Marketing Science

Real-Time Evaluation of E-mail Campaign Performance

André Bonfrer, Xavier Drèze,

Please scroll down for article—it is on subsequent pages

Real-Time Evaluation of E-mail Campaign Performance

André Bonfrer
Lee Kong Chian School of Business, Singapore Management University,
Singapore 178899, andrebonfrer@smu.edu.sg

Xavier Drèze
The Wharton School of the University of Pennsylvania, Philadelphia, Pennsylvania 19104,
xdreze@wharton.upenn.edu

We develop a testing methodology that can be used to predict the performance of e-mail marketing campaigns in real time. We propose a split-hazard model that makes use of a time transformation (a concept we call virtual time) to allow for the estimation of straightforward parametric hazard functions and generate early predictions of an individual campaign's performance (as measured by open and click propensities). We apply this pretesting methodology to 25 e-mail campaigns and find that the method is able to produce in an hour and fifteen minutes estimates that are more accurate and more reliable than those that the traditional method (doubling time) produces after 14 hours. Other benefits of our method are that we make testing independent of the time of day and we produce meaningful confidence intervals. Thus, our methodology can be used not only for testing purposes, but also for live monitoring. The testing procedure is coupled with a formal decision theoretic framework to generate a sequential testing procedure useful for the real time evaluation of campaigns.

Key words: database marketing; e-mail; pretesting; advertising campaigns
History: Received: November 2, 2006; accepted: January 25, 2008; processed by Michel Wedel. Published online in *Articles in Advance* October 7, 2008.

1. Introduction

E-mail can be a powerful vehicle for marketing communications. Many marketers favor this medium because it provides a cheaper and faster way to reach their customers. Further, the online environment allows marketers to more accurately measure consumers' actions (Weible and Wallace 2001). This is a boon for marketing scientists in their desire to increase the effectiveness of marketing efforts and measure the return on investment (ROI) of marketing expenditures.

Although e-mail response rates started out high (especially when compared with those reported for online and offline advertising), they declined over time and are now below 2.5% (Direct Marketing Association 2005). Finding ways to raise these response rates is critical for e-mail marketers. A useful tool to achieve this is an effective e-mail testing methodology. Identifying potential strengths and weaknesses of the content (the e-mail creative) and the target population *before* the e-mail is sent at full scale can help marketers improve the response rates for their campaigns.

Testing elements of the marketing mix is not new to marketing scientists. In new product development (and distribution) the ASSESSOR model (Silk and Urban 1978, Urban and Katz 1983) has been used

for decades to forecast the success of new products through laboratory based test marketing. Moe and Fader (2002) use advance purchase orders made via the CDNOW website to generate early indicators of new product sales for music CDs. In advertising, the efficacy of an advertising campaign is assessed using a battery of tests designed to identify the best creative campaign to use (e.g., the most persuasive or the most memorable) using selected members of the target audience. Field experiments with split-cable television technology have also been used to study the impact of advertising on brand sales (Lodish et al. 1995, Blair 1988).

Testing is also popular in Internet marketing applications (Drèze and Zufryden 1998). Online advertisers track banner ad performance in real time to identify the appeal (click-through) of various advertising creatives. Click-stream models can be implemented to test the appeal of content by measuring the click-through rates (CTR) or website *stickiness* (Bucklin and Sismeiro 2003). Eye tracking technology may be used to identify where (and if) a customer is viewing the advertising message embedded on a webpage (Drèze and Hussherr 2003).

In direct marketing, modeling techniques have been developed to help marketers select the right customer

group for a given content. Bult and Wansbeek (1995) build a regression model to predict each customer's likelihood of responding to a direct marketing communication, and then select which customers should be contacted by explicitly maximizing the expected profits generated by each communication. Bitran and Mondschein (1996) take the profit maximization objective a step further by incorporating inventory policies (inventory and out-of-stock costs) into the decision. Gönül and Shi (1998) extend this model by allowing customers to optimize their purchase behavior over multiple periods (i.e., both the firm and the customer are forward looking). Recognizing that customers can order from old catalogs and that one can still garner new sales from old customers who were not sent a new catalog, Gönül et al. (2000) propose a hazard function model of purchases where customers are sent a catalog only if the expected profits with the additional mailing exceeds the profits without the mailing. Elsner et al. (2004) use Dynamic Multilevel Modeling (DMLM) to simultaneously optimize customer segmentation and communication frequency. Gönül and Ter Hofstede (2006) find the optimal mailing policy for a catalog marketer given individual level predictions of customers' order incidence and volume.

While some of these methods may be adapted to the context of e-mail marketing, some unique features of e-mail present several new modeling challenges. First, many firms have implemented tracking technologies for e-mail campaigns that monitor whether and when a customer responds to an e-mail. Given the goal of real-time testing, it is essential that we make full use of this continuous time data.

Second, in contrast to a typical click-stream setting, e-mail communications are initiated by the firm rather than the customer. This adds a layer of complexity in that, while the delivery of an e-mail is often close to instantaneous, there is a delay between the time the e-mail is sent and the time it is opened. This delay will depend on how often users check their e-mail.

A third difference in e-mail marketing involves the lead time for generating both the creative and the execution of the campaign. While e-mail delivery is neither free nor instantaneous, even a large e-mail campaign can be sent at relatively low cost and delivered in a matter of hours. Consequently, campaigns are short lived and often run with short lead times and compressed deadlines—such as weekly (e.g., American Airlines, Travelocity, The Tire Rack), bi-weekly (e.g., Longs Drugstore), or even daily (e.g., Sun Microsystems). These short lead times place significant constraints on testing.

For these reasons, effective e-mail marketing communication requires a testing methodology that can generate predictions of open incidence and the CTR (CTR = clicks/opens) of any e-mail campaign as quickly and accurately as possible. Our paper describes the development and performance of such a model. We begin by developing a split-hazard model of open behavior using a log-logistic hazard function to predict the distribution of open times. Click behavior is then modeled using both a censored split-hazard model and a simpler binomial model. To help produce stable estimates even when data are sparse (a common occurrence when trying to test campaigns in a short amount of time), we use Bayesian shrinkage estimation with correlated priors between open and clicks.

When applying our model to field data, we find it necessary to account for intraday variations in customer responses (e.g., to account for fewer e-mails opened at night). Consequently, we develop a concept of virtual time that allows us to produce a model that fits the data well while keeping the specification simple. Virtual time involves adjusting the speed of time through the day to adapt to the marketers' and customers' availability. For instance, compare an e-mail sent at noon to an e-mail sent at two in the morning. While both e-mails might ultimately be opened, it is likely that the delay between send and open will be smaller for the e-mail sent in the middle of the day than for the e-mail sent in the middle of the night simply because the average recipient is more likely to be awake and checking his e-mail during the day than during the night. To reflect that recipients' availability fluctuates throughout the day, we speed up and slow down time. An hour of real time at noon (when customers are available) could be expanded to two hours of virtual time to reflect that a lot can happen during this interval while an hour at two in the morning (when most customers are asleep) could be compressed to half an hour to reflect that little would happen in the middle of the night. Using virtual time allows us to keep the model specification simple. This makes shrinkage straightforward and allows for an easy interpretation of the model parameters.

Our model offers a number of substantial practical benefits. (1) Fast evaluation of a campaign allows for early warnings about the probable success or failure of the campaign. This can lead to timely go/no-go decisions for either campaigns or creative the selected e-mail. (2) The model provides diagnostic information that can be integrated in a formal decision process to improve the results of an underperforming campaign or discard any campaign that does not perform above some threshold level of response. (3) Only a small sample is required for testing. The small sample size makes it easy to test the effectiveness of multiple advertising copies or to test the reaction of different target groups of customers. (4) Our process incorporates historical data and thus leads to better decisions as more campaigns are tested.

