PUNEET MANCHANDA, PETER E. ROSSI, and PRADEEP K. CHINTAGUNTA*

Sales response models are widely used as the basis for optimizing the marketing mix. Response models condition on the observed marketing-mix variables and focus on the specification of the distribution of observed sales given marketing-mix activities. The models usually fail to recognize that the levels of the marketing-mix variables are often chosen with at least partial knowledge of the response parameters in the conditional model. This means that contrary to standard assumptions, the marginal distribution of the marketing-mix variables is not independent of response parameters. The authors expand on the standard conditional model to include a model for the determination of the marketing-mix variables. They apply this modeling approach to the problem of gauging the effectiveness of sales calls (details) to induce greater prescribing of drugs by individual physicians. They do not assume *a priori* that details are set optimally, but instead they infer the extent to which sales force managers have knowledge of responsiveness, and they use this knowledge to set the level of sales force contact. The authors find that their modeling approach improves the precision of the physician-specific response parameters significantly. They also find that physicians are not detailed optimally; high-volume physicians are detailed to a greater extent than low-volume physicians without regard to responsiveness to detailing. It appears that unresponsive but high-volume physicians are detailed the most. Finally, the authors illustrate how their approach provides a general framework.

Response Modeling with Nonrandom Marketing-Mix Variables

Researchers in marketing have made a great deal of progress in formulating and applying sales response models to elements of the marketing mix. However, in virtually all cases, the sales response models are conditional or "regression-style" models in which the values of the independent variables are fixed at observed values. The models make the implicit assumption that the marketing-mix variables are set *independently* of the response function parameters. This is at variance with the theory of optimal marketing-mix allocation, which requires that marketing-mix variables be set so as to equalize the ratio of marginal benefit to cost across all mix activities. In practice, we also observe nonrandom allocation of marketing-mix variables on the basis of various proxies for responsiveness. The strategic or nonrandom setting of the marketing-mix variables produces a simultaneity problem in the estimation of the sales response model. We develop a general framework for attacking this particular simultaneity problem, and we apply this framework to data on allocation of sales force efforts in the pharmaceutical industry.

Our empirical application is based on data on sales calls (termed "details") made to physicians for the purpose of inducing them to prescribe a specific drug. Detailing is the single largest marketing activity in the pharmaceutical industry; it has more than three times the expenditure of the second largest activity: direct-to-consumer advertising (Wittink 2002). In theory, sales managers should allocate

*Puneet Manchanda is Associate Professor of Marketing (e-mail: puneet.manchanda@chicagogsb.edu), Peter E. Rossi is Joseph T. Lewis Professor of Marketing and Statistics (e-mail: peter.rossi@chicagogsb.edu), and Pradeep K. Chintagunta is Robert Law Professor of Marketing (e-mail: pradeep.chintagunta@chicagogsb.edu), University of Chicago. The authors thank Sridhar Narayanan for help with the data and an anonymous firm for providing the data. The two anonymous *JMR* reviewers, participants at the Marketing Science conference in Edmonton, seminar participants at Stanford, Olin, Cornell, New York University, Wharton, University of Iowa, University of California at Berkeley, Kellogg School of Management, University of California at Los Angeles, University of California at Riverside, University of California at Irvine, Wyeth Pharmaceuticals, Xavier Dréze, and Rob McCulloch provided valuable comments on previous drafts. The authors thank the Kilts Center for Marketing at the Graduate School of Business, University of Chicago, for research support. The first author also thanks the Beatrice Foods Company Faculty Research Fund for research funding.

Journal of Marketing Research
Vol. XLI (November 2004), 467–478

detailing efforts across the many thousands of regularly prescribing physicians so as to equalize the marginal impact of a detail across doctors (under the assumption of equal marginal cost, which is reasonable according to industry sources).

