by 0.3013 for Rhythm and Blues (R&B) relative to the other music types, indicating that sales for this music type are 35% higher relative to the other music types ($\exp(0.3013) - 1$). Similarly, market potential is expected to be higher by 24% for each additional AMG rating point and by 5% for each unit increase in promotion of the album. Finally, market potential is affected by the number of gold and platinum albums held by the artist. Although the coefficient for the number of gold albums is negative, this factor is highly correlated with the number of platinum albums ($r = 0.65$), and the interpretation of these related coefficients, as usual, assumes the other factor is held constant. Holding platinum albums constant, artists with a greater number of gold albums had lower total sales than artists with fewer gold albums. This result may have occurred since a larger number of gold albums relative to platinum albums would indicate inferior sales potential.

The remaining results for δ pertain to the parameters of the hazard function. The results indicate that male soloists have a significantly different hazard

Figure 4 Hazard Function Comparisons by Album Type



function from other artists, since the nonlinear time parameter is significantly different from zero. Panel 1 of Figure 4 illustrates how the different hazards affect the shape of the sales curve of male soloists relative to other artists. The hazard for males decreases more rapidly than does that for females and groups, meaning that albums for males tend to sell, percentagewise, earlier in the life cycle of the album.

The different genres of music have significantly different hazards as well, as shown in Panel 2 of Figure 4. R&B has a rapidly increasing hazard relative to other music styles; the culture of listeners of different music types may account for the variations in adoption patterns. Differences in AMG ratings also impact hazard shapes, as shown in Panel 3 of Figure 4, where higher-rated albums have hazards that are almost linearly increasing, not showing the initial decrease in hazard of the (relatively) lower-rated albums.

Finally, note that although all of the hazards shown in Figure 4 have fairly similar shapes, we have only shown the effect of individual album characteristics on the hazards. These album characteristics combine together to yield a variety of hazards for our dataset: constant, decreasing, decreasing/increasing, and increasing.

It is also interesting in some cases to note which δ parameters were not significant. In particular, we expected that music genre dummy variables would affect market potential, since the sizes of audiences differ. However, after accounting for AMG ratings and for album promotion, music genre does not significantly impact market potential.[3] In retrospect, since AMG ratings and promotion spending are based on past album success (sales) of artists, such measures are functions of audience sizes (market potential) and leave little variance to be explained by remaining album characteristics such as music genre.

[3] In models with music genre dummies but no promotion or AMG characteristics, the genre dummies significantly impact market potential.

Table 4 Comparison of Fit (MAPE) with Alternative Hazard Formulations

	Estimation			Forecasting		
	Generalized Gamma model	Generalized Bass model	Proposed model	Generalized Gamma model	Generalized Bass model	Proposed model
MAPE						
Average	0.267	0.196	0.178	0.896	0.799	0.700
Median	0.183	0.151	0.097	0.741	0.784	0.522
Country	0.301	0.235	0.188	0.841	0.783	0.712
POP	0.225	0.183	0.132	1.169	0.715	0.586
RAP	0.315	0.190	0.253	0.673	0.714	0.430
Alternative	0.224	0.186	0.128	0.742	0.770	0.466
ROCK	0.160	0.175	0.086	0.893	0.749	0.410
R&B	0.259	0.203	0.161	0.752	0.807	0.276
Hard rock	0.406	0.199	0.371	1.200	1.054	0.608

In order to assess the value of the flexible hazard in our model, we apply different types of diffusion models to each of the albums in our calibration sample and compare the estimation fit. Table 4 shows the fit comparison with (i) a generalized Bass model (Bass et al. 1994) with airplay as time-varying covariate, (ii) a generalized gamma hazard model, which nests exponential, gamma, log-normal, and Weibull distribution as submodels, and (iii) our proposed logistic hazard formulation (in Equation (4)). This comparison shows that the proposed formulation of the hazard outperforms the others, on an album-by-album basis, both in terms of the average and median album-level mean absolute percentage error. Also, it fits better for 194 out of 245 albums compared to the generalized Bass model and 234/245 to a gamma hazard. Therefore, we conclude that the proposed model shows the most appropriate specification for the hazard function, one that is flexible enough to accommodate the diverse patterns of adoption across all albums. In the next section, we discuss the forecasting results based on our hierarchical Bayes random-coefficient model.