Table 1   Campaign Summary Statistics

	Mean	Minimum	Maximum	Std. dev.
E-mails				
Sent/campaigns	24,681	5,171	84,465	20,393
Opens/campaigns	4,457	600	13,116	3,877
Clicked/campaigns	387	53	1,444	360
Open rate	0.181	0.081	0.351	0.073
Click-through | open	0.096	0.035	0.303	0.061
Click-through rate	0.016	0.005	0.105	0.021
Doubling time (hours)	13.76	4	29	5.57
First open (minutes)	6.41	0.07	14.68	2.08
First click (seconds)	2.08	1.0	6.0	1.656
Number of campaigns	25			

Note. All campaign statistics are reported based on real time.

2. Research Setting and Data Description

We calibrate and test our models using a database of 25 e-mail campaigns sent as part of the online newsletter of an entertainment company (see Table 1 for summary statistics of the e-mail campaigns). Most of the e-mails are in the form of promotions aimed at inducing customers to purchase a movie title online or offline, or to click on links to access further content. (Different campaigns have different purposes.) Each e-mail has a subject line displaying the purpose of the promotion (e.g., "*Spiderman III* now out on DVD!"). The main body of the e-mail is only visible after the recipient has opened the e-mail. Within the body of the e-mail, recipients can click on various links to learn more about the promotion or go to a movie-specific website. It is important to note that clicks can only occur if a recipient opens the e-mail.

Figures 1(a) and 1(b) present histograms of the time (in hours) it takes for customers to open the e-mail after it is sent and the time (in minutes) it takes customers to click on the e-mail once opened. Given our objective of reducing the time allocated to testing, several features of our data are highly pertinent to model construction.

1. Most e-mails are opened within 24 hours of sending; clicks occur within a minute of opening.

2. It takes a few hours for response rate to build. There is a relatively low level of e-mail activity immediately after a campaign is launched, followed by a buildup around two hours later.

3. The histogram of the delay between send and open (Figure 1(a)) reveals a multimodal pattern during the first 24 hours after an e-mail is sent. This pattern is also visible on individual campaign histograms.

The first feature requires that a rapid testing model of open and click rate work well with censored data. Indeed, by shortening the testing time, we reduce the amount of uncensored data available to us. The second feature suggests that we must be careful about

Figure 1   Histograms of Elapsed Time Between Sent and Open (a) and Between Open and Click (b) Events, Across All Campaigns
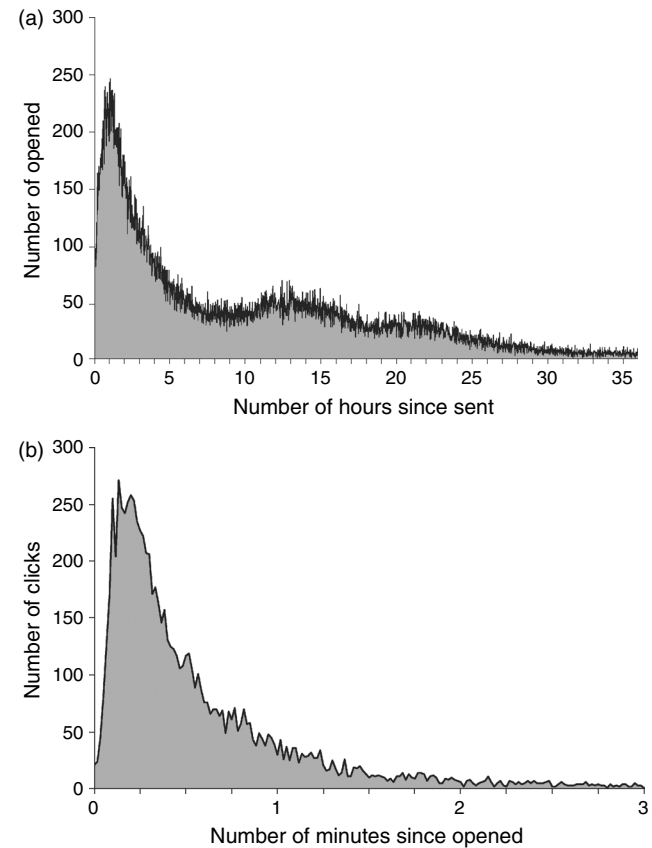
our assumptions about the data generation process when building our model. This is particularly important in our case, as we are trying to make predictions about the entire distribution of e-mail activity based on only a few hours of activity.

The multimodal pattern found in the time until opening is troublesome as it does not conform to any standard distribution and might be difficult to capture with a simple model. To understand what may be driving this multimodal pattern, we plot the distribution of e-mail openings throughout the day (see Figure 2). This graph shows considerable variation. There are fewer e-mails opened late at night and early in the morning than during the day. We refer to this pattern as intraday seasonality. In the next section we show how this seasonality may be the cause of the multimodal feature of Figure 1(a).

3. Model Setup

We develop a rapid testing methodology for a specific application: the testing of online e-mail campaigns. Rapid testing provides early feedback on whether a campaign is likely to be successful or not. In the spirit of traditional testing models, it is important that our methodology consumes as few resources as possible. Ideally, the model would also be parsimonious.

Figure 2   Distribution of E-mail Opens Through the Day (Pacific Standard Time)
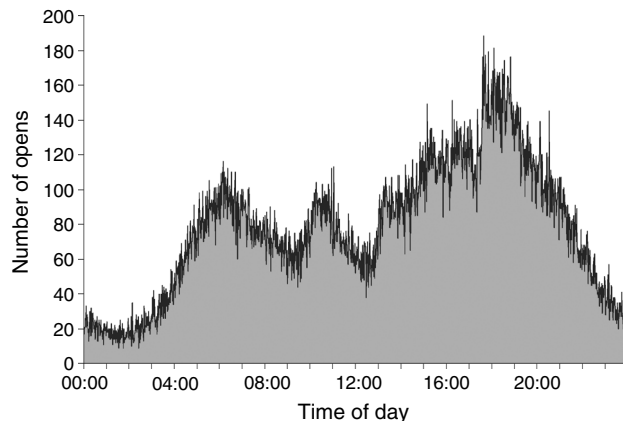
It would estimate quickly such that a test could be implemented in real time and would allow for the monitoring of a campaign as it is being sent. Indeed, an overly complex or overparameterized model that takes hours to generate predictions would defeat the purpose of rapid testing.

We first describe in more detail how the model accommodates intraday seasonality. Next, we develop a split-hazard model of open and click probabilities that takes into account the possibility that some e-mails are never opened or clicked on. We then derive the shrinkage estimators for the open and click models and state the likelihood function used in the estimation.

3.1.   From Physical Time to Virtual Time

A traditional approach to handling seasonality, such as that displayed in Figures 1(a) and 2, is to introduce time-varying covariates in the model. There are two main problems with this approach. First, the covariates are often ad hoc (e.g., hourly dummies). Second, they often make the other parameters less interpretable (e.g., a low open rate during peak hours could be larger than a high open rate during off peak hours). To alleviate these concerns, we build on the approach proposed by Radas and Shugan (1998), (hereafter RS), who deseasonalized a sales process by changing the speed at which time flows. They showed that by speeding up time during high seasons, and slowing down time during low seasons, one can create a new (virtual) time series that is devoid of seasonality. The benefits of this approach, assuming that one has the right seasonality pattern, is that one can use straightforward models in virtual time and easily interpret the meaning of the parameters of these models.