The barrier to implementing optimal allocation of detailing effort is the availability of reliable estimates of the marginal impact of a detail. Although individual physician-level data are available on the writing of prescriptions from syndicated suppliers such as IMS Health and Scott-Levin, practitioners do not fit individual physician models because of the paucity of detailing data and extremely noisy coefficient estimates obtained from the data. Instead, practitioners pool data across physicians in various groups, usually on the basis of total drug category volume. Detailing targets are announced for each group. In general, higher-volume physicians receive greater detailing attention. Even if detailing had no effect on prescription behavior, volume-based setting of the detailing independent variable would create a spurious detailing effect in pooled data. In addition to general rules that specify that detailing levels are related to volume, individual sales force managers adjust the level of detailing given informal sources of knowledge about the physician. This has the net effect of making the levels of detailing a function of baseline volume and, possibly, detailing responsiveness. Thus, the independent variable in our analysis has a level that is related to parameters of the sales response function.

Given the need for physician-specific detailing effects, it might seem natural to apply Bayesian hierarchical models to this problem. Bayesian hierarchical models "solve" the problem of unreliable estimates from individual physician models by a form of "shrinkage," or partial pooling, in which information is shared across models. A Bayesian hierarchical model can be viewed as a particular implementation of a random-coefficients model. The simultaneity problem we discuss herein is a generalization of Chamberlain's (1980, 1984) formulation, in which the independent variable is made a function of random intercepts. Chamberlain shows that there can be asymptotic bias for models estimated with only the conditional likelihood function. In this context, the standard Bayesian hierarchical models suffer from simultaneity bias because the independent variables are functions of both random intercepts and slopes.

If detailing levels are functions of sales response parameters, standard Bayesian hierarchical models will be both biased and inefficient. The inefficiency, which can be substantial, results from the *level* of the independent variable having information about the response coefficients. The standard approach simply does not use this information. We supplement the sales response function by an explicit model for the distribution of detailing, which has a mean that is related to response coefficients. In our application, given that sales (prescriptions) and detailing are count data, we use a negative binomial distribution (NBD) regression as the sales response function and a Poisson distribution for detailing. We demonstrate that this joint model provides much more precise estimates of the effects of detailing and improved predictive performance. Rather than impose optimality conditions on our model, we estimate the detailing policy function that sales managers use.

The problem we consider herein is a special case of the more general setting in which the marginal distribution of the independent variable is not independent of the conditional distribution of the dependent variable given the independent. This general class of problems can be characterized by the situation in which the independent variables are set strategically or in which the independent variables are endogeneous. It should be emphasized that our problem is different from the problem of price endogeneity, as in the work of Villas-Boas and Winer (1999) or Nevo (2001). In the case of price endogeneity, prices are set strategically as a function of a *common* demand shock. In our case, the detailing levels are set as a function of *physician-specific* response parameters. Our approach is closer to that of Bronnenberg and Mahajan (2001), who postulate that marketing-mix variables are set as a function of the baseline level of sales.

The article is organized as follows: We first describe our data. We then formulate the standard sales response model (i.e., in which the marketing-mix elements are independent of the sales response coefficients) in a hierarchical Bayesian framework. We use the stated policy for determining detailing to develop our approach to the simultaneous determination of both prescriptions and detailing. We then contrast the results from our simultaneous model with the results we obtained from the standard conditional approach. We demonstrate the superiority of our approach through predictive validation. The specific parametric models introduced herein are special cases of a general framework for situations in which the marketing-mix variables are set with some knowledge of individual response parameters. We discuss this general framework next. Finally, we provide concluding remarks.

DATA

A major U.S. pharmaceutical firm made data on physician prescription behavior and sales force effort for a specific drug available to us (the firm has requested that we do not reveal any other information). The data represent a detailed record of physicians' prescription behavior for the drug in question (which we refer to as Drug X) over the period from June 1999 to June 2001. Drug X belongs to a mature product category and was under patent during the time our data were collected. The number of affected people in Drug X's therapeutic category is estimated to be approximately 19 million, which makes it one of the top three prescription categories in the United States.

Data on the number of prescriptions written for Drug X and for the balance of the Drug X category are available by physician on a monthly basis. Data on sales contacts or details are also available from internal sales records. Our data only provide the number of details (measured as an office visit with physician contact) per month in which Drug X is detailed. The only information available about the nature of the contacts is about the number of free samples (which typically comprise one course of treatment) of Drug X given to each physician by month. Thus, our analysis should be interpreted as pertaining to an "average" detail and sample drop.

