SANGKIL MOON, WAGNER A. KAMAKURA, and JOHANNES LEDOLTER*

This study addresses a problem commonly encountered by marketers who attempt to assess the impact of their sales promotions—namely, the lack of data on competitive marketing activity. In most industries, competing firms may have competitive sales data from syndicated services or trade organizations, but they seldom have access to data on competitive promotions at the customer level. Promotion response models in the literature either have ignored competitive promotions, focusing instead on the focal firm's promotions and sales response, or have considered the ideal situation in which the analyst has access to full information about each firm's sales and promotion activity. The authors propose a random coefficients hidden Markov promotion response model, which takes the competitor's unobserved promotion level as a latent variable driven by a Markov process to be estimated simultaneously with the promotion response model. This enables the authors to estimate cross-promotion effects by imputing the level of competitive promotions. The authors test the proposed model on synthetic data through a Monte Carlo experiment. Then, they apply and test the model to actual prescription and sampling data from two main competing pharmaceutical firms in the same therapeutic category. The two tests show that compared with several benchmark models, the proposed random coefficients hidden Markov model successfully imputes unobserved competitive promotions and, accordingly, reduces biases in the own- and cross-promotion parameters. Furthermore, the proposed model provides better predictive validity than the benchmark models.

Estimating Promotion Response When Competitive Promotions Are Unobservable

Researchers in marketing have devoted considerable attention to the estimation of promotion response models to assess the effects of promotional efforts by a brand on its own and competitors' sales (for a comprehensive review, see Hanssens, Parsons, and Schultz 2001). A vast literature already exists in areas such as the effects of advertising (Dekimpe and Hanssens 1995; Lodish et al. 1995), personal selling (Parsons and Abeele 1981), sales promotion (Blattberg and Wisniewski 1989), and price discounts (Farris and Quelch 1987) on brand sales at the national and market levels. Most studies on the development and application of promotion response models in marketing focus on packaged goods, for which data on sales and marketing efforts by the leading brands competing in each product category are often available from syndicated sources. This wealth of data enables researchers to develop elegant and sophisticated models of brand competition that provide estimates of own- and cross-elasticities. Under this situation, the brand manager can draw inferences about the impact of marketing efforts on his or her own brand and competitor brands and, furthermore, about his or her vulnerability to competitive efforts (Cooper 1988). Disaggregate analyses using single-source data provide an even clearer picture of cross-competitive effects, enabling the manager to tailor his or her efforts to each individual household on the basis of individual-level estimates of own- and cross-elasticities.

*Sangkil Moon is Assistant Professor of Marketing, College of Management, North Carolina State University (e-mail: smoon2@ncsu.edu). Wagner A. Kamakura is Ford Motor Co. Professor of Global Marketing, Fuqua School of Business, Duke University (e-mail: kamakura@duke.edu). Johannes Ledolter is Maxwell C. Stanley Professor of Management Sciences, Henry B. Tippie College of Business, University of Iowa, and Professor of Statistics, Vienna University of Economics and Business Administration (e-mail: johannes-ledolter@uiowa.edu). The authors thank the two anonymous *JMR* reviewers. They also thank Gary Russell for feedback. Russell Winer served as interim and guest editor for this article.

To read and contribute to reader and author dialogue on JMR*, visit http://www.marketingpower.com/jmrblog.*

ISSN: 0022-2437 (print), 1547-7193 (electronic)

Journal of Marketing Research
Vol. XLIV (August 2007), 503–515

These recent developments in promotion response modeling are of great value to the packaged goods industry and to a few other industries in which full information is available on the sales and promotional efforts by the major competing brands at the market or national level. However, these response models are of limited value to industries in which each firm has access to its competitors' sales data through syndicated sources but cannot obtain information on competitors' promotional efforts. For example, a manufacturer of consumer electronics may have monthly sales data for all competing brands at each regional market from syndicated sources, such as GfK, but may not know about its competitors' slot allowances or other promotions. This is a common situation for most manufacturers that sell their products through distributors or retailers. As another example, a manufacturer of pharmaceuticals may have prescription data for all competing brands at the physician level from syndicated sources, such as NDC Health or IMS Health, but may know only which physicians were visited (or detailed, as these visits are named by the industry) and/or received free samples from its own sales force. (Some syndicated sources provide competitive data on detailing and sampling but only for a small sample and typically on a rotating panel.) This is also a typical situation for any firm trying to assess the effectiveness of its salespeople when it might have data on clients' purchases in the product category but not on their exposure to competitors' salespeople. Furthermore, there is a growing emphasis on customer relationship management, in which marketing effort is customized at the customer level. In most customer relationship management implementations, the firm has a wealth of information regarding its own contacts with individual customers and customers' responses to these contacts but a dearth of data on the contacts these customers might have had with its competitors. In other words, many firms do not have access to the data on their competitors' marketing efforts that would make it possible for them to apply a wide variety of the promotion response models found in the marketing literature. A solution (the most commonly used) would be to ignore competitive efforts and estimate a model using only the firm's own marketing efforts. However, it is well known that ignoring a predictor leads to biased estimates of the response parameters when the missing predictor is correlated with the observed one (Goldberger 1991, pp. 189–90). Therefore, this simple and common solution not only prevents the brand manager from having estimates of promotion cross-elasticities but also leads to biased estimates of his or her own promotion elasticities.

The main purpose of this study is to propose and test an approach aimed at resolving the problem of unobserved data on competitors' promotional efforts. We propose a hidden Markov model (HMM) that can effectively estimate the impact of both own- and cross-promotional efforts, even when competitive promotion data are not available. The model simultaneously imputes the missing competitive promotional efforts and estimates the own- and cross-promotional effects on sales. The Markov process of the unobserved competitive promotions enables us to estimate cross-promotional effects through our imputation of the competitive promotions. Our goal in this study is to introduce this idea of estimating a full cross-effects promotion response model while imputing the competitive promotions. It achieves the ostensibly simple but inherently difficult goal of estimating promotion cross-effects for a firm that has only sales data on its competitors in a market. In the following section, we introduce the HMM. We then extend the model into a random coefficients HMM that estimates both unobserved heterogeneity and cross-promotional effects at the individual client level. We compare these models with alternative formulations on synthetic data and then follow up with a real-world data application that estimates the effects of promotion on physician prescription behavior. Finally, we discuss the findings and suggest directions for further research.

AN HMM OF RESPONSE TO UNOBSERVABLE COMPETITIVE PROMOTIONS

Our goal is to estimate how individual clients (e.g., physicians) respond to marketing communications when a brand manager knows his or her own marketing behavior—that is, whether each client is exposed to the brand's own promotions. However, the manager usually has no information on whether or when clients are exposed to competitive promotions. This is a common situation encountered in various industries, including the pharmaceutical industry. In the pharmaceutical industry, a firm can obtain sales data for each physician (i.e., client) from syndicated services (e.g., NDC or IMS), but the firm knows only when its own salespeople visited each physician. For the purpose of clear illustration, we consider the case of a firm with only one major competitor in a duopoly market. This model can be extended to a market in which there is any number of competitors. In such a general case, we can infer the general competitive promotion level rather than the competitor's promotion behavior in a duopoly market. We discuss this generalization subsequently.

For each client n, we want to estimate the following system of response functions in a duopoly market:

(1a) $$y_{nat} = \mu_{na} + \beta_{na} z_{nt} + \delta_{na} x_{nt} + e_{nat}, \text{ and}$$

(1b) $$y_{nbt} = \mu_{nb} + \beta_{nb} z_{nt} + \delta_{nb} x_{nt} + e_{nbt},$$

where

y_{ndt} = sales by client n for brand d (where a denotes the focal brand and b denotes the competitor brand) in month t,

z_{nt} = a variable that indicates whether client n is exposed to promotion by the focal firm in month t ($z_{nt} = 1$) or not ($z_{nt} = 0$),

x_{nt} = a variable that indicates whether client n is exposed to promotion by the competitor in month t ($x_{nt} = 1$) or not ($x_{nt} = 0$),

μ, β, δ = the response parameters to be estimated,

$e_{nat} \sim N(0, \sigma^2_{na})$ and $e_{nbt} \sim N(0, \sigma^2_{nb})$ are regression residuals, and

σ_{nab} = the covariance of the two regression residuals, e_{na} and e_{nb}.