The effectiveness of the RS approach hinges on having a good handle on the seasonal pattern present in the data. In their application (the movie industry) they produce seasonal adjustments by combining past sales data with industry knowledge. A shortcoming of this approach is that some of the seasonality may be endogenous to the firms' decisions. For instance, if movie studios believe that Thanksgiving weekend is a "big" weekend, they may choose to release their movies during that weekend rather than during off-peak weekends (Ainslie et al. 2005). Thus, part of the seasonality observed during Thanksgiving will stem from the fact that more consumers have the time and desire to see movies on that weekend (consumer-induced seasonality), and part of the seasonality will stem from the fact that more movies are available (firm-induced seasonality). If one uses past data as a base for seasonal adjustment without considering the decisions of the firm, one can potentially overcorrect and attribute all the seasonal effects to consumer demand while it in fact also partly stems from firm supply.

In our case, we are confronted by both consumer- and firm-induced seasonality. For instance, the average consumer is much less likely to open e-mails at four in the morning than at four in the afternoon. Similarly, firms do not work 24 hours a day. If we look at when the firm sends its e-mail (Figure 3), we observe little (but some) activity during the night, then a peak at eight in the morning, a peak at noon, and a lot of activity in the afternoon. It is likely that these peaks are responsible for some of the increase in activity we see in Figure 2 at similar times.

To separate consumer- from firm-induced seasonality, we benefit from two features of our modeling environment not present in RS. First, we have continuous time individual level data. While RS had to work with aggregate weekly measures, we know the exact time each e-mail is sent and opened. Second, while a movie can open on the same day throughout the country, e-mails cannot all be sent at the same time. E-mails are sent sequentially; for example, a million-e-mail campaign can take up to 20 hours to

Figure 3   Distribution of E-mail Sends Through the Day (Pacific Standard Time)
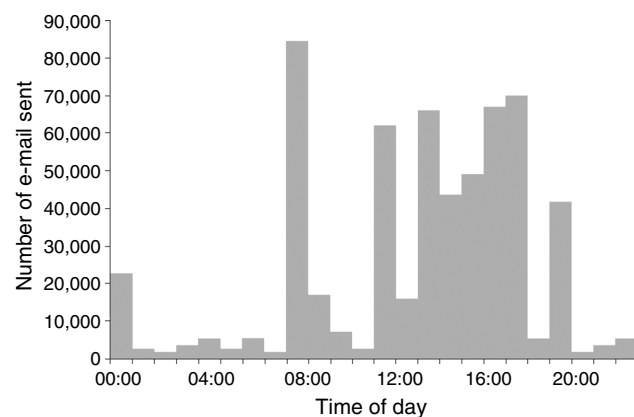
send. Thus, we can simulate an environment that is devoid of firm based seasonality by resampling our data such that the number of e-mails sent at any point in time is constant through the day (i.e., Figure 3 for such a firm would be flat).

To resample the data, we proceed in three steps. First, for each minute of the day, we collect all e-mails that were sent during that minute. Second, we randomly select with replacement 100 e-mails from each minute of the day (144,000 draws). Third, we order the open times of these 144,000 e-mails from 00:00:00 to 23:59:59 and associate with each actual open time a virtual time equal to its rank divided by 144,000. The relationship between real and virtual time based on their cumulative density functions is shown in Figure 4. This represents the passing of time as seen by consumers independent of the actions of the firm.

We can use the relationship depicted in Figure 4 to compute the elapsed virtual time between any two events. For instance, if an e-mail were sent at midnight and opened at two in the morning, we would compute the elapsed virtual time between send an open by taking the difference between the virtual equivalent of two A.M. (i.e., 00:29:44 virtual) and midnight (i.e., 00:00:00 virtual) to come up with 29 minutes and 44 seconds. Similarly, if the e-mail had been sent at noon and opened at two P.M., then the elapsed virtual time would be 11:05:10–09:08:30 = 1 hour 56 minutes and 40 seconds.

Our approach does not account for another form of endogeneity that could arise if the firm were to strategically send good and bad e-mail campaigns at different times of the day such that the two types of campaigns would not overlap. Fortunately, this was not company policy in our application—e-mail campaigns were sent out whenever they were ready.

Figure 4   Within-Day Cumulative Density Functions of Real and Virtual Time
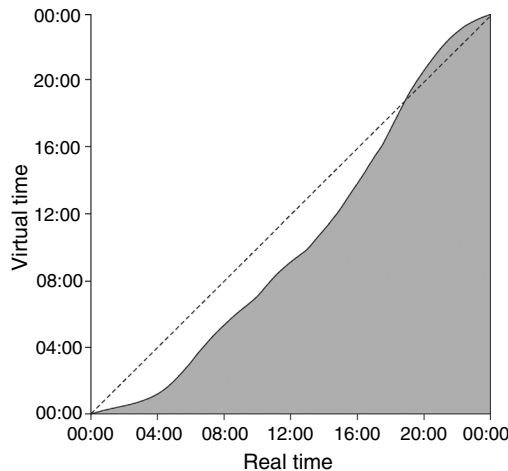
Note. Real time is distributed uniformly throughout the day, resulting in a 45-degree line.

3.2. A Split-Hazard Model of Open and Click Time

The time it takes for a customer to open an e-mail from the time it is sent and the time it takes to click on an e-mail from the time a customer opens it are both modeled using a standard duration model (e.g., Moe and Fader 2002, Jain and Vilcassim 1991). Because both actions can be modeled using a similar specification, we discuss them interchangeably. Starting with opens, we account for the fact that in an accelerated test, a failure to open an e-mail means one of two things. Either recipients are not interested in the e-mail or they have not had a chance to see it yet (i.e., the data is censored). Of course, the shorter the time allocated to a test, the higher the likelihood that a nonresponse is indicative of censoring rather than lack of interest. Thus, we model the open probability and the open time simultaneously in a right-censored split-hazard model (Kamakura et al. 2004, Sinha and Chandrashekaran 1992).

The probability that a customer will open or click an e-mail varies from campaign to campaign and is denoted with δ_k^e, where e is a superscript identifying different campaigns, and the subscript k denoting an open ($k = o$) or click ($k = c$). The likelihood function is constructed as follows. We start with a basic censored hazard rate model of the open or click time distribution:

$$L_k^e(t_k^e, T^e, R_k^e \mid \Theta_k^e) = \prod_{i=1}^{N^e} f(t_{ik}^e \mid \Theta_k^e)^{R_{ik}^e} S(T^e \mid \Theta_k^e)^{1-R_{ik}^e}, \quad (1)$$

where

e	is a superscript that identifies a specific e-mail campaign,
k	is a subscript that identifies the model used ($k \in \{o = \text{open}, c = \text{click}\}$),
i	is an index of recipients,
N^e	is the number of recipients for e-mail e,
R_{ik}^e	is 1 if recipient i opened/clicked e-mail e before the censoring point T^e,
T^e	is the censoring point of campaign e,
t_{ik}^e	is the elapsed time between send and open (open and click) in the event that the recipients opened (clicked) the e-mail,
$f(t \mid \Theta)$	is the pdf for time t, given a set of parameters Θ,
$S(t \mid \Theta)$	is the corresponding survival function at time t,
Θ_k^e	is a vector of parameters to be estimated.

The likelihood function in (1) needs to be adjusted to account for the fact that some recipients will never open or click on the e-mail. Let δ_k^e denote the probability that e-mail e will be opened or clicked.