In our analysis, we drew a sample of 1000 physicians from a restricted population of 112,088 regularly prescribing physicians who have received at least two details in 1 of the 24 months of observation. We impose this restriction to ensure that there is sufficient variability in detailing to estimate detailing effects. We have a total of 24,000 observa-

tions, that is, 24 months of data for 1000 physicians. On average, a physician in our sample writes approximately five new prescriptions for Drug X and receives 1.8 details and 5.6 (product) samples per month.

In terms of specialty, on the basis of our discussions with industry specialists, there are three types of physicians across which we expect to find differences in prescription behavior. These are the specialty directly related to the drug benefit or patient problem (labeled SPE), primary care/family physicians (PCPs), and all other specialties (OTH). Of our sample of physicians, 18.5% are SPEs, and 60.1% are PCPs.

CONDITIONAL MODELING APPROACH

A conditional model for the distribution of prescriptions written given detailing and sampling is the starting point for our analysis. Our data are count data with most observations at less than ten prescriptions in a given month. Figure 1 (Panel A) plots the mean and variance (computed over the 24 monthly observations) for each of the 1000 physicians and shows evidence that the variance often exceeds the mean. This does not conclusively prove that the conditional distribution of prescriptions given detailing is overdispersed, but it is consistent with this. If there are significant differences between physicians in the parameters of the conditional distribution, we observe a mixture of conditional distributions that could appear to be overdispersed. We adopt the NBD as the base model for the conditional distribution, and we couple this model with a model of the distribution of coefficients over physicians. The NBD model is flexible in that it can exhibit a wide range of degrees of overdispersion, thereby enabling the data to resolve the issue. An NBD distribution with mean λ_{it} and overdispersion parameter α is provided in Equation 1.

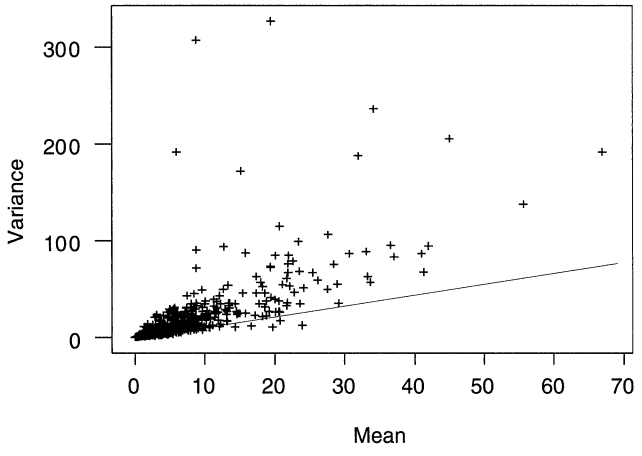
$$(1)\qquad \Pr(y_{it}=k|\lambda_{it})=\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(k+1)}\left(\frac{\alpha}{\alpha+\lambda_{it}}\right)^{\alpha}\left(\frac{\lambda_{it}}{\alpha+\lambda_{it}}\right)^{k},$$

where y_{it} is the number of new prescriptions written by physician i in month t. As α approaches infinity, the NBD distribution approaches the popular Poisson distribution.