If y, z, and x were all observed, the estimation of the system in Equation 1 would be straightforward as a system of seemingly unrelated regressions (SURs). However, the focal firm, a, does not observe the promotion efforts by the competitor at the client level, and therefore x_{nt} is unavailable. If the focal firm ignores this competitive promotion, it will not be able to assess how the competitor affects its own sales. Moreover, ignoring competitive promotion leads to a biased assessment of its own promotion effects (β_{na}) if there is a correlation between its promotion (z_n) and the competitor's unobserved promotion (x_n).

We treat the unobserved competitive promotions x_{nt} as missing data to be inputed at the same time when we estimate the SUR system of response functions in Equation 1. For this purpose, we assume that the unobserved competitive promotions (x_{nt}) follow a two-state, first-order Markov-switching process with initial-state probabilities $\Pi_n = (\pi_{n0}, \pi_{n1})$ and the following transition probabilities:

$$(2)\qquad A_n = \begin{bmatrix} q_n & 1 - q_n \\ 1 - p_n & p_n \end{bmatrix},$$

where

$$q_n = a_{n00} = \Pr[x_{nt} = 0 | x_{n,t-1} = 0] \text{ and}$$

$$p_n = a_{n11} = \Pr[x_{nt} = 1 | x_{n,t-1} = 1].$$

We assume that the competitive promotion (x_{nt}) follows a Markov process because the manager's current promotion decision is likely to be influenced by the state of promotion at the previous time.

We present the promotion response model in Equations 1 and 2 as a two-state (promotion versus no promotion) HMM in a duopoly market to simplify our exposition and to match our empirical application, which we present subsequently. However, this model can be generalized to an S-state HMM in a C-competitor market without extensive work. We accomplish the state extension by extending Equation 2 to the S-state case (Du and Kamakura 2006). For example, the promotion states might be expanded to "no promotion," "low promotion," "medium promotion," and "high promotion," increasing the number of transition probability parameters significantly. The competitor extension is fulfilled by simply redefining the hidden states of the competitive promotion (x_{nt}). Specifically, we can define the state of the competitive promotion at a given time as the general promotion level of all the competitor brands in the same industry. For example, if the focal firm, A, has two main competitors (B and C), our regression system in Equation 1 would expand to three sales equations, and the hidden states of competitive promotion would include "no competitive promotion," "only B's promotion," "only C's promotion," and "both B's and C's promotions." Adding more promotion levels and more competitors would result in a much larger number of states in the HMM and considerably heavier data requirements.

Hidden Markov models, such as the promotion response model we propose here, have their origin in electroacoustics, particularly for the purpose of speech recognition (for an introduction, see Rabiner and Juang 1986; for a review of this literature, see Juang and Rabiner 1991). Conversely, their applications in marketing are more recent. For example, we refer to the work of Brangule-Vlasgma, Pieters, and Wedel (2002), who use the framework for dynamic value segmentation; Liechty, Pieters, and Wedel (2003), who identify instances of local versus global visual attention when readers are exposed to print advertisements; Montgomery and colleagues (2004), who use a hidden Markov process to identify unobservable goals that drive Web-browsing behavior; and Du and Kamakura (2006), who define household life cycles as latent states in a multistate hidden Markov process. Traditionally, HMMs have been estimated by maximum likelihood using an expected maximization (EM) algorithm (Rabiner 1989). More recently, Bayesian estimation procedures have been developed (Kim and Nelson 1999), which can efficiently lead to a full integration of unobserved heterogeneity. In the following section, we describe how our promotion response HMM can be estimated with the Markov chain Monte Carlo (MCMC) method.

ESTIMATING THE HMM

Estimating the HMM with the EM Algorithm

Given that we have a time series of observed monthly sales for both competing brands and the promotion indicator for the focal brand at the client level, we have enough data to estimate the HMM in Equations 1 and 2. The main purpose of our HMM is to estimate the transition probabilities in Equation 2 for the hidden Markov process while estimating the parameters of the promotion response functions in Equation 1. The likelihood function for one client is given by

$$(3)\ P\big[(e_0, e_1, ..., e_T); A, B, \Pi\big] = \sum_{x_0, x_1, ..., x_T} P\big[(e_0, e_1, ..., e_T) | (x_0, x_1, ..., x_T); A, B, \Pi\big] P\big[(x_0, x_1, ..., x_T); A, B, \Pi\big],$$

where $(e_0, e_1, e_2, \ldots, e_T)$ is the vector of residuals from the system of promotion response functions in Equation 1; $(x_0, x_1, x_2, \ldots, x_T)$ is the vector of missing competitive promotions to be imputed by the hidden Markov process; and A contains Markov transition probabilities, Π contains the initial-state probabilities, and B contains the promotion response parameters.

Note that the likelihood function in Equation 3 depends on a sequence of unobserved events—that is, the competitor promotion. To compute the likelihood, we make use of the first-order Markov process assumed for this missing variable,

$$(4)\quad P[(e_0, e_1, ..., e_T); A, B, \Pi] = \sum_X [b_{x_0}(e_0) b_{x_1}(e_1) \ldots b_{x_T}(e_T)] [\pi_{x_0} a_{x_0 x_1} a_{x_1 x_2} \ldots a_{x_{T-1} x_T}],$$

where $b_{x_t}(e_t) = f[e_t | x_t, B]$. We estimate A, B, and Π with the Baum–Welch EM algorithm (for details about this algorithm, see Baum 1972; Rabiner 1989; Rabiner and Juang 1986). Finally, the Baum–Welch algorithm is guaranteed to converge only to local optima. To find better optima, Kirkpatrick, Gelatt, and Vecchi (1983) developed simulated annealing, which is essentially an MCMC that can increase the performance, especially in high dimensional spaces (Miklos and Meyer 2005).

Estimating the Random Coefficients HMM

We now extend the HMM to allow for unobserved heterogeneity in the response coefficients. The individual-level model does not make the most efficient use of the available data when there are a limited number of clients. Furthermore, the problem of limited data for each individual client can become aggravated with HMM because the process requires that we estimate the state of each time (i.e., monthly promotion state). To overcome this data limitation, we adopt a random coefficients extension of the proposed model in Equations 1 and 2, in which we simultaneously estimate the population distribution of the parameters of the promotion response functions and the individual-level shrinkage estimates from a hierarchical Bayesian model (Allenby and Rossi 1999).

Specifically, Kim and Nelson (1999) develop a Bayesian estimation procedure to estimate an HMM that involves a single regression equation. In this section, we suppress the subscript n for each client to simplify our notation unless it is necessary for clarification. The Kim–Nelson (KN) estimation algorithm treats both the parameters of the response model and the Markov-switching variable, $X = (x_1, x_2, \ldots, x_T)'$, as random variables, where X is the vector of all the unobserved promotion values over the entire period. Thus, in contrast to the classical approach, such as the Baum–Welch algorithm, inference on X is based on a joint distribution, not a conditional distribution. Albert and Chib (1993) make the Bayesian analysis of Markov-switching models easy to implement by using Gibbs sampling, in which they treat both the model parameters and the unobserved Markov-switching variables as missing data generated from appropriate conditional distributions. When we apply this procedure to estimate our model in Equations 1 and 2, it amounts to dealing with T + K variates at the individual client level, where T is the total number of times and K is the number of parameters in the model, as follows:

$$(5)\quad g(X, \mu_a, \beta_a, \delta_a, \sigma_a, \mu_b, \beta_b, \delta_b, \sigma_b, \sigma_{ab}, q, p|D)$$
$$= g(X|D)g(q, p|X)g(\mu_a, \beta_a, \delta_a, \sigma_a, \mu_b, \beta_b, \delta_b, \sigma_b, \sigma_{ab}|D, X),$$

where D indicates all the observed data points. In our case, D is a combination of the focal brand promotion (z) and both brands' prescriptions (y_a and y_b). We assume that conditional on X, the transition probabilities (q and p) are independent of both the response model parameters and the data D. Conditional on D and X, regressions are a system of SURs with known dummy variables, D and X. These conditioning features of the model enable us to employ the simulation tool of Gibbs sampling for Bayesian inference. Furthermore, we assume that the collection of regression parameters for client n, $\theta_n = \{\mu_{na}, \beta_{na}, \delta_{na}, \sigma_{na}, \mu_{nb}, \beta_{nb}, \delta_{nb}, \sigma_{nb}, \sigma_{nab}\}$, follows the following multivariate normal distribution (MVN) across all the clients:

$$(6)\quad \theta_n|D, X \sim \text{i.i.d. } MVN(\gamma_p, \Sigma_p),$$

where γ_p and Σ_p are defined as the population means and variances–covariances (Allenby and Rossi 1999; Gelman et al. 1995).