Then the likelihood component for action k at censor time T^e is

$$L_k^e(t_k^e, T^e, R_k^e \mid \Theta_k^e) = \prod_{i=1}^{N^e} [\delta_k^e f(t_{ik}^e \mid \Theta_k^e)]^{R_{ik}^e}$$
$$\cdot [\delta_k^e S(T^e \mid \Theta_k^e) + (1 - \delta_k^e)]^{(1-R_{ik}^e)}. \quad (2)$$

The estimation of δ_k^e and Θ_k^e for any parametric hazard function can be performed by maximum likelihood estimation.

We tested a variety of candidate hazard rate specifications (including the exponential, Weibull, log-normal, and the more general Box-Cox specification) to model the open and click processes (see the Technical Appendix, which can be found at http://mktsci.pubs.informs.org, for details and results). Overall, the tests revealed that the log-logistic distribution is best suited for our application. The probability density function and the survivor function for the log-logistic are (see Kalbfleisch and Prentice 1985)

$$f(t \mid \alpha, \lambda) = \frac{\lambda \alpha (\lambda t)^{\alpha-1}}{(1 + (\lambda t)^\alpha)^2} \quad \text{and}$$
$$S(t \mid \alpha, \lambda) = \frac{1}{1 + (\lambda t)^\alpha}, \quad (3)$$

where $\lambda > 0$ is a location parameter and $\alpha > 0$ is a shape parameter. Consistent with previous notation, we refer to the shape and location parameters for any given campaign (e) and e-mail response action ($k \in \{o, c\}$) as α_k^e, and λ_k^e, respectively. Depending on the value of α, the log-logistic hazard is either monotonically decreasing ($\alpha \leq 1$) or inverted U-shape ($\alpha > 1$) with a turning point at $t = (\alpha - 1)^{1/\alpha}/\lambda$.

3.3. Shrinkage Estimators

As in most practical applications, we benefit from having data available from past campaigns. This information can be used to improve the performance of our model. Specifically, we use the parameters estimated from past campaigns to build priors for new campaigns. This is especially useful at the beginning of a campaign when data are sparse.

Because we use the log-logistic hazard function in our application, each of the split-hazard models has three parameters (α, λ, δ). When building priors for these parameters, it is reasonable to assume that the open and click rates for a campaign are correlated. Indeed, broad appeal campaigns should yield both high open and high click rates while unappealing campaigns would exhibit both low open and low click rates. To accommodate the possibility of correlated open and click rates, we use a bivariate beta as the prior distribution for the (δ_o, δ_c). As in Danaher and

Hardie (2005), we implement the bivariate distribution proposed by Lee (1996)

$$(\delta_o, \delta_c) \sim g(x_1, x_2) = f(x_1 \mid a_1, b_1) f(x_2 \mid a_2, b_2)$$
$$\times [1 + \omega(x_1 - \mu_1)(x_2 - \mu_2)],$$

where $f(x_i \mid a_i, b_i)$ is the univariate beta density for x_i given parameters (a_i, b_i), μ_i is the mean of the univariate beta for x_i, $\omega = \text{Corr}(x_1, x_2)/(\sigma_1.\sigma_2)$ is a function of the correlation between x_1 and x_2, and where $\sigma_i^2 = a_i b_i/[(a_i + b_i)^2(a_i + b_i + 1)]$ is the variance of the univariate beta.

There is no reason to believe that the hazard rate parameters for clicks and opens will be correlated because one process (opens) depends on the availability of recipients to open e-mails while the other process (clicks) is conditional on the recipients having opened the e-mail (thus being available), and depends on the amount of processing needed to understand the e-mail and react to it. It is possible, however, that the shape and location parameters of each hazard rate are correlated. Thus, we use a bivariate log-normal distribution for the prior on α_k and λ_k

$$(a_k, \lambda_k) \sim \text{Log-Normal}(\mu_k, \Sigma_k). \quad (4)$$

For a given campaign, we use the method of moments to estimate the parameters $(a_o, b_o, a_c, b_c, \text{Corr}(Open, Click), \mu_o, \Sigma_o, \mu_o, \Sigma_o)$ based on the parameters $(\alpha_o, \lambda_o, \delta_o, \alpha_c, \lambda_c, \delta_c)$ obtained from all other campaigns. The correlation, $\rho_{\alpha, \lambda}^{\text{LN}}$, between parameters α_k and λ_k of the log-normal distribution is adjusted for possible small sample bias using the correction factor described in Johnson and Kotz (1972, p. 20):

$$\tilde{\rho}_{\alpha, \lambda}^{\text{LN}} = \frac{\exp(\rho_{\alpha, \lambda}^{\text{LN}} \sigma_\alpha \sigma_\lambda) - 1}{\sqrt{\{\exp(\sigma_\alpha^2) - 1\}\{\exp(\sigma_\lambda^2) - 1\}}}. \quad (5)$$

The combination of the likelihood functions given in (2) with the log-logistic hazard specification in (3) and the bivariate beta and log-normal priors gives the following likelihood function to be estimated for a given campaign (omitting the e superscript for clarity):

$$L(t_o, R_o, t_c, R_c, T \mid \Theta)$$
$$= \text{LN}(\lambda_o, \alpha_o \mid \mu_o, \Sigma_o)\text{LN}(\lambda_c, \alpha_c \mid \mu_c, \Sigma_c)$$
$$\times \text{Beta}(\delta_o \mid a_o, b_o)\text{Beta}(\delta_c \mid a_c, b_c)$$
$$\cdot \left[1 + w\left(\delta_o - \frac{a_o}{a_o + b_o}\right)\left(\delta_c - \frac{a_c}{a_c + b_c}\right)\right]$$
$$\times \prod_{i=1}^N \left[\delta_o \frac{\lambda_o \alpha_o (\lambda_o t_{io})^{\alpha_o - 1}}{(1 + (\lambda_o t_{io})^{\alpha_o})^2}\right]^{R_{io}}$$
$$\cdot \left[1 - \delta_o\left(1 - \frac{1}{1 + (\lambda_o T)^{\alpha_o}}\right)\right]^{(1-R_{io})}$$

$$\times \prod_{i=1}^{N}\left(\left[\delta_c \frac{\lambda_c \alpha_c (\lambda_c (t_{ic} - t_{io}))^{\alpha_c - 1}}{(1 + (\lambda_c (t_{ic} - t_{io}))^{\alpha_c})^2}\right]^{R_{ic}}\right.$$

$$\left.\cdot \left[1 - \delta_c \left(1 - \frac{1}{1 + (\lambda_c (T - t_{io}))^{\alpha_c}}\right)\right]^{(1 - R_{ic})}\right)^{R_{io}}. \quad (6)$$

3.4. An Alternative Approach to Estimating Click Rates

Although theoretically sound, using a split-hazard model to estimate the parameters of the click times (conditional on an open) might be overly complex. Because most consumers are likely to click on an e-mail within seconds of opening it, few click observations are right censored. Thus, we can test a simpler model for the estimation of click rates that does not take censoring of clicks into account. Our hope is that this more parsimonious model will perform better at the beginning of a test, when few data points are available.