Prelaunch and Updated Forecasts

To test our modeling framework in prelaunch forecasts, we tried to emulate the manager's situation as closely as possible. For example, since weekly airplay is not known beforehand and is not directly under control of the manager, we do not consider weekly airplay in our predictive tests. Table 4 shows the fit comparison of out-of-sample predictions with other models using the same background characteristics. Our proposed model clearly outperforms them, both in terms of overall fit and category-specific predictions. In Figure 5 we show prelaunch forecasts over 24 weeks for albums in the holdout sample[4]. The plotted points show the actual sales, the solid line represents the prelaunch forecast, and the Generalized Bass model forecasts are shown in an alternating dash-and-dot pattern. As would be expected, there are cases in which our model over- or underpredicts market potential (Reba McEntire's "Starting Over"), while accurately predicting the shape of the sales curve. In other cases, the sales curve itself is difficult to predict, while our model provides a relatively accurate market potential parameter (such as in Tim McGraw's "All I Want"). Even so, across a diverse set of albums our model provides relatively accurate predictions of the sales patterns as well as the total sales, more accurate than those of the Generalized Bass model.

Also in Figure 5, we use a dashed line to show forecasts that use the first 12 weeks to project the remaining weeks. In cases where the actual sales follows a relatively smooth function, such as Pink Floyd's album "Pulse," the updated forecasts are extremely accurate. When sales are noisier, however, such as for Collin Raye's "Extremes," the updated forecast is less accurate, possibly less accurate than even

[4] For some albums there are fewer than 24 weeks plotted, since sales died down in fewer than 24 weeks.