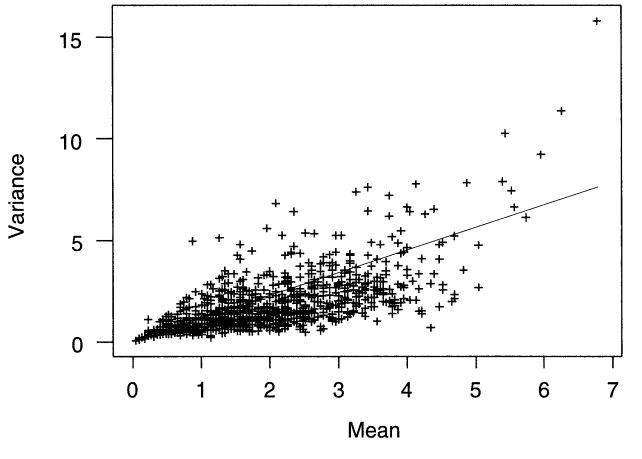
The specification of the conditional mean is determined by the nature of detailing effects. Many analyses of aggregate sales and detailing use cumulative detailing measures that are the cumulative discounted number of details (see Neslin 2001; Wittink 2002). The notion is that the effects of detailing not only affect current month's prescriptions but also carry over to future months' prescription behavior. Rather than use cumulative detailing measures that involve somewhat ad hoc choices of the discount or smoothing parameters, we include a lagged prescriptions term in the conditional mean function to allow for carryover effects, as in much of the time-series literature on advertising effects (Clarke 1976). The literature on detailing effects has emphasized diminishing returns to detailing (Gönül et al. 2001; Manchanda and Chintagunta 2000). This can be accomplished by adding a detailing-squared term to the regression function. We considered a model with physician-specific squared terms, but we found that we could not estimate these coefficients reliably. Because detailing rarely varies outside a range of 0–5 details per month, it may be too much to expect this data set to estimate physician-level

Figure 1
OVERDISPERSION DIAGNOSTICS: PHYSICIAN MEANS AND VARIANCES

A: New Prescriptions per Month (1000 Physicians over 24 Months)



B: Details per Month (1000 Physicians over 24 Months)



diminishing returns. Therefore, we did not include squared terms in the models we report.

We adopt the standard log-link function and specify that the log of the mean of the conditional distribution is linear in the parameters.

$$(2)\qquad \lambda_{it}=E[y_{it}|x_{it}]=\exp(x_{it}'\beta_i),$$

and

$$(3)\qquad \ln(\lambda_{it})=\beta_{0i}+\beta_{1i}\mathrm{Det}_{it}+\beta_{2i}\ln(y_{it-1}+d).$$

The lagged log-prescriptions term, $(y_{it-1}+d)$, in Equation 3 allows the effect of detailing to influence not only the current period but also subsequent periods. We add "d" to the lagged level of prescriptions to remove problems with zeros in the prescription data. The smaller the number added, the more accurate is the Koyck solution as an approximation. The problem is that the log of small numbers can be of large magnitude, which would give the zeros

in the data undue influence on the carryover coefficients. We choose d = 1 as the smallest number that will not create large outliers in the distribution of $\ln(y_{it} + d)$.

To complete the conditional model, we specify a distribution of coefficients across physicians. This follows a standard hierarchical formulation.

$$\beta_i = \Delta z_i + \nu_i, \text{ and} \quad (4)$$

$$\nu_i \sim N(0, V_\beta).$$

The z vector includes information on the nature of the physician's practice and level of sampling (note that we use PCPs as the base physician type): $z' = (1, SPE, OTH, SAMP)$, where SAMP is the mean (per-physician) number of monthly samples, divided by ten. This specification of z and the model in Equation 4 allows for a main effect and an interaction for both physician specialty type and sampling. We might expect that SPEs will have a different level of prescription writing. In addition, detailing may be more or less effective, depending on the physician's specialty. Sales calls may include the provision of free Drug X samples. The effect of sampling is widely debated in the pharmaceutical industry; some people argue that it enhances sales, and others argue that cannibalization is a major effect. Most people believe that sampling is of secondary importance to detailing. Sampling is conditional on detailing in that it cannot occur without a detail visit. For this reason, we include the average sampling variable in the mean of the hierarchy, which creates an interaction term between detailing and sampling.

BEYOND THE CONDITIONAL MODEL

The firm that produces Drug X does not set detailing levels randomly. The firm recognizes that detailing targets for the sales force should be set at the physician level and not at some higher level, such as sales territory. According to discussions with the firm, detailing is set primarily on the basis of the physician decile, computed by IMS Health, for the quarter before the annual planning period.[1] IMS Health assigns each physician to a decile based on the physician's total prescription writing for all drugs in the therapeutic class. The physician level annual targets are then adjusted quarterly on the basis of previous quarter deciles. The quarterly adjustments tend to be minor.