Formally, we assume that clicks follow a binomial process. Using the same bivariate beta prior for (δ_o, δ_c) as we did in the full model, the likelihood function for this simplified model is

$$L_S(t_o, R_o, t_c, R_c, T \mid \Theta)$$
$$= \text{LN}(\lambda_o, \alpha_o \mid \mu_o, \Sigma_o)\text{LN}(\lambda_c, \alpha_c \mid \mu_c, \Sigma_c)$$
$$\times \text{Beta}(\delta_o \mid a_o, b_o)\text{Beta}(\delta_c \mid a_c, b_c)$$
$$\cdot \left[1 + w\left(\delta_o - \frac{a_o}{a_o + b_o}\right)\left(\delta_c - \frac{a_c}{a_c + b_c}\right)\right]$$
$$\times \prod_{i=1}^{N}\left[\delta_o \frac{\lambda_o \alpha_o (\lambda_o t_{io})^{\alpha_o - 1}}{(1 + (\lambda_o t_{io})^{\alpha_o})^2}\right]^{R_{io}}$$
$$\cdot \left[1 - \delta_o \left(1 - \frac{1}{1 + (\lambda_o T)^{\alpha_o}}\right)\right]^{(1 - R_{io})}$$
$$\cdot \left[\delta_c^{R_{ic}} (1 - \delta_c)^{(1 - R_{ic})}\right]^{R_{io}}. \quad (7)$$

3.5. Comparisons with Benchmarks: The Doubling Method

Before discussing the application of our model, we would like to draw a comparison with existing approaches for predicting the success rate of direct marketing campaigns. The most common model used by practitioners is the *doubling* method (Nash 2000; it is sometimes referred to as the half-life analysis as in Hughes 2006). This method involves first looking at the responses of past direct marketing campaigns and computing the amount of time it takes for 50% of the responses to be received (the *doubling point*). The analyst then uses the heuristic that for any future campaigns, the predicted total number of responses is equal to double the number of responses observed at the doubling point. In direct marketing, for first class mailing firms will wait a minimum of two weeks; for

third class mailing, they will wait four weeks. Once the test results are in, the decision maker needs to make a go/no-go decision based on profitability. In our case, the doubling point is 14 hours (see Table 1).

The doubling method is a powerful and simple heuristic. It makes three implicit assumptions. First, it assumes that not everybody will respond. Second, it assumes that it takes time for people to respond. Third, it assumes that the timing of the responses is independent of the rate of response and constant across campaigns. As a nonparametric method, it makes no assumptions about the underlying response process, nor does it provide ways to test whether the current campaign conforms to the data collected from previous campaigns or runs faster or slower than expected. Hence, it does not provide ways to evaluate whether a current test should be run longer or could be finished early. Our model provides this important piece of information.

In essence, the doubling method aggregates time into two bins—each containing half of the responses. This aggregation loses vital timing information that could be used to better model the response process. If one wanted to speed up the doubling method, any other quantile could be used to perform a test. For example, the 25th percentile would give a *quadrupling* method.

Another benchmark examined in this paper is waiting for a predetermined time period and counting the number of opens and clicks observed at that time. The campaign would be deemed a success if these numbers exceed a predefined threshold. While this benchmark does not predict open and click rates, it can be used as a decision rule. We discuss this and other decision rules in more detail in §3.2.

4. Application of the Model to E-mail Campaign Pretesting

We now apply our models to the data from the e-mail campaigns described earlier and evaluate the relative predictive validity of various models. Estimation of the parameters is based on a numerical maximization of the censored likelihood functions in Equations (6) and (7) using SAS (with the quasi-Newton optimization algorithm). For both real and virtual time, we compare the predictions of our models with those of the doubling method. We then use our model in a campaign selection simulation and benchmark it against other possible heuristics.

4.1. Simulation and Validation

The main purpose of the simulation and validation stages is to validate the models proposed in the paper by studying the accuracy of the predictions they make out-of-sample. We also want to find out which of

the models has the best predictive performance and whether we can generate estimates that are useful for decision making within a short amount of time (say hours) such that testing is feasible for campaign planning. When doing so, we compare the models based on real time (no time transformation) and virtual time, for both the full model and the simplified click formulation. In summary, we fit and validate the following four models:

(1) Full model in real time (using Equation (6) and real time).

(2) Full model in virtual time (using Equation (6) and transformed time).

(3) Simplified model in real time (using Equation (7) and real time).

(4) Simplified model in virtual time (using Equation (7) and transformed time).

In each simulation, we adopt the perspective of a marketer who wishes to pretest campaigns before committing to the final send. To this end, we look at each campaign assuming that the remaining $(E - 1)$ campaigns have been completed. Prior to the test, we know nothing about a focal campaign except the number of e-mails that need to be sent. However, we can use all the information collected through the other campaigns to build our priors.

We set the sample test size at 2,000 e-mails. As a robustness check, we varied the sample size between 1,000 and 2,000 in 200-e-mail increments; this did not yield any substantive difference in findings. We simulated different test lengths in 30-minute increments, ranging from 30 minutes to 8 hours. For each test length, any e-mail that had been opened prior to the simulated end of test was used in the noncensored component of the log-likelihood. All other observations are coded as censored (e.g., if an open were recorded 45 minutes after an e-mail was sent, this observation would be coded as censored when we simulate a 30-minute test, and as an open for a 60-minute test and any subsequent tests). Based on this set of censored and uncensored observations, the parameters of the full and simplified models are estimated using priors based on all other campaigns.

As a practical matter, the distributions of open and click times exhibit long tails, such that some responses continue to come in long after a campaign has run its course. Historical data reveal that 99% of all e-mail responses are observed within 3 weeks of being sent. Typically, the company conducts postcampaign debriefing 2–3 weeks after the e-mails are sent. Thus, the cutoff date is set at 3 weeks (504 hours), and the number of opens and clicks observed at that time is used as the true value the models need to predict. Our forecast of the number of opens and clicks at 3 weeks is constructed using the parameter estimates for each

of the censored samples; standard errors are calculated using the Delta method. The cumulative distribution of opens and clicks at 504 hours is calculated using

(1) $\widehat{Open}_{504} = \hat{\delta}_o^e \times F(504 \text{ hours} \mid \hat{\alpha}_o^e, \hat{\lambda}_o^e) \times$ (Number sent for campaign e), and

(2) $\widehat{Click}_{504} = \hat{\delta}_c^e \times F(504 \text{ hours} \mid \hat{\alpha}_c^e, \hat{\lambda}_c^e) \times$ (Estimated opens for campaign e).

4.1.1. Predictive Performance of the Models. The full results for each campaign and each censoring point consist of a set of parameters $(\delta_o^e, \alpha_o^e, \lambda_o^e)$ for opens and $(\delta_c^e, \alpha_c^e, \lambda_c^e)$ for clicks. To assess the performance of the models in the precampaign tests, we calculate the mean absolute deviation (MAD) for the predicted number of clicks and opens at each censoring point. The MAD expresses the absolute prediction error in terms of the number of clicks and number of opens averaged across campaigns and is calculated by comparing the predicted opens and clicks with the true opens and clicks.

The MAD statistics for the first four hours are reported in Table 2. For all the open models tested, the MAD declines as testing time increases. Based on the MAD prediction error, the virtual-time models outperform the real-time models. The virtual-time models achieve a prediction error that is equal or better than that of the doubling method in as little as three hours. This represents a reduction in test time of close to 80%. For real time, it takes slightly over four hours before the tested model outperforms the doubling time model. At the doubling time point (14 hours), all the tested models were superior in predictive performance to the doubling method.

Although, in the interest of space, we do not present the standard errors for the various campaigns' parameter estimates, we find that using virtual time also produces tighter and more stable estimates (i.e., smaller confidence intervals) than the model using real-time data. Because virtual-time models also estimate faster, they are well suited to test the performance of a campaign in compressed-time applications.