In line with the KN algorithm, we apply three key changes to the basic model that leads to our random coefficients HMM. First, our estimation incorporates a random coefficients component into the KN procedure to pool information across all the clients in the data set. This extension is required to make full use of the observed data. Typically, we can obtain data only on a limited time length for each client. Because we must estimate the promotion state (promotion/no promotion) of each point, model estimation makes extremely high demands on the data. By extending the individual-level KN procedure into a hierarchical model to pool data across clients, we can improve estimation accuracy significantly. Subsequently, we show this point in our tests using synthetic data and actual prescription data alike. Second, our model consists of a system of regressions, whereas the KN procedure considers only a single regression. When the explanatory variables of the two regressions are identical, as is the case in our model, the generalized least squares and ordinary least squares estimators coincide, and each of the two regressions can be estimated separately in our model (Goldberger 1991, p. 327). However, this does not mean that the two regression residuals are independent; the correlation of the two regression residuals has an impact on estimating the hidden state of the competitive promotion, as we show in the Appendix. Third, the original KN procedure generates state-specific regression parameter estimates, whereas our model is simplified to generate state-common estimates for the purpose of parsimony.

To implement the Gibbs-sampling estimator, we need to derive the distributions of the blocks of each of the T + K variates, conditional on all the other blocks of variates at the individual client level. Furthermore, we estimate population regression parameters in a separate step because it is a random coefficients model. Using arbitrary starting values for the model parameters, we can repeat four steps until convergence occurs. This is a significant extension of the KN procedure, which does not include Step 1 (specific details for each step appear in the Appendix).

Step 1: Generate population regression parameters $\varphi_p = \{\mu_{pa}, \beta_{pa}, \delta_{pa}, \sigma_{pa}, \mu_{pb}, \beta_{pb}, \delta_{pb}, \sigma_{pb}, \sigma_{pab}\}$ from $g(\varphi_p|D, X, \theta)$, where θ is a collection of all the individual clients' regression parameters.

Step 2: Generate each x_t from $g(x_t|X_{t-}, \mu_a, \beta_a, \delta_a, \sigma_a, \mu_b, \beta_b, \delta_b, \sigma_b, \sigma_{ab}, q, p, D)$, t = 1, 2, …, T, where $X_{t-} = \{x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_T\}'$ refers to a vector of X variables that excludes x_t.

Step 3: Generate the transition probabilities, q and p, from $g(q, p|X)$.

Step 4: Generate individual clients' regression parameters $\mu_a, \beta_a, \delta_a, \sigma_a, \mu_b, \beta_b, \delta_b, \sigma_b, \sigma_{ab}$ from $g(\mu_a, \beta_a, \delta_a, \sigma_a, \mu_b, \beta_b, \delta_b, \sigma_b, \sigma_{ab}|D, X)$.

TESTING THE PROPOSED PROMOTION MODEL ON SYNTHETIC DATA

Before applying our proposed promotion response model to actual data, we present the results from a Monte Carlo simulation study in which we generate synthetic data based on known parameter values and compare different formulations of the promotion response model in terms of their ability to recover the "true" response parameters and to impute the promotion activity of the competing brand. Furthermore, we compare the predictive performance of the models.

Data Generation

For each client n, we draw "true" transition probabilities q_n and p_n from the uniform distribution of U[.05, .95]. Then, we draw promotion response parameters μ_{na}, β_{na}, and δ_{na} for the first regression in Equation 1 from i.i.d. normal distributions $\phi(\mu = 10, \sigma^2 = 5^2)$, $\phi(5, 2.5^2)$, and $\phi(-3, 2^2)$, respectively, and for the second regression (μ_{nb}, β_{nb}, and δ_{nb}) from i.i.d. normal distributions $\phi(10, 5^2)$, $\phi(-3, 2^2)$, and $\phi(5, 2.5^2)$, respectively. We chose all these parameter values to be close to their counterparts in the prescription data set used in our subsequent empirical analysis. Given these parameters for each client n, we then generate the synthetic promotion and sales time series for 50 periods (i.e., t = month) using the following steps:

1. Using the drawn transition probabilities for client n, q_n and p_n, we draw a series of the binary promotion variable (promotion/no promotion) for the competitor brand, x_{nt}. We draw the initial states for each client from a Bernoulli distribution with a promotion probability of .7. We assume that this series is unobserved in the model estimation, but we use it to generate synthetic data here. Conversely, we generate a series of the binary promotion variable for the focal brand, z_{nt}, from Bernoulli draws with proportions varying for each client from the uniform distribution of [.05, .95]. We assume that the series of the focal firm's promotion variable is observed in the model estimation.
2. We draw the two regression residuals, e_{at} and e_{bt}, from $\phi(0, 2.5^2)$ and $\phi(0, 2.3^2)$, respectively.
3. Using the promotion predictors and residuals for the month and the regression coefficients for the client, we generate the sales of the two brands for the client and the month.
4. We compute correlations between the two promotion variables, x_{nt} and z_{nt}. According to the correlation, we draw 200 clients with a low correlation (.0 ~ .3) and another 200 clients with a high correlation (.3 ~ .6). We use the two client groups to consider situations in which the bias caused by ignoring competing promotions will be low or high. It is known that the correlations of the two predictor variables can affect the magnitude of coefficient estimation biases in a misspecified model (Goldberger 1991, pp. 189–90). The client-specific correlation between the two regression residuals ranged from –.41 to .35 in the synthetic data.

Benchmark Models

To assess the performance of the two versions of our proposed promotion response HMM (i.e., the individual HMM and the random coefficients HMM), we consider the three following benchmark models:

1. *SUR*: This is the system of regressions in Equation 1, which assumes that we know the promotions for both the focal and the competing brands for each client. This is the ultimate benchmark because it uses full information about the competitor's promotion.
2. *Regressions without competitor*: This is the most naive benchmark because it ignores the competing brand's unobservable promotions.
3. *Latent variable regressions*: For this type of regression, we take the competitor's promotion as unobservable, and we impute it as a time-independent latent variable. We estimate this model as a random coefficients model; the only difference between this benchmark model and our proposed random coefficients HMM is that the former does not link the competitor brand's unobserved promotion to prior promotions, an essential part of our HMM. We expect that our random coefficients HMM will be more effective in recovering the unobserved promotion values because it uses additional information from previous unobserved promotions through the hidden Markov process.

Performance Measures

Our proposed models have three main purposes: (1) to estimate own- and cross-promotion effects for both the focal and the competing brand (parameter recovery), (2) to impute the competitor's unobservable promotions (imputation of unobservable promotion), and (3) to predict sales level on the basis of the intended promotion activity by the focal brand and the projected promotion for the competing brand (predictive fit). Therefore, we compare our two proposed HMMs with the three benchmark models on multiple performance criteria and address these three objectives as follows:

1. *Parameter recovery:*

$$\text{MAD} = \sum_{n=1}^{N} |\hat{\theta}_n - \theta_n|/N, \text{ and MAP} = \frac{\sum_{n=1}^{N} \frac{|\hat{\theta}_n - \theta_n|}{|\theta_n|}}{N},$$

where MAD = mean absolute deviation, MAP = mean absolute percentage, and θ_n and $\hat{\theta}_n$ are actual and estimated response coefficients (μ, β, and δ) for client n.

2. *Imputation of unobservable promotion*:

HR = % of correct imputations of the missing promotions, and

$$\text{APP} = \sum_{n=1}^{N} \left[\frac{\sum_{t=1}^{T} \gamma_t(x_{nt}^*)}{T} \right] /N,$$

where HR = hit ratio, APP = average predicted probability, and $\gamma_t(x^*_{nt})$ is the posterior probability of the competitor's promotion to client n at period t.