Figure 5 Forecasts

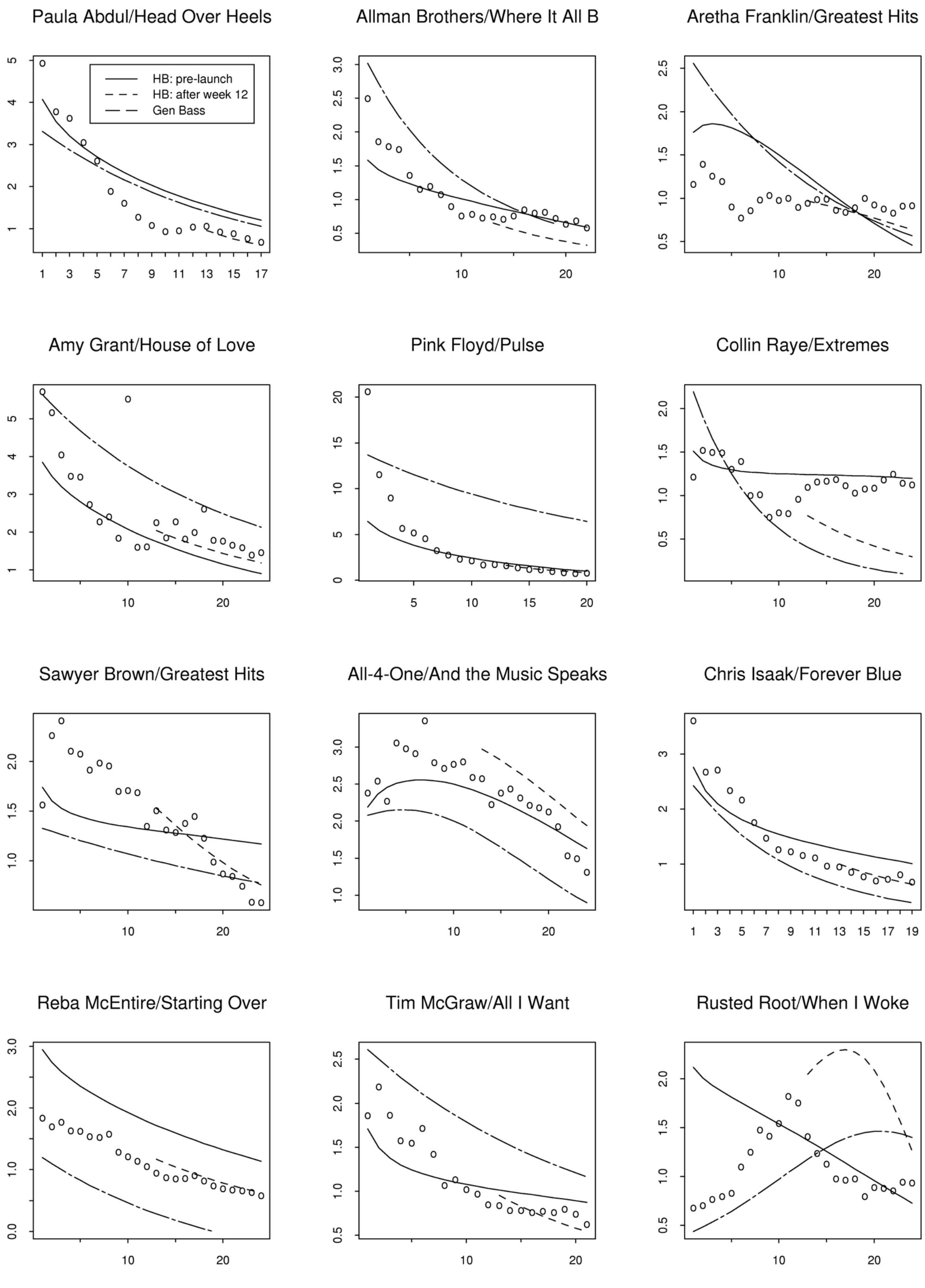


Table 5 MAPE of 12-Week-Ahead Forecasts

Forecast time	MAPE (%)
Prelaunch without album characteristics	69
Prelaunch	52
After 1 week of observed sales data	29
After 2 weeks	31
After 3 weeks	30
After 4 weeks	26
After 6 weeks	28
After 8 weeks	34

Note. The MAPEs reported here are median MAPEs across albums.

the prelaunch forecast. However, in most cases the updated forecasts are very close to the actual weekly sales. As for statistics on the forecast accuracy over all 50 albums in the holdout sample, we provide MAPE calculations in Table 5. We calculate the MAPE for each album over periods of 12 weeks using the posterior means of the parameters; we report the median MAPE over the set of 50 albums. For the forecasts using the first week of data, for instance, we forecast Weeks 2 through 13 for each album, calculate the MAPE over this period for each album, and report the median of these MAPEs in Table 5. For record companies, the prelaunch forecasts are arguably the most important forecasts of those in the table. Our results show that the set of album characteristics provides a reliable basis for an informative prior for the parameters of the holdout sample, for the set of covariates improved the prelaunch forecast MAPE from 69% down to 52%.

The remaining MAPE statistics in Table 5 use initial sales observations to update the informative prior. The first week of sales greatly enhances the accuracy of the forecasts, improving the MAPE from 52% to 29%. After the first week, though, remaining weeks provide little incremental forecasting improvement, for the MAPE results after observing additional sales do not continue to decrease far below 29%, but hover in the low 30% and upper 20% range. Considering that the value of forecasts diminishes as weeks go by, we find the improvement of fit after observing one week to be dramatic. So, the covariates provide significant information for accurate prelaunch forecasts, and the first observation of sales, when it becomes available, can reliably be used to update the initial forecasts.