The prediction errors for clicks exhibit a similar pattern. The simplified model slightly outperforms the full model in the early stages of testing (up to about 5 hours for the real-time model, and 1.5 hours for the virtual-time model). The simplified model in virtual time achieves a predictive performance similar to the doubling method in only 2.5 hours—an 82% reduction in testing time.

4.2. A Campaign Selection Decision Rule
The results reported in the preceding section demonstrate that our models can produce estimates of the open and click rates faster and more accurately than the doubling method. While these results show a clear improvement in speed and accuracy, a natural

Table 2   **Mean Absolute Deviation (MAD), by Model, for Split-Hazard Parameters and for Selected Censoring Points**

	Method					
	Opens (MAD for $\hat{\delta}_o$)		Clicks (MAD for $\hat{\delta}_c$)			
Censor time (hh:mm)	Real time, simple model (Equation (7))	Virtual time, simple model (Equation (7))	Real time, simple model (Equation (7))	Real time, full model (Equation (6))	Virtual time, simple model (Equation (7))	Virtual time, full model (Equation (6))
0:30	2,308.7	2,545.8	258.411	254.6	293.9	303.5
1:00	1,522.6	2,084.4	185.152	181.8	249.8	254.9
1:30	1,222.3	1,791.5	170.594	176.2	199.3	201.2
2:00	1,165.1	1,446.2	187.473	195.2	167.7	167.1
2:30	1,161.4	1,110.4	161.224	165.2	**146.4**	**148.8**
3:00	1,101.5	**906.6**	**147.196**	**148.1**	121.7	122.1
3:30	977.9	763.7	128.365	129.5	99.8	100.0
4:00	**914.5**	703.9	126.8	128.7	95.5	95.2

Notes. For the doubling time method, MAD at 14 hours for clicks is 156.7 and for opens it is 917.5. MAD numbers in boldface indicate the first time (in 30-minute increments) that the model outperforms the doubling time method.

question arises: Are the benefits of the new models substantial from a managerial standpoint? To investigate this in the context of our application, we develop a formal decision rule based on the testing methodology developed in §2. We take the perspective of a campaign manager who wishes to improve the ROI of his marketing actions by eliminating underperforming campaigns. One possible reason for not wanting to use a low-yield campaign is that the e-mails cost more to send than the expected returns. Such underperforming campaigns also represent an opportunity cost in that they tie up resources that could be used to send more profitable campaigns. Furthermore, sending undesirable material could lead to higher customer attrition (Hanson 2000).

Imagine that our manager wishes to eliminate any campaigns with a CTR lower than the historical average of 2% (across 25 campaigns). Clearly, rejecting any campaigns with response rates lower than the historical average will have the consequence of improve response rates across future campaigns and help the firm achieve growth objectives (Pauwels and Hanssens 2007). If the manager wishes to further improve response rates, it is also possible to set this threshold higher. To make a go/no-go decision on campaigns, the manager could consider one of the following four decision rules:

Rule 1: Doubling Method. Under the doubling method, the decision rule is quite simple. For a given campaign, send out the test sample, wait 14 hours, and observe the CTR. If the CTR is greater than 1% (1/2 of 2%), then run the full campaign; otherwise cancel it.

Rule 2: The Horse Race.[1] If the manager needs only to make a go/no-go decision rather than predict the actual outcome of the campaign, he or she can compare the results of the current campaign to

the other campaigns at some point during the testing phase. If these results are better than average, go ahead with the full campaign, or else, cancel it. In essence, the manager is entering the campaign in a horse race, pitching the current campaign against previous ones.

This approach is simple and parameter free. It has the advantage of being faster than the doubling method and is simpler than our proposed model. If all one cares about is picking the right campaign, it may be sufficient to use this approach. We implemented this approach by using the same 2.5-hour cutting point as we use in Rule 3.

Rule 3: The Proposed Model. Under Rule 3, the manager will run the test for 2.5 hours, then fit our model and use its output to predict the long-run CTR. If the predicted result is larger than 2%, then the manager proceeds with the campaign; otherwise it is cancelled.

Rule 4: Sequential Statistical Test Under Proposed Model. Rule 3 makes a go/no-go decision based on the CTR predicted by our model. In doing so, it ignores the uncertainty around this estimate. A more astute manager would consider this uncertainty when making the go/no-go decision. There are many ways to do this. One way is to continue the test until there are enough data to conduct a statistical test that would reject the null hypothesis that the CTR is equal to 2% (we call this Rule 4). This approach is in the spirit of *adaptive learning*, i.e., technology used by the firm to adapt in real time to tested customer preferences (Sun 2006). A related approach is to incorporate a loss function into the decision (see Calabria and Pulcini 1996, Pandey 1997, Zellner 1986). This is Rule 5, which is discussed in the Technical Appendix at http://mktsci.pubs.informs.org.

Under Rule 4, the decision process adopted by the manager is as follows:

Step 1. Send 2,000 test e-mails.
Step 2. Wait *M* minutes.

---

[1] We thank an anonymous reviewer for proposing this alternative methodology.

Step 3. Fit our proposed model.

Step 4. Predict the end of campaign CTR.

Step 5. Use standard error around the prediction to test the hypothesis that CTR = 2%.

Step 6. If the test fails to reject the null hypothesis, go back to Step 2.

Step 7. If the test shows that CTR > 2% then proceed with the campaign, otherwise cancel it.

This method has the advantage of making the most of the data available to the analyst. It considers not only the point estimate CTR, but also the uncertainty around the estimate. Finally, it adapts to each campaign by cutting short the test if the results are clearly significant, or running the test longer if more precision is needed.

In our test, we set the significance level for the statistical test at $p = 0.05$ (two-tailed) and we set M at 30 minutes to limit the number of models we need to fit (25 campaigns in 30-minute increments leads to 400 runs for an 8-hour test). In practice, one could use a much shorter pooling interval. Our model estimates in a matter of seconds. One could even consider running the test in real time and reestimating the model when a new open or click is recorded.

4.2.1. Comparison of Decision Rules. The results of decision rules 1 to 4 are reported in Table 3. Seven of the 25 campaigns have a true CTR greater than 2%. If only these seven campaigns were selected, then the average CTR would be 4.45%—see the column labeled *Actual*. We use the doubling method (Rule 1) as the method to beat. In our case, the doubling method selects 11 campaigns, the 7 correct campaigns plus 4 underperforming campaigns, yielding a CTR of 3.34%.

Table 3   **Results from the Decision Rule Simulation**

	Actual	Rule 1 Doubling method	Rule 2 Horse race	Rule 3 Proposed model	Rule 4 Adaptive model
Number of campaigns selected	7	11	14	7	6
True positives		7	6	6	6
False positive		4	8	1	0
False negative		0	1	1	1
Average testing time (hours)		14:00	02:30	02:30	02:57
Minimum testing time		14:00	02:30	02:30	01:00
Maximum testing time		14:00	02:30	02:30	06:00
Click-through rate (%)	4.45	3.34	2.66	4.29	4.77
Improvement over no rule (%)	123	67	33	114	139
Revenue per e-mail sent ($)	0.027	0.016	0.009	0.025	0.031
Revenue per name (= revenue per e-mail sent × no. of campaigns) ($)	0.190	0.172	0.122	0.178	0.183

The horse race (Rule 2) does not perform as well as the doubling method and leads to a decrease in response rate of 20% (2.66% versus 3.34% CTR); it is not very discriminate and retains too many campaigns. Looking at the campaigns that are added erroneously, we find that these campaigns are the ones with the highest λ parameters in the open-hazard function. That is, these campaigns start off strong, but quickly deteriorate. Because Rule 2 only looks at the overall clicks after 2.5 hours of testing without adjusting for the timing of the clicks, it is not surprising that it is less able to discriminate between the campaigns that deteriorate quickly and overall strong performers.