3. *Predictive fit*:

$$\text{K-step-ahead MSE(K)} = \sum_{n=1}^{N} (y_{nT} - \hat{y}_{n,T-K,T})^2/N, \text{ and}$$

$$\text{K-step-ahead MAPE(K)} = \sum_{n=1}^{N} \left(\frac{|y_{nT} - \hat{y}_{n,T-K,T}|}{y_{nT}} \right) /N,$$

where MSE = mean squared error, MAPE = mean absolute percentage error, and $\hat{y}_{n,T-K,T}$ is the K-step-ahead forecast produced for client n at time T – K for time T based on the observed promotion by the focal firm and the projected promotion by the competitor at time T – K.

Empirical Results from Synthetic Data

We compare the overall performance of the two proposed HMMs (the individual HMM and the random coefficients HMM) and the three benchmarks (SUR, regressions without competitor, and latent variable regressions) on synthetic data in Table 1 in terms of the three main goals (i.e., parameter recovery, imputation of unobservable promotion, and predictive fit).

Because the SUR model uses all the available information, including data on the competitor promotions, it shows the best parameter recovery in terms of the MAD and the MAP deviation from the "true" parameters. The results from the naive model (regression without competitor) show that ignoring competitor promotions not only prevents us from estimating the effect of the competitor's promotion but also substantially increases the biases in the parameter estimates for the focal brand. The random coefficients latent-variable-regressions model enables us to impute the missing competitor promotions and to estimate the effect on the dependent variable. However, although we obtain the individual-level parameters through shrinkage, the bias is as high as with the naive model. Parameter recovery for the individual HMM is fairly similar to that of the latent-variable-regressions model. We obtain better results with the random coefficients HMM, showing the benefits of combining the first-order Markov structure on the unobserved promotions and individual-level estimates of the response coefficients through shrinkage.

Table 1, Panel B, shows that the performance of the random coefficients HMM is also superior to that of the latent-variable-regressions model and the individual HMM in

Table 1

EMPIRICAL RESULTS FROM SYNTHETIC DATA

A: Parameter Recovery on Synthetic Data

	SUR (Using All Data)	*Regressions Without Competitor*	*Latent Variable Regressions*	*Individual HMM*	*Random Coefficients HMM*
Low Correlations Between Observed and Unobserved Promotions					
MAD					
μ_a	.51	1.29	1.21	1.88	.67
β_a	.54	.61	.58	.60	.57
δ_a	.62	—	2.60	3.34	.85
μ_b	.45	2.04	1.17	1.05	.81
β_b	.53	.71	.67	1.09	.60
δ_b	.63	—	4.19	3.20	1.05
MAP					
μ_a	11%	22%	24%	25%	14%
β_a	14%	16%	17%	16%	15%
δ_a	61%	—	105%	189%	74%
μ_b	6%	31%	20%	19%	18%
β_b	29%	44%	40%	59%	33%
δ_b	18%	—	92%	67%	30%
High Correlations Between Observed and Unobserved Promotions					
MAD					
μ_a	.46	1.11	1.09	1.74	.67
β_a	.62	1.15	1.07	1.06	.82
δ_a	.64	—	3.12	3.84	.85
μ_b	.40	1.80	1.77	.99	.58
β_b	.55	1.75	1.58	2.11	1.19
δ_b	.59	—	4.94	3.89	1.43
MAP					
μ_a	8%	21%	20%	24%	12%
β_a	17%	32%	31%	27%	21%
δ_a	35%	—	98%	139%	55%
μ_b	8%	44%	50%	34%	15%
β_b	37%	118%	95%	119%	68%
δ_b	17%	—	94%	76%	39%

B: Imputation of Unobservable Promotion and Predictive Fit on Synthetic Data

	SUR (Using All Data)	*Regressions Without Competitor*	*Latent Variable Regressions*	*Individual HMM*	*Random Coefficients HMM*
Low Correlations Between Observed and Unobserved Promotions					
Imputation					
HR	—	—	59%	59%	79%
APP	—	—	49%	54%	74%
Predictive Fit for Brand A (Focal Brand)					
MSE(1)	7.83	9.70	8.86	9.08	8.23
MAPE(1)	61%	84%	90%	78%	49%
MSE(2)	7.86	9.77	8.75	9.13	8.29
MAPE(2)	63%	85%	76%	77%	64%
Predictive Fit for Brand B (Competitor Brand)					
MSE(1)	6.27	13.24	11.45	10.81	8.99
MAPE(1)	33%	48%	46%	44%	35%
MSE(2)	6.28	13.33	12.13	12.92	10.91
MAPE(2)	33%	48%	46%	46%	40%

Table 1
CONTINUED

	SUR (Using All Data)	Regressions Without Competitor	Latent Variable Regressions	Individual HMM	Random Coefficients HMM
High Correlations Between Observed and Unobserved Promotions					
Imputation					
HR	—	—	56%	59%	83%
APP	—	—	58%	54%	77%
Predictive Fit for Brand A (Focal Brand)					
MSE(1)	7.08	10.54	9.57	9.61	9.19
MAPE(1)	83%	102%	95%	94%	89%
MSE(2)	7.13	10.59	9.67	9.80	9.59
MAPE(2)	83%	102%	93%	93%	92%
Predictive Fit for Brand B (Competitor Brand)					
MSE(1)	5.12	13.62	11.82	12.95	9.17
MAPE(1)	25%	45%	43%	42%	31%
MSE(2)	5.14	13.72	12.02	10.96	10.19
MAPE(2)	25%	45%	42%	39%	36%

terms of its ability to impute the unobservable promotions by the competitor. The individual HMM shows essentially the same performance in imputing the missing data as the random coefficients latent-variable-regressions model, again suggesting that the HMM makes too heavy demands on the data when it is estimated at the individual level. Table 1, Panel B, also shows that in most tested cases, the random coefficients HMM is second only to the full-information model (SUR) in terms of predictive fit.

In Table 1, Panels A and B, we study two groups of simulated clients that are based on the correlation between the unobserved and observed promotion variables, x_{nt} and z_{nt}. Specifically, we divide people into a low-correlation group (.0 ~ .3) and a high-correlation group (.3 ~ .6). This division is to confirm empirically that the biases on the focal brand's promotion variable coefficients (β_a and β_b) are significantly amplified for the high-correlation case when the competitor brand's promotion information is ignored, as is the case in the regression-without-competitor model because of omitted-variable bias (Goldberger 1991, pp. 189–90). A comparison of the MAP measures of β_a and β_b between the two correlation groups confirms this bias. Specifically, for the regression-without-competitor model, the MAP of β_a for the high-correlation group (32%) is twice as large as the same measure for the low-correlation group (16%). The bias increase is even larger for β_b in the high-correlation sample (MAP = 118%) than in the low-correlation sample (MAP = 44%). Such amplified biases in β_a appear to have influenced the predictive fit for the focal brand (Brand A). Again, for the regressions-without-competitor model, the MAPE(1) measures (102%) for the high-correlation physician group is much larger than the same measure (84%) for the low-correlation group. We observe the same pattern for the MAPE(2) measures. More important, in the high-correlation group, our random coefficients HMM can significantly reduce the biases relative to the naive model for both β_a (MAP = 32% to 21%) and β_b (MAP = 118% to 68%). This bias correction results in better predictive performance for our random coefficients HMM as well.

In general, our performance tests are consistent and robust across all the three distinct measures: (1) regression coefficient bias reduction, (2) unobserved promotion imputation, and (3) sales prediction. The general conclusions that can be drawn from this simulation experiment are threefold. First, ignoring competitive promotions can lead to biased estimates of promotion responses, depending on how correlated they are with the focal firm's observed promotions. Second, our random coefficients promotion response HMM is able to impute what the competition is doing in terms of promotion effort and, accordingly, leads to a less biased estimate of promotion response. Third, our proposed model can be used for forecasting future sales of both the focal firm and the competitor firm in the duopoly market. Improved sales forecasting can be used to develop the focal firm's future marketing strategies. In the next section, we test the same models using real-world data from the pharmaceutical industry.