Conclusion

Most companies recognize that the development of new products is accompanied by high costs and risks. One way of controlling these risks is the use of sound explicit models for planning and forecasting new product sales. Because several hundred new singles and albums are released each month, it is rather difficult and cumbersome to produce forecasts for each album in a reliable and consistent manner. It is a common problem in other product categories as well (e.g., motion pictures, books, CD-ROMS, pharmaceutical drugs), in which new products are frequently introduced (Sawney and Eliashberg 1996, Jones and Ritz 1991). The purpose of our study is to reduce the uncertainty by leveraging past sales information for previous albums via a hierarchical model to provide relatively accurate weekly sales predictions as well as estimates of total market potential for a new set of albums. Because most of the quantitative forecasting methods would rely on existing data for a specific album to establish model parameters for that album, they may not be well suited for predicting sales of a new product before its introduction. In this paper, we use the experience obtained from prerecorded (previous) albums to present a forecasting method based upon a hierarchical Bayesian model of the logistic diffusion model.

In sum, our model can be used to identify generalized sales patterns for the purpose: (1) of forecasting sales of a new album before its introduction and (2) to improve these forecasts as new information (album-specific sales) becomes available. In the context of new product forecasting before introduction, the album-specific information we have is limited to the relevant background characteristics of a new album. Knowing only the general attributes of a new album, the Bayesian approach proposed here takes into account the informed prior on the dynamics of duration, the effects of marketing variables, and the unknown market potential, where the prior is informed via a past set of albums.

The accuracy of forecasts does suffer from limitations that are common to models in general. Though the overall performance of the proposed model depends on the strength of the relationship between album/artist characteristics and the album-specific market potential and hazard parameters, there are albums which do not follow the typical patterns that establish the empirical generalization. For instance, the artist with an impressive track record such as multiple gold/platinum albums will be off to a good start due to loyal fans. Though logical and mostly true, such a pattern is not without limitation in explaining the emergence of a blockbuster debut of a one-hit wonder. Similarly, a consecutive, sizable two-month increase (Rusted Root's "When I Woke") or decrease (Collin Raye's "Extremes") in sales dictated more prolonged patterns in our model, patterns which turned out to be inaccurate. Predicting a sudden burst of consumer interest or any other deviations from the norm is not the strength of any model, for the cost of the consistency of a model is the rigidity of the (partial) information set used to calibrate the model (Blattberg and Hoch 1990). What our model offers is a sound decision-theoretic framework, a framework that updates to incorporate new information, accounts for a wide range of sales patterns, and produces accurate forecasts of album sales far in advance of launch.

References

Allison, P. D. 1995. *Survival Analysis Using the SAS System: A Practical Guide*, SAS Institute Inc, Cary, NC.

Bake, A. 1992. *The Music Business*. B. T. Batsford Ltd., London, U.K.

Barnard, J., R. McCulloch, X. Meng. 2000. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statist. Sinica* **10**(4) 1281–1311.

Barwise, P. 1995. Good empirical generalisations. *Marketing Sci.* **14**(2) G29–G35.

Bass, F. 1969. A new product growth model for consumer durables. *Management Sci.* **15**(1) 215–227.

——. 1995. Empirical generalizations and *Marketing Science*: A personal view. *Marketing Sci.* **14**(2) G6–G19.

——, T. V. Krishnan, D. C. Jain. 1994. Why the Bass model fits without decision variables. *Marketing Sci.* **13**(2) 203–223.

Bayus, B. L. 1993. High-definition television: Assessing demand forecasts for a next generation consumer durable. *Management Sci.* **39**(11) 1319–1333.

Blattberg, R. C., S. J. Hoch. 1990. Database models and managerial intuition: 50% model + 50% manager. *Management Sci.* **36**(8) 887–899.

Cox, D. R., D. Oakes. 1984. *Analysis of Survival Data.* Chapman & Hall, London, U.K.

Draper, D. 2001. Bayesian hiearchical modeling, 2nd draft. Working paper, www.bath.ac.uk/~masdd/, 72.