Our model (Rule 3) performs better than the doubling method. The average campaign response rate increases by 28% (4.29% CTR versus 3.34%) by using our model and waiting 2.5 hours of virtual time. One can gain a further 15% (4.77% versus 3.63%, for a total improvement of 43%) by using the sequential testing rule (Rule 4) rather than a fixed time stopping rule. Rule 4 actually overperforms by selecting only six of the seven correct campaigns and having no false positives.

Looking at CTR only does not fully reflect the economic consequences of sending more or fewer campaigns. One can always improve click-through rates by being more cautious and having a high cutoff rate. This of course leads to fewer campaigns selected and thus less income for the firm (Bult and Wansbeek 1995, Steenburgh et al. 2003). To consider the economic impact of each decision rule, one can use a per e-mail sent profit function, $\Pi(\widehat{\Theta}, T) = (\delta_o b + \delta_c p) - mc$, where b represents the revenue to the firm from an open only (i.e., a billboard effect, the recipients might read the e-mail, then buy from a third party), p represents the revenue to the firm when a click occurs, and mc is the cost of sending the e-mail. For the purpose of illustration, we set mc at \$0.02, b at \$0.01, and p at \$1.00.

The last two rows of Table 3 show the average revenue per e-mail sent and the total revenue per name (assuming the same names are used in all campaigns). The economic consequences of the various decision rules are best exemplified by looking at Rules 2 and 4. Rule 2 has a revenue per e-mail sent that is only one third of the ideal (\$0.009 versus \$0.027). Nevertheless, because it results in twice as many campaigns being selected, it generates total revenue that is almost two thirds of ideal (\$0.122 versus \$0.190). In contrast, Rule 4 has a revenue per e-mail sent that is larger than the ideal (\$0.031 versus \$0.027), but it sends fewer e-mails, and thus yields a slightly smaller overall profit (\$0.183 versus \$0.190).

It should be noted that, when using a statistical test to decide when to stop testing, the testing time varies

widely across campaigns. On average, the adaptive testing rule requires 2 hours 57 minutes. The fastest campaign reaches statistical significance in as little as 1 hour. The slowest campaign needs 6 hours. In short, our tests demonstrate that our method performs better than a simple comparison of campaign performance after a fixed time (be it the doubling method rule or the horse race). Further, by taking advantage of the fact that our method produces not only a point estimate but also a distribution around the point estimate, one can further improve decision making either through a sequential testing method or the use of a loss function that reflects manager beliefs.

4.3. When Is the Best Time to Test?

Our comparison of the predictive ability for the split-hazard rate model suggests that, on average, we can learn as much in 2.5 hours as we can learn from the doubling method in 14 hours. However, it is important to remember that these 2.5 hours are measured in virtual time. In real time, the test will require more or less time depending on the time of day when it is conducted. Figure 5 shows how long 2.5 virtual hours correspond to in real time, depending on when the test starts. There appears to be a *sweet spot* in the afternoon, between 13:00 and 19:00 hours where a 2.5-virtual-hour test can be carried out in fewer than 2 actual hours (the shortest it could take would be 1 hour and 15 minutes by starting at 17:34 hours). Starting after 19 hours will impose delays as the test because it is unlikely to be finished before people go to bed; if the test is started at 22:12 hours it will take almost 6.5 hours to complete.

Given the speed at which the tests can be carried out, an e-mail marketer could implement a sequential refinement approach starting with a series of concepts tested at 09:00 hours. The results would come in by noon. Between noon and 13:00 hours, the marketer could refine the winning concept by producing

Figure 5   Test Length as a Function of Time of Day

Note. This graph plots the 2.5-hour virtual waiting time in real time based on the time of day a test is performed.

different variations of it. At 13:00 hours, the refined concepts can be tested, with the results coming in by 15:00 hours. The winning concepts of this second phase are used to produce a final set of concepts that are tested at 16:00 hours with the results out by 18:00 hours. At that time the marketer can select the final creative and send it out. This series of concept creation/testing/selection can be carried out in a day and should produce much better results than sending out untested creatives. A key advantage of our methodology is the ability to develop multiple results in quick succession, a feature that is crucial for a sequential testing procedure.

5. Discussion and Conclusion

The value of information increases with its timeliness. Knowing quickly whether a campaign is going to be successful provides the opportunity to correct potential problems before it is too late or even to stop the campaign before it is completed. It is therefore imperative to develop methods that improve both the accuracy and speed with which campaign testing can be done. In this article, we study a testing procedure that can be implemented for the fast evaluation of e-mail campaign performance.

The performance of an e-mail campaign is defined by its open and CTR. The methodology we propose predicts these rates quickly based on estimates produced with small samples of the main campaign. We propose to send 2,000 e-mails and to wait fewer than two hours to produce estimates of how the campaign will perform after three weeks. In two hours, fewer than 100 opens and fewer than 10 clicks are typically observed. The key to successful prediction of the ultimate results of an e-mail campaign based on so few data points lies in using the information to its fullest potential.

There are three elements that make our methodology successful: (1) using the appropriate model specification, (2) transforming time to handle intraday seasonality, and (3) using informative priors. Each of these three elements provides its own unique contribution to the overall fit and predictive performance of the model.

The appropriate hazard function is critical because our compressed-time tests produce observations that are heavily right censored. Thus, we are often fitting a whole distribution based only on its first quartile (or even less). We find that the best fitting parametric model for open times is the log-logistic. For modeling clicks, we find that the straight binomial process is a good descriptor of the phenomenon given that consumers respond quickly after opening an e-mail. Thus, the CTR (the total number of clicks for a campaign, unconditional on open) is best predicted using
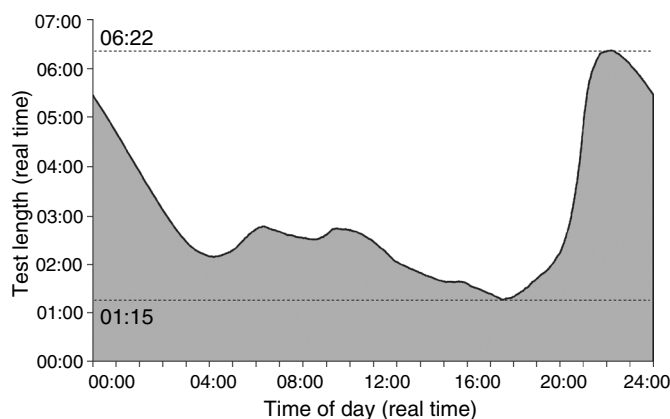
a combination of a binomial model for the clicks, and a log-logistic split-hazard model for the opens.

We apply our split-hazard model in a virtual-time setting. The virtual-time transformation removes intraday seasonality and makes our testing procedure invariant to time of day. This is a key factor in the robustness of our model in that it allows us to bypass the need to handle seasonality directly in the model and allows for a straightforward specification with only three parameters for opens and one for clicks. By limiting the number of parameters, we make the best use of our limited data and produce parameters that are directly interpretable.

Another benefit of our time transformation is that by making each campaign independent of the time of day, it is possible to compare results across campaigns, and to easily build informative priors for each of the parameters. This yields a procedure that produces meaningful estimates and confidence intervals with a minimal amount of data. It also allows a firm to conduct tests serially. That is, the firm could choose to modify either a campaign's creative or target population as the result of a test, then retest the campaign and compare the new results to the first.