EMPIRICAL ANALYSIS: PHYSICIANS' RESPONSES TO PROMOTION

For pharmaceutical products, detailing (i.e., a visit by a salesperson) and sampling (i.e., delivery of free drug samples) are well known as two of the most effective and widely used forms of promotion. According to IMS Health, detailing is estimated to cost more than $4 billion each year to the pharmaceutical industry. According to some sources (e.g., Mizik and Jacobson 2004), the retail value of free samples distributed during these detailing visits is estimated at $10 billion each year. A single visit by a salesperson, including a sample drop, represents a marketing investment of approximately $250 per visit for a pharmaceutical manufacturer (Kamakura, Kossar, and Wedel 2004); therefore, the deployment of these promotion tools at the physician level is the subject of careful scrutiny by this industry. Fortunately, firms in the pharmaceutical industry have at their disposal comprehensive monthly data on the prescription behavior of each physician in the United States from at least two sources (i.e., NDC and IMS). Furthermore, each pharmaceutical manufacturer has its own promotion information about the physicians it visits each month. However, it does not have access to the competitor brand's promotion behavior at the physician level. At best, the manufacturer might be able to obtain limited data on competitor promotions for a small rotating panel of physicians. Previous studies of promotion response in the pharmaceutical industry reflect these data constraints; specifically, they either use full information (across competing brands) from a small panel of physicians (Gonul et al. 2001; Narayanan, Manchanda, and Chintagunta 2004) or ignore competitive promotions (Manchanda and Chintagunta 2004; Manchanda, Rossi, and Chintagunta 2004; Mizik and Jacobson 2004). The reality that competitors in the pharmaceutical industry face is similar to that faced in many other industries; data might be available on sales revenues for all competing brands but only in limited scope on marketing efforts. We attempt to address this problem with our proposed promotion response HMM.

Data Description

We obtained data from two of the leading competing brands in a particular therapeutic class for which we had both monthly prescription data and information on whether each of 156 physicians received samples of the two competing drugs during a 24-month period (see Table 2). This provides us with a rare opportunity to test our proposed models with real market data and to validate the models' predictive validities. Ideally, we want to fit a model that incorporates both detailing and sampling as promotional tools. However, we have no access to detailing data for the competing brand, which prevents us from testing the imputation of unobservable detailing visits. Furthermore, because we have partial data on the focal firm's detailing, we know that the correlation between sampling and detailing is .72, which shows a strong collinearity between the two promotional tools over time. Because of this strong collinearity and because we do not have any data on detailing for the competing brand, we used sampling in model testing because it is our best available proxy measure; even if we had data on detailing for both brands, the strong correlation between detailing and sampling would present the problem of severe multicollinearity when we included both sampling and detailing in the response model (Greene 1997, pp. 418–27).

In summary, with our prescription data, the unobserved variable (x_{nt}) in Equation 1 indicates the presence/absence of the competitor's promotion activity, including both detailing and sampling, which is common in the pharmaceutical industry. We use only the competitor brand's sampling data to test a model's performance in terms of unobserved state imputation because it is the only available information to us regarding the competitor's promotion behavior. We present summary statistics for the data used in our empirical application in Table 2.

Inferring the Effects of Unobservable Sampling Promotions

Our testing of the two proposed promotion response HMMs (individual HMM and random coefficients HMM) on actual physician prescription behavior proceeded in a similar way to the procedure applied to the synthetic data, except for parameter recovery. With this actual prescription data set application, we do not have any information about the "true" parameter values. Therefore, our parameter recovery performance criteria now use the estimates from the SUR model, which makes use of all the available data, including the competitive promotions as the proxy of the "true" parameters. The results from our model comparisons appear in Table 3, Panels A and B.

As in our tests that use synthetic data, we attained the best parameter recovery (compared with the SUR model in this application) with the random coefficients HMM. The biases still seem substantial in terms of the MAP measures. Notably, in terms of the MAD measure, the biases do not appear to be substantial. The large numbers of the MAP measures occur partially because of the small coefficient values in the SUR estimates that are used as a benchmark model; because these estimates occupy the denominator in the MAP formula, small increases in the bias resulted in dramatic increases in the MAP, a typical distortion in such a relative measure. Our random coefficients HMM also produces the best imputations of the missing promotions with an HR of 76%. Our model also generates predictive fits that are second only to the full-information SUR model, which obviously has the benefit of knowing competitor promotions. In general, performance in terms of predictive fit improves as we move from the most naive model (regression without competitor) to the most advanced model (random coefficients HMM).

Figure 1 shows the estimates of promotion response of the focal brand to its own promotions (β_a) and the unobservable competitor promotions (δ_a) we obtained with our

Table 2
SUMMARY STATISTICS OF THE PRESCRIPTION DATA

Brand	*Variable*	*Data Description*
Focal brand (Brand A)	Prescription	M = 5.07 SD = 6.44
	Sampling	Binary data (sampling/no sampling) Proportion 53.6%
	Detailing	(The number of monthly visits per physicians) M = 1.20 SD = 1.13
Competitor brand (Brand B)	Prescription	M = 2.86 SD = 3.70
	Sampling	Binary data (sampling/no sampling) Proportion 70.8%
	Detailing	Data not available

Notes: The actual brand names cannot be released because of confidentiality request of the data provider.

Table 3
EMPIRICAL RESULTS FROM ACTUAL PRESCRIPTION DATA

A: Parameter Recovery on Actual Physician Prescription Data

	Regressions Without Competitor	*Latent Variable Regressions*	*Individual HMM*	*Random Coefficients HMM*
MAD				
μ_a	1.36	1.24	2.09	1.01
β_a	.34	.52	.58	.30
δ_a	—	1.75	1.78	1.46
μ_b	2.08	2.09	.77	.65
β_b	.58	.66	.84	.46
δ_b	—	2.61	2.83	.85
MAP				
μ_a	49%	47%	98%	36%
β_a	47%	58%	61%	41%
δ_a	—	257%	278%	193%
μ_b	891%	954%	349%	234%
β_b	180%	201%	295%	145%
δ_b	—	87%	101%	38%

B: Imputation of Unobservable Promotion and Predictive Fit on Actual Physician Prescription Data

	SUR (Using All Data)	*Regressions Without Competitor*	*Latent Variable Regressions*	*Individual HMM*	*Random Coefficients HMM*
Imputation					
HR	—	—	61%	63%	76%
APP	—	—	51%	54%	72%
Predictive Fit for Brand A (Focal Brand)					
MSE(1)	11.79	23.65	26.34	18.95	14.01
MAPE(1)	63%	96%	98%	85%	71%
MSE(2)	14.35	25.59	26.35	20.02	16.04
MAPE(2)	69%	100%	98%	87%	78%
Predictive Fit for Brand B (Competitor Brand)					
MSE(1)	4.25	9.09	7.32	7.80	7.31
MAPE(1)	53%	105%	91%	93%	091%
MSE(2)	5.24	13.06	13.23	9.72	8.89
MAPE(2)	60%	126%	124%	107%	103%

Notes: The results in Panel A use the SUR estimates as the proxy of the true parameter values.

proposed model and the full-information SUR model. On the one hand, this scatterplot shows a close recovery of the own-promotion effects. On the other hand, there are some discrepancies in the response coefficients for the unobservable promotion (δ_a). We believe that these discrepancies are caused by the imperfect estimation of the hidden-state variable values by the HMM (HR = 76%).

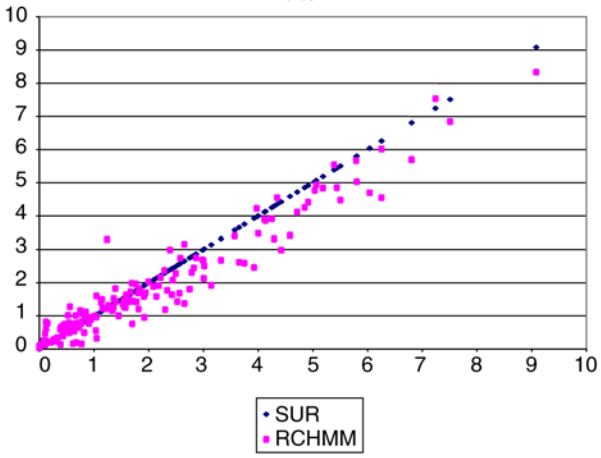
In Table 4, we provide a summary of the individual-level estimates we obtained with our random coefficients HMM, as well as those we obtained with the current practice of ignoring unobservable competitive promotion (regressions without competitor) and the benchmark estimates from the full-information SUR model. In general, this summary confirms what we found in our model comparisons in Table 3, but the differences are seemingly less accentuated. The most dramatic differences between the three models are found in the two intercepts (μ_a and μ_b). Regarding this issue, we present the following two points: First, ignoring the competitor's sampling leads to the underestimation of the focal brand's intercept (μ_a) because of the overestimation of its own promotion effect (β_a). Second, ignoring the competitor's sampling leads to the overestimation of the competitor's intercept (μ_b) because of the overestimation of the focal firm's promotion (β_b). Thus, our point is confirmed again that ignoring the competitor's sampling leads to an incomplete and inaccurate view of the focal firm's sales response. In contrast, our proposed HMM shows how each physician responds to the focal firm's promotion depending on the competitor's promotion activity.