Efron, B. 1982. *The Jackknife, the Bootstrap, and the other Resampling Plans.* National Science Foundation-Conference Board of the Mathematical Sciences Monograph 38, SIAM, Philadelphia, PA.

Erlewine, M. 1994. *All Music Guide: The Expert's Guide to the Best Releases from Thousands of Artists in All Types of Music*. Miller Freeman Books, San Francisco, CA.

Farley, J. U., D. R. Lehmann. 1986. *Meta-Analysis in Marketing: Generalizing from Response Models.* Lexington Books, Lexington, MA.

——, ——, A. Sawyer. 1995. Empirical marketing generalizations using meta-analysis. *Marketing Sci.* **14**(2) G36–G46.

Fink, M. 1996. *Inside the Music Industry: Creativity, Process, and Business*. Schirmer Books, New York.

Flinn, C., J. Heckman. 1982. Models for the analysis of labor force dynamics. *Adv. Econometrics* **1** 35–95.

Gilks, W. R., S. Richardson, D. J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, New York, 10–12.

Jones, M. J., C. J. Ritz. 1991. Incorporating distribution into new product diffusion models. *Internat. J. Res. Marketing* **8** 91–112.

Judge, G. G., W. E. Griffiths, R. C. Hill, H. Luktepohl, T. Lee. 1988. *The Theory and Practice of Econometrics.* John Wiley and Sons, New York.

Kalbfleisch, J. D., R. L. Prentice. 1980. *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Kiefer, N. M. 1988. Economic duration data and hazard functions. *J. Econometric Lit.* **26** 649–679.

Lenk, P. J., A. G. Rao. 1990. New models from old: Forecasting product adoption by hierarchical Bayes procedures. *Marketing Sci.* **9**(4) 42–53.

Mahajan, V., Y. Wind. 1986. *Innovation Diffusion Models of New Product Acceptance.* Ballinger Publishing Company, Cambridge, MA.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller. 1953. Equations of state calculations by fast computating machines. *J. Chem. Phys.* **21** 1087–1092.

Montgomery, A. 1997. Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Sci.* **16**(4) 315–337.

Neelamegham, R., P. Chintagunta. 1999. A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Sci.* **18**(2) 115–136.

Prentice, R., J. Gloecker. 1978. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* **34** 57–67.

Putsis, Jr., W. P., V. Srinivasan. 2000. Estimation techniques for macro diffusion models. V. Mahajan, E. Muller, Y. Wind, eds, *New Product Diffusion Models.* Kluwer Academic Publishers, Boston, MA.

Rossi, P. E., R. E. McCulloch, G. M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Sci.* **15**(4) 321–340.

Sawhney, M. S., J. Eliashberg. 1996. A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Sci.* **15**(1) 113–131.

Smith, A. F. M., A. E. Gefland. 1992. Bayesian statistics without tears: A sampling-resampling perspective. *Amer. Statist.* **46** 84–88.

Sueyoshi, G. T. 1995. A class of binary response models for grouped duration data. *J. Appl. Econometrics* **10**(3) 411–431.

Sultan F., J. U. Farley, D. R. Lehmann. 1990. A meta-analysis of applications of diffusion models. *J. Marketing Res.* **32**(February) 70–77.

Talukdar, D., K. Sudhir, A. Ainslie. 2002. Investigating new product diffusion across products and countries. *Marketing Sci.* **21**(1) 97–114.

Urban, G. L., J. R. Hauser. 1993. *Design and Marketing of New Products*, 2nd ed. Prentice Hall, Englewood Cliffs, NJ.

Vanhonacker, W., D. R. Lehmann, F. Sultan. 1990. Combining related and sparse data in linear regression models. *J. Bus. Econom. Statist.* **10**(October) 427–435.

Accepted by Dipak C. Jain; received July 2, 1999. This paper was with the authors 23 months for 2 revisions.