Overall, by putting these three elements together, the model is capable of running a test in one hour and fifteen minutes that produces similar results to a traditional test in 14 hours (a 91% decrease in testing time). The model can be estimated within a matter of seconds, and could therefore be used in real time. Thus, our methodology can be used not only for testing but also for live monitoring. A manager could terminate an underperforming campaign early or change the campaign's creative midstream. If done right, this could significantly improve average response rates by limiting the detrimental impact of poor performing campaigns.

Our model represents a first step towards better online marketing testing. As such, we see several avenues for further research in this area. For instance, because of the lack of individual level information in our data set, we could not include covariates in our models. It is likely that adding such information, when available, would improve the fit and predictive power of the model. Further, if the data set contained many data points per recipient (we have an average of two e-mails sent per name) it would be possible to incorporate unobserved heterogeneity.

One could also include campaign covariates in the model to account for the fact that different movies will appeal to different audiences. We attempted to do so by using different priors for different genres of movies (e.g., action, romance, and others). However, because of the small number of campaigns at our disposal the analysis did not produce meaningful results (the details of the analysis and the results

are reported in the Technical Appendix, which can be found at http://mktsci.pubs.informs.org).

Another issue is the link between click and purchase behavior. The assumption is that click behavior is a measure of interest and is highly correlated to purchase. As many campaign managers are evaluated based on clicks, we feel our analysis is appropriate. However, in future applications and with better data, it should be possible to link click and purchase behavior, and thus optimize purchases rather than clicks.

In conclusion, our model has demonstrated strong performance advantages over extant methods used in practice. Our methodology leverages features unique to e-mail as a form of direct marketing. These features include compressed-time intervals available for testing, near real-time response measurement, and individual recipient tracking. Beyond e-mail marketing, emerging technologies in marketing that also share these characteristics (e.g., mobile phone communications) could benefit from the testing methodology developed in this paper.

Acknowledgments

References

Ainslie, A., X. Drèze, F. Zufryden. 2005. Modeling movie life cycles and market share. *Marketing Sci.* **24**(3) 508–517.

Bitran, G. D., S. V. Mondschein. 1996. Mailing decisions in the catalog sales industry. *Management Sci.* **42**(9) 1364–1381.

Blair, M. H. 1988. An empirical investigation of advertising wearin and wearout. *J. Advertising Res.* **28**(6) 45–50.

Bucklin, R. E., C. Sismeiro. 2003. A model of web site browsing behavior estimated on clickstream data. *J. Marketing Res.* **40**(33) 249–267.

Bult, J. R., T. Wansbeek. 1995. Optimal selection for direct mail. *Marketing Sci.* **14**(4) 378–394.

Calabria, R., G. Pulcini. 1996. Point estimation under asymmetric loss functions for left-truncated exponential samples. *Comm. Statist. Theory Methods* **25**(3) 585–600.

Danaher, P. J., B. G. S. Hardie. 2005. Bacon with your eggs? Applications of a new bivariate beta-binomial distribution. *Amer. Statistician* **59**(4) 282–286.

Direct Marketing Association. 2005. The DMA 2005 response rate report. Report, Direct Marketing Association, New York.

Drèze, X., F.-X. Hussherr. 2003. Internet advertising: Is anybody watching? *J. Interactive Marketing* **17**(4) 8–23.

Drèze, X., F. Zufryden. 1998. A Web-based methodology for product design evaluation and optimization. *J. Oper. Res. Soc.* **49**(10) 1034–1043.

Elsner, R., M. Krafft, A. Huchzermeier. 2004. Optimizing Rhenania's mail-order business through dynamic multilevel modeling (DMLM) in a multicatalog-brand environment. *Marketing Sci.* **23**(2) 192–206.

Gönül, F., F. Ter Hofstede. 2006. How to compute optimal catalog mailing decisions. *Marketing Sci.* **25**(1) 65–74.

Gönül, F., M. Z. Shi. 1998. Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models. *Management Sci.* **44**(9) 1249–1262.

Gönül, F., B.-D. Kim, M. Z. Shi. 2000. Mailing smarter to catalog customers. *J. Interactive Marketing* **14**(2) 2–16.

Hanson, W. A. 2000. *Principles of Internet Marketing*. South-Western College Publishing, Cincinnati.

Hughes, A. M. 2006. *Strategic Database Marketing*, 3rd ed. McGraw-Hill, New York.

Jain, D. C., N. J. Vilcassim. 1991. Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Sci.* **10**(1) 1–23.

Johnson, N. L., S. Kotz. 1972. *Distributions in Statistics*: *Continuous Multivariate Distributions*. John Wiley & Sons, New York.

Kalbfleisch, J. D., R. L. Prentice. 1985. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York.

Kamakura, W. A., B. S. Kossar, M. Wedel. 2004. Identifying innovators for the cross-selling of new products. *Management Sci.* **50**(8) 1120–1133.

Lee, M.-L. T. 1996. Properties and applications of the Sarmanov family of bivariate distributions. *Comm. Statist. Theory Methods* **25**(6) 1207–1222. http://www.informaworld.com/smpp/content~content=a780019996~db=all.

Lodish, L. M., M. Abraham, S. Kalmenson, J. Livelsberger, B. Lubetkin, B. Richardson, M. E. Stevens. 1995. How T.V. advertising works: A meta-analysis of 389 real world split cable T.V. advertising experiments. *J. Marketing Res.* **32**(2) 125–139.

Moe, W. W., P. S. Fader. 2002. Using advance purchase orders to forecast new product sales. *Marketing Sci.* **21**(3) 347–364.

Nash, E. 2000. *Direct Marketing*. McGraw-Hill Eds, New York.

Pandey, B. N. 1997. Testimator of the scale parameter of the exponential distribution using LINEX loss function. *Comm. Statist. Theory Methods* **26**(9) 2191–2202, http://www.informaworld.com/smpp/content~content=a780134236~db=all.

Pauwels, K., D. M. Hanssens. 2007. Performance regimes and marketing policy shifts. *Marketing Sci.* **26**(3) 293–311.

Radas, S., S. M. Shugan. 1998. Seasonal marketing and timing introductions. *J. Marketing Res.* **35**(3) 296–315.

Silk, A. J., G. L. Urban. 1978. Pre-test-market evaluation of new packaged goods: A model and measurement methodology. *J. Marketing Res.* **15**(2) 171–191.

Sinha, R. V., M. Chandrashekaran. 1992. A split hazard model for analyzing the diffusion of innovations. *J. Marketing Res.* **29**(1) 116–127.

Steenburgh, T. J., A. Ainslie, P. H. Engebretson. 2003. Massively categorical variables: Revealing the information in zip codes. *Marketing Sci.* **22**(1) 40–57.

Sun, B. 2006. Technology innovation and implications for customer relationship management. *Marketing Sci.* **25**(6) 594–597.

Urban, G. L., G. M. Katz. 1983. Pre-test-market models: Validation and managerial implications. *J. Marketing Res.* **20**(3) 221–234.

Weible, R., J. Wallace. 2001. The impact of the internet on data collection. P. S. Richardson, ed. *Internet Marketing*: *Readings Online Resources*. McGraw-Hill, Irwin, Boston, 274–281.

Zellner, A. 1986. Bayesian estimation and prediction using asymmetric loss functions. *J. Amer. Statist. Assoc.* **81**(394) 446–451.