In Table 5, we provide a summary of the expected prescriptions for the focal firm across all physicians, conditional on the two competitors' sampling activities. As would be expected, our HMM (as well as the full-information model) predicts that the focal firm's sampling will have a more positive impact on its own prescriptions if the competitor does not distribute a sample at the same time. Furthermore, the expected prescriptions for the focal firm are the lowest when the competitor brand promotes and the focal brand does not promote. In contrast, the expected prescriptions for the focal firm are the highest in the opposite case in which the competitor brand does not promote and the focal brand promotes. We cannot analyze these two-by-two scenarios with the regressions-without-competitor model, because it ignores competitive sampling. This naive model simply predicts an average of 6.6 prescriptions if the focal brand promotes versus 4.5 prescriptions if it does not promote. Our proposed model produces predictions that are closer to the full-information benchmark model (i.e., SUR).

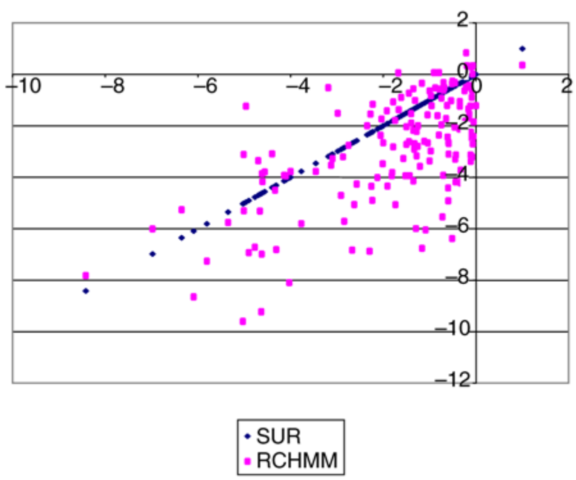
In conclusion, the results with actual prescription and sampling data mirror those we obtained with synthetic data

Figure 1
INDIVIDUAL PHYSICIANS' ESTIMATED PROMOTION RESPONSIVENESS

β_a



δ_a



Notes: RCHMM = random coefficients HMM. The SUR points constitute a 45-degree line as a model with complete information, and the RCHMM demonstrates the best estimates among the four compared models in Table 3, Panel A.

in the previous section; our random coefficients promotion response HMM leads to less biased estimates of promotion response than both the regressions-without-competitor model, which ignores competitive promotions, and the latent-variable-regressions model, which treats unobservable promotions as a latent variable without considering the Markovian nature of the promotion changes over time. This is possible because the random coefficients HMM is better at imputing unobservable competitive promotions than the other competing models. Finally, the structure of our HMM allows us to obtain better predictions for future prescription sales of both the focal and the competitor brands.

DISCUSSION AND DIRECTIONS FOR FURTHER RESEARCH

In this study, we attempted to address a common problem faced by marketers who need to assess the impact of their promotion efforts when they do not have access to data on their competitors' promotion behavior. In some industries (e.g., packaged goods), competitive promotions are available for a sample of stores or customers, and therefore a full cross-effect promotion response model can be estimated from the sample. Conversely, for most other industries, such data are not available, and the analyst has no option other than ignoring competitive promotions. We propose an alternative solution to this problem: Treat the unobserved promotions as missing data to be imputed by a hidden Markov process, and estimate a full own- and cross-effect promotion response model. We believe that our proposed approach addresses a situation that is more typical in the marketplace than the one depicted in most promotion response models developed for the packaged goods industry with complete competitive promotion information. The empirical results, which are based on both synthetic and actual data, suggest that the proposed model provides less biased estimates of promotion response and allows the manager to estimate promotion cross-effects on sales. Given that promotions in many industries do not generate primary demand but are more likely to shift market shares, models that ignore promotion cross-effects would be of less value as a tool for strategic and tactical deployment of marketing resources.

The main purpose of our study was to introduce the idea of estimating a full cross-effect promotion response model while imputing the unobserved competitive promotion. To this end, we tested it on actual market data that enabled us to compare the performance of our proposed random coefficients hidden Markov approach with results obtained with full information about the competition. For this reason and for the purpose of simple exposition, we considered only a relatively simple duopoly market with two competing brands that matched the actual data we had at our disposal. As we explained previously, however, our model can be generalized to an S-state HMM in a C-competitor market.

Table 4
SUMMARY OF THE PARAMETER ESTIMATES ACROSS PHYSICIANS

	SUR (Using All Data)		*Regressions Without Competitor*		*Random Coefficients HMM*	
Parameter	*M*	*SD*	*M*	*SD*	*M*	*SD*
μ_a	5.83	5.86	4.53	4.99	6.68	5.13
β_a	2.04	1.82	2.23	1.96	1.87	1.63
δ_a	–1.78	1.73	—	—	–2.91	2.16
μ_b	1.04	1.26	3.11	2.05	1.73	1.27
β_b	–1.00	1.23	–1.24	1.61	–.88	1.33
δ_b	2.82	1.33	—	—	2.61	1.51

Table 5
EXPECTED PRESCRIPTIONS OF THE FOCAL BRAND ACROSS ALL PHYSICIANS BY THE SAMPLING CONDITION

SUR (Using All Data)			
		Competitor Brand (Brand B)	
		No Promotion	*Promotion*
Focal brand (Brand A)	No promotion	5.4 (6.3)	4.1 (4.9)
	Promotion	7.7 (6.6)	6.1 (5.4)
Random Coefficients HMM			
		Competitor Brand (Brand B)	
		No Promotion	*Promotion*
Focal brand (Brand A)	No promotion	4.6 (5.0)	4.3 (5.1)
	Promotion	6.9 (6.4)	6.3 (5.6)
Regressions Without Competitor			
Focal brand (Brand A)	No promotion	4.5 (5.0)	
	Promotion	6.6 (5.7)	

Notes: Numbers outside the parentheses are means, and numbers inside the parentheses are standard deviations.

For example, separating out each competitor promotion behavior in the multiple-competitor market would require specification of a model structure that is different from our SUR structure in Equations 1 and 2. Suppose that the focal firm has two competitors (B and C). Our system of two regressions in Equation 1 would then be extended to three equations, and the HMM representing the promotional "state" of the two competitors would include four possible states: no competitive promotion, both competitors' simultaneous promotion, only B's promotion, and only C's promotion. We leave such a context for further research.

When developing our HMM, we considered the unobserved competitor promotion as a discrete event. As in previous applications of HMM in marketing, we could accommodate only discrete "hidden" states because of the discrete nature of the model we used. A possible extension of the proposed promotion response model for "continuous" unobserved promotions may arise from the continuous HMM (Digalakis, Monaco, and Murveit 1996).