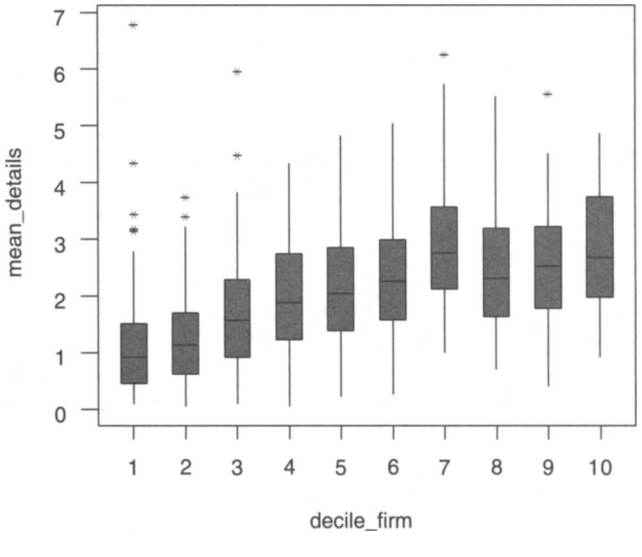
Conditional modeling approaches rely on the assumption that the marginal distribution of the independent variables does not depend on the parameters of the conditional distribution specified in Equation 3. If total category volume is correlated with the parameters of the conditional response model, this assumption will be violated. We believe that it is highly likely that physicians who write a large volume of Drug X prescriptions regardless of detailing levels (e.g., have high value of the intercept in Equation 3) also have higher-than-average category volume. This means that marginal distribution of detailing depends, at the minimum, on the intercept parameter in Equation 3. This dependence is the origin of the spurious correlation that can occur if higher-volume physicians are detailed more often.

In addition, although detailing targets are set on an annual basis (and revised quarterly), there is much month-to-month variation in detailing as a result of factors outside of the control of the sales force managers. In addition, even though detailing targets are set at a high level in the firm, sales force district or territory managers may change the actual level of detailing on the basis of their own specialized knowledge about specific physicians. Figure 2 shows the distribution of the mean level of detailing for each physician in our sample, plotted against the deciles that the firm uses to determine detailing targets. There is a relationship between decile and mean detailing in which the higher deciles (which represent higher category volume) are detailed at approximately twice the level of the lowest deciles. However, there is a great deal of variation *within* each decile.

The high level of intradecile variation displayed in Figure 2 shows that there are other considerations of the sales force in allocating detailing resources. If sales force managers had full knowledge of the functional form and parameters of the detailing response function, detailing would be allocated so as to equalize the marginal effects across physicians. Given that the current industry practice is not to compute individual physician estimates, it is unreasonable to assume that firms use a full-information optimal allocation approach.

We adopt a specification of the detailing distribution that allows for some partial knowledge of detailing response parameters. A simple but flexible approach would be to assume that detailing is i.i.d. with mean set as a function of the long-term response parameters from Equation 3. Note that the average first-order autocorrelation for detailing is less than .3. Monthly detailing is a count variable with

Figure 2
DISTRIBUTION OF MEAN DETAILING BY TOTAL CATEGORY (FIRM) DECILE



Notes: We computed the mean detail for all physicians across the 24-month period.

[1]Note that the firm also uses other behavioral measures and overlays these on the volumetric deciles to generate a more complex segmentation plan, which then becomes the basis for carrying out the detailing allocation. However, the volumetric decile typically explains the largest amount of variance in the detailing plan across physicians. The firm also takes into consideration the portfolio of products that it markets to a given physician in setting an individual call plan for that physician.

rarely more than five details per month. Figure 1 (Panel B) shows that the mean and variance of detailing are approximately equal, which indicates that overdispersion is not a problem for the detailing variable. We model detailing as an i.i.d. draw from a Poisson distribution with a mean that is a function of baseline sales and the long-term response to detailing.