Regarding our empirical data from the pharmaceutical industry, some readers could argue that a Poisson regression would be more appropriate than our normal regression approach because we used the number of prescriptions written each month as our dependent variable (Manchanda and Chintagunta 2004; Manchanda, Rossi, and Chintagunta 2004). Despite the merit of such an argument, we decided to maintain the current normal linear regression component for three reasons. First, as Equation 1 suggests, the normal regression can deal with a more general case in which sales is the dependent variable, whereas the Poisson regression is limited to count data. Because we have count data of a large magnitude, we can use a normal regression because a Poisson distribution in such a case will approximate a normal distribution. Second, a typical Poisson regression might be inappropriate to our case for two reasons: overdispersion and excessive zero observations in both brands' prescriptions. Whereas there is a relatively simple remedy for overdispersion, with the negative binomial distribution, there is no straightforward solution for the excessive zero observations problem in association with the structure of the HMM. Mullahey (1986) suggests a hurdle model as a solution to the zero observations for the Poisson model, in which the zero outcome from the data-generating process is considered different from the positive ones (regime-splitting mechanism; Greene 1997). The combination of this regime-splitting mechanism and the HMM mechanism will complicate the estimation procedure substantially. Third, we determined that our normal model was empirically robust enough for the prescription data, given the observed range of monthly prescriptions (up to 65 and 66 for the focal brand and the competitor's brand, respectively). This argument is empirically supported by our Wald test for the regression residual normality test (Greene 1997). Specifically, the regression from Equation 1a passed the residual normality test for 83% of the 156 physicians in the data, and the regression from Equation 1b passed the same test for 63% of the same physicians.

In our model development and empirical illustration, we assumed that competitor promotions were unobservable to the focal firm, thus precluding the possibility that the two competing firms reacted to each other's promotions. Another potentially worthwhile extension of our promotion response model would be to assume that the hidden states (i.e., competitive promotions) depend on the previous promotions by both the competing brand and the focal brand rather than only the competing brand, as we assumed in this research. In other words, the assumption would be that the focal firm can somehow perceive the competitive promotions, even though they are not directly observable to the analyst. Such an extension of our proposed model would be possible by making the transition probabilities in Equation 2 a function of the previously observed promotion by the focal brand. This would enable managers not only to measure the effect of their promotions on sales of all brands but also to understand better how competitors react to their own promotional efforts. We leave this extension for further research.

APPENDIX: THE MCMC ESTIMATION OF THE RANDOM COEFFICIENTS HMM

We provide specific details of each step of the MCMC estimation procedure for the random coefficients HMM outlined in the "Estimating the Random Coefficients HMM" subsection.

Step 1: Generate population regression parameters $\phi_p = \{\mu_{pa}, \beta_{pa}, \delta_{pa}, \sigma_{pa}, \mu_{pb}, \beta_{pb}, \delta_{pb}, \sigma_{pb}\}$ from $g(\phi_p|D, X, \theta)$.

The hyperpriors of the population regression parameters are set up as follows:

$$\gamma_l|\gamma_0, \Sigma_0 \sim MVN(\gamma_0, \Sigma_0), \text{ and } \Sigma_l \sim IW_{v0}(\Lambda_0^{-1}),$$

where IW indicates the inverse Wishart distribution. In our analysis, specifically, $\gamma_0 = [0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]'$, $\Sigma_0 = 1000 \times I_8$, $\nu_0 = 100$, and $\Lambda_0 = 100 \times I_8$. We combine this with the normal likelihood $MVN(\gamma_l, \Sigma_l)$, where γ_l and Σ_l are the means and variances–covariances of all the posterior regression parameters of all the individual clients from Step 4. Then, we can obtain the posteriors from the following MVN (Allenby and Rossi 1999; Gelman et al. 1995):

$$\varphi_p|D, X, \theta \sim MVN(\gamma_p, \Sigma_p),$$

where

$$\gamma_p = \frac{\Sigma_0^{-1}\gamma_0 + \Sigma_l^{-1}\gamma_l}{\Sigma_0^{-1} + \Sigma_l^{-1}} \text{ and}$$

$$\Sigma_p = (\Sigma_0^{-1} + \Sigma_l^{-1})^{-1}.$$

In the remainder of the Appendix, we suppress the subscript n for each client to avoid cluttering notation.

Step 2: Generate each x_t from $g(x_t|X_{t-}, \mu_a, \beta_a, \delta_a, \sigma_a, \mu_b, \beta_b, \delta_b, \sigma_b, \sigma_{ab}, q, p, D)$.

Single-move Gibbs sampling, originally motivated by the work of Albert and Chib (1993), dictates simulating x_t one by one from each of the following T conditional distributions (Kim and Nelson 1999):

$$g(x_t|X_{t-}, \mu_a, \beta_a, \delta_a, \sigma_a, \mu_b, \beta_b, \delta_b, \sigma_b, \sigma_{ab}, q, p, D), t = 1, 2, \ldots, T.$$

When we suppress the conditioning parameters,

$$g(x_t|X_{t-}, D) \propto g(y_{at}, y_{bt}|x_t)g(x_t|x_{t-1})g(x_{t+1}|x_t), t = 1, 2, 3, \ldots, T,$$

where

$$g(y_{at}, y_{bt}|x_t) = BVN(\rho_a, \rho_b, \sigma_a^2, \sigma_b^2, \sigma_{ab}),$$

where BVN denotes the bivariate normal distribution; ρ_a and ρ_b indicate the means of the two regressions, respectively; and $g(x_t|x_{t-1})$ and $g(x_{t+1}|x_t)$ are given by the transition probabilities. Using this conditional distribution,

$$Pr(x_t = j|X_{t-}, D) = \frac{g(x_t = j|X_{t-}, D)}{\sum_{i=0}^{1} g(x_t = i|X_{t-}, D)}.$$

After $Pr(x_t = j|X_{t-}, D)$ is calculated, it is straightforward to generate x_t using a uniform distribution. Specifically, a random number from U[0, 1] is generated. If the generated number is less than or equal to the calculated value of $Pr(x_t = j|X_{t-}, D)$, x_t is set to be 1 (promotion). Otherwise, x_t is set to be 0 (no promotion).

Step 3: Generate the transition probabilities, q and p, from $g(q, p|X)$.

Conditional on X, q and p are independent of the data set D and the model's other parameters. We use two independent beta distributions as conjugate priors for the transition probabilities q and p, respectively:

$$q \sim beta(u_{00}, u_{01}), \text{ and}$$

$$p \sim beta(u_{11}, u_{10}),$$

and $g(q, p) \propto q^{u00-1}(1-q)^{u01-1}p^{u11-1}(1-p)^{u10-1}$. In our analysis, specifically, $u_{00} = 1$, $u_{01} = 1$, $u_{11} = 1$, and $u_{10} = 1$.

The likelihood function for q and p is given by

$$L(q, p|X) = q^{n00}(1-q)^{n01}p^{n11}(1-p)^{n10},$$

where n_{ij} refers to the transitions from state i to j, which can be easily counted given X. Combining the prior distribution and the likelihood function, we get the following posterior distributions:

$$g(q, p|X) = g(q, p)L(q, p|X) \propto q^{u00+n00-1}$$

$$(1-q)^{u01+n01-1}p^{u11+n11-1}(1-p)^{u10+n10-1},$$

which implies that the posterior distribution is given by the two independent beta distributions:

$$q|X \sim beta(u_{00} + n_{00}, u_{01} + n_{01}), \text{ and}$$

$$p|X \sim beta(u_{11} + n_{11}, u_{10} + n_{10}),$$

from which q and p are drawn.

Furthermore, we assume that the chains of the model have a unique stationary (steady-state) distribution and that the initial-state distribution is equal to this steady-state distribution. The conditions of positive recurrent aperiodic chain guarantee the existence of a steady-state distribution (Turner, Cameron, and Thomson 1998, pp. 111–12). Given that this assumption holds, the initial-state distribution is completely determined by the transition equation $\pi'A = \pi'$, where A is the transition probabilities matrix in Equation 2 and $\pi = (\pi_0, \pi_1)'$ is the initial-state probabilities vector (Ross 1997, p. 180). In our two-state case, $\pi_0 = (1 - p)/(2 - q - p)$ for State 0, and $\pi_1 = 1 - \pi_0$ for State 1.

Step 4: Generate individual clients' regression parameters μ_a, β_a, δ_a, σ_a, μ_b, β_b, δ_b, σ_b, and σ_{ab} from $g(\mu_a, \beta_a, \delta_a, \sigma_a, \mu_b, \beta_b, \delta_b, \sigma_b, \sigma_{ab}|D, X)$.

This step can be further divided into two steps as follows:

Step 4a: Generate μ_a, β_a, δ_a, μ_b, β_b, and δ_b from $g(\mu_a, \beta_a, \delta_a, \mu_b, \beta_b, \delta_b|\sigma_a, \sigma_b, \sigma_{ab}, D, X)$.

Step 4a is applied to estimate all regression coefficients. Because the ordinary least squares estimation can be applied for each of the two regressions instead of general least squares estimation on the whole regression system when the two regressions have the same predictors, as in our case (Goldberger 1991, p. 327), Step 4a results in estimating two separate regressions independently. In Step 4a, we take the regression in Equation 1a first. We set the prior as follows:

$$\mu_a, \beta_a, \delta_a|\sigma_a \sim N(b_1, D_1),$$

where N indicates a normal distribution and both b_1 and D_1 are obtained from the population distribution in Step 1. Combining this with the normal likelihood

$$L(y_{at}|\mu_a, \beta_a, \delta_a, \sigma_a, z_t, x_t) \sim N(b_2, D_2)$$

results in the following normal posterior distribution:

$$\mu_a, \beta_a, \delta_a|\sigma_a, D, X \sim N(b_3, D_3),$$

where

$$b_3 = \frac{D_1^{-1}b_1 + TD_2^{-1}b_2}{D_1^{-1} + TD_2^{-1}} \text{ and}$$

$$D3 = (D_1^{-1} + TD_2^{-1})^{-1}.$$

We can follow exactly the same procedure for the second regression in Equation 1b.

Step 4b: Generate σ_a, σ_b, and σ_{ab} from $g(\sigma_a, \sigma_b, \sigma_{ab}|\mu_a, \beta_a, \delta_a, \mu_b, \beta_b, \delta_b, D, X)$.

To obtain the standard deviations and covariance, we consider $g(\sigma_a|\mu_a, \beta_a, \delta_a, D, X)$ first. We assume a gamma distribution as a conjugate prior for σ_a^{-2}:

$$\sigma_a^{-2}|\mu_a, \beta_a, \delta_a \sim \Gamma(\nu_1/2, \delta_1/2),$$

where Γ refers to a Gamma distribution. In our analysis, $\nu_1 = 6$, and $\delta_1 = 2$. Then, the posterior distribution for σ_a^{-2} is obtained as follows:

$$\sigma_a^{-2}|\mu_a, \beta_a, \delta_a, X, D \sim \Gamma(\nu_2/2, \delta_2/2),$$

where

$$\nu_2 = \nu_1 + T \text{ and}$$

$$\delta_2 = \delta_1 + \sum_{t=1}^{T}[y_{at} - (\mu_a + \beta_a z_t + \delta_a x_t)]^2.$$

We follow a similar procedure for σ_b and σ_{ab}, respectively.

REFERENCES

Albert, James H. and Siddhartha Chib (1993), “Bayesian Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts,” *Journal of Business and Economic Statistics*, 11 (1), 1–15.

Allenby, Greg M. and Peter E. Rossi (1999), “Marketing Models of Consumer Heterogeneity,” *Journal of Econometrics*, 89 (1–2), 57–78.

Baum, L.E. (1972), “An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of a Markov Process,” *Inequalities*, 3, 1–8.

Blattberg, Robert and K.J. Wisniewski (1989), “Price-Induced Patterns of Competition,” *Marketing Science*, 8 (Fall), 291–309.

Brangule-Vlasgma, Kristine, Rik Pieters, and Michel Wedel (2002), “The Dynamics of Value Segments: Modeling Framework and Empirical Illustration,” *International Journal of Research in Marketing*, 19 (3), 267–85.

Cooper, Lee G. (1988), “Competitive Maps: The Structure Underlying Asymmetric Cross-Elasticities,” *Management Science*, 34 (6), 707–723.

Dekimpe, Marnik and Dominique Hanssens (1995), “The Persistence of Marketing Effects on Sales,” *Marketing Science*, 14 (1), 1–21.

Digalakis, V., P. Monaco, and H. Murveit (1996), “Genones: Generalized Mixture Tying in Continuous Hidden Markov Model-Based Speech Recognizers,” *IEEE Transactions on Speech and Audio Processing*, 4 (4), 281–89.

Du, Rex Y. and Wagner A. Kamakura (2006), “Household Life Cycles and Lifestyles in the United States,” *Journal of Marketing Research*, 43 (February), 121–32.

Farris, Paul W. and John A. Quelch (1987), “In Defense of Price Promotion,” *Sloan Management Review*, 20 (1), 63–69.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (1995), *Bayesian Data Analysis*. New York: Chapman & Hall.

Goldberger, Arthur S. (1991), *A Course in Econometrics*. Boston: Harvard Business School Press.

Gonul, F., F. Carter, E. Petrova, and K. Srinivasan (2001), “Promotion of Prescription Drugs and Its Impact on Physicians’ Choice Behavior,” *Journal of Marketing*, 65 (July), 79–90.

Greene, William H. (1997), *Econometric Analysis*, 3d ed. Upper Saddle River, NJ: Prentice Hall.

Hanssens, Dominique M., Leonard J. Parsons, and Randall L. Schultz (2001), *Market Response Models: Econometric and Time Series Analysis*. Boston: Kluwer Academic Press.

Juang, B.H. and L.R. Rabiner (1991), “Hidden Markov Models for Speech Recognition,” *Technometrics*, 33 (August), 251–72.

Kamakura, Wagner A., Bruce Kossar, and Michel Wedel (2004), “Identifying Innovators for the Cross-Selling of New Products,” *Management Science*, 50 (August), 1120–33.

Kim, Chang-Jin and Charles R. Nelson (1999), *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge, MA: MIT Press.

Kirkpatrick, S., C.D. Gelatt Jr., and M.P. Vecchi (1983), “Optimization by Simulated Annealing,” *Science*, 220 (4598), 671–680.

Liechty, John, Rik Pieters, and Michel Wedel (2003), “Global and Local Covert Visual Attention: Evidence from a Bayesian Hidden Markov Model,” *Psychometrika*, 68 (December), 519–41.

Lodish, Leonard M., Magrid Abraham, Stuart Kalmenson, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens (1995), “How T.V. Advertising Works: A Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments,” *Journal of Marketing Research*, 32 (May), 125–39.

Manchanda, Puneet and Pradeep K. Chintagunta (2004), “Responsiveness of Physician Prescription Behavior to Salesforce Effort: An Individual Level Analysis,” *Marketing Letters*, 15 (2–3), 129–45.

———, Peter E. Rossi, and Pradeep K. Chintagunta (2004), “Response Modeling with Nonrandom Marketing-Mix Variables,” *Journal of Marketing Research*, 41 (November), 129–45.

Miklos, Istvan and Irmtraud M. Meyer (2005), “A Linear Memory Algorithm for Baum-Welch Training,” *BMC Bioinformatics*, (September 19), 231.

Mizik, Natalie and Robert Jacobson (2004), “Are Physicians ‘Easy Marks’? Quantifying the Effects of Detailing and Sampling on New Prescriptions,” *Management Science*, 50 (12), 1704–1715.

Montgomery, Alan, Shibo Li, Kannan Srinivasan, and John C. Liechty (2004), “Modeling Online Browsing and Path Analysis Using Clickstream Data,” *Marketing Science*, 23 (4), 579–86.

Mullahey, J. (1986), “Specification and Testing of Some Modified Count Data Models,” *Journal of Econometrics*, 33 (3), 341–65.

Narayanan, Sridhar, Puneet Manchanda, and Pradeep K. Chintagunta (2004), “Temporal Differences in the Role of Marketing Communication in New Product Categories,” working paper, Graduate School of Business, University of Chicago.

Parsons, Leonard and Piet Vanden Abeele (1981), “An Analysis of Sales Call Effectiveness,” *Journal of Marketing Research*, 17 (February), 107–113.

Rabiner, Lawrence R. (1989), “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition,” *Proceedings of the IEEE*, 77 (2), 257–85.

——— and B.H. Juang (1986), “An Introduction to Hidden Markov Models,” *IEEE ASSP Magazine*, (January), 4–16.

Ross, Sheldon M. (1997), *Introduction to Probability Models*, 6th ed. San Diego: Academic Press.

Turner, T. Rolf, Murray A. Cameron, and Peter J. Thompson (1998), “Hidden Markov Chains in Generalized Linear Models,” *Canadian Journal of Statistics*, 26 (1), 107–125.