Equation 3 shows that the long-term effects of detailing depend on the size of the lagged log coefficient. However, the exact expressions for the cumulative or long-term effect of a change in detailing do not exist in closed form. For ease of interpretation, we provide the following intuition as to the size of these long-term effects through an approximate solution. Note that we use this approximation only to characterize the long-term impact of detailing. The long-term mean level of prescriptions, μ_i^*, is approximately the solution to the following equation:[2]

$$\ln(\mu_i^*) \doteq \beta_{0i} + \beta_{1i}\text{Det} + \beta_{2i}\ln(\mu_i^* + d). \tag{5}$$

The exact solution to Equation 5 depends on the level of μ_i^*. However, we can approximate the solution by the standard Koyck lag solution.

$$\ln(\mu_i^*) \doteq \left[\frac{\beta_{0i}}{(1-\beta_{2i})}\right] + \left[\frac{\beta_{1i}}{(1-\beta_{2i})}\right]\text{Det}. \tag{6}$$

The i.i.d. model of detailing is as given by the Poisson distribution:

$$\Pr(x_{it} = m|\eta_i) = \frac{\eta_i^m \exp(-\eta_i)}{m!}. \tag{7}$$

The mean of this Poisson distribution is a function of the (approximate long-term) coefficients (Equation 6) as

$$\ln(\eta_i) = \gamma_0 + \gamma_1\left[\frac{\beta_{0i}}{(1-\beta_{2i})}\right] + \gamma_2\left[\frac{\beta_{1i}}{(1-\beta_{2i})}\right]. \tag{8}$$

The specification in Equation 8 allows for various possibilities. If detailing is set with no knowledge of responsiveness to detailing, we should expect that $\gamma_2 = 0$. In the case of $\gamma_2 = 0$, mean detailing is a function of baseline sales, or sales in the absence of detailing activity. This can be regarded as an approximation to the stated policy of setting detailing as a function of total category volume. We note that the actual policy of detailing appears to differ materially from the stated policy, as is revealed in Figure 2. For this reason, we prefer the specification in Equation 8, which provides flexibility to describe various policies rather than to include past y values in the equation. However, if detailing is set with some knowledge of responsiveness to detailing, we should expect that γ_1 and γ_2 have posteriors massed away from zero. There are a variety of functional forms for the relationship between the mean level of detailing and the response parameters. We regard our specification as exploratory and as a general linear approximation to some general function of long-term effects.

In summary, our approach is to enlarge the conditional model by specifying a model for the marginal distribution of detailing, which depends on conditional response parameters. Using the standard notation for conditional distributions in hierarchical models, we can express the new model as follows:

$$y_{it}|\text{Det}_{it}, \beta_i, \alpha, \quad \text{NBD regression}, \tag{9}$$

and

$$\text{Det}_{it}|\beta_i, \gamma, \quad \text{Poisson marginal}. \tag{10}$$

This dependence of marginal distribution on the response parameters alters the standard conditional inference structure of hierarchical models. In the standard conditional model given only by Equation 9, inference about the response parameters, β_i, is based on time-series variation in detailing for the same physician and through similarities between physicians, as expressed by the random effects or the first-stage prior. However, when Equation 10 is added to the model, inferences about β_i change as new information is available from the level of detailing. That is, the additional structure added by the marginal model of detailing in Equation 10 can systematically change our estimates. Under the specification in Equations 9 and 10, the conditional estimates based on Equation 9 alone exhibit asymptotic bias.

The marginal model in Equation 10 implies that the level of detailing is informative about responsiveness, and this information is incorporated into the final posterior on β_i. For example, suppose that $\gamma_2 < 0$ in Equation 8, then detailing is set so that less responsive physicians are detailed at higher levels, which provides an additional source of information that will be reflected in the β_i estimates. Thus, the full model consisting of Equations 9 and 10 can deliver improved estimates of physician-level parameters by exploiting information in the levels of detailing. The model specified by Equations 9 and 10 is conditional on β_i. We add the heterogeneity distribution on β_i as given by Equation 4.

Another way to appreciate this point is to observe that likelihood for β_i has two components: the NBD regression and the Poisson marginal model.

$$\ell(\{\beta_i\}) = \prod_i \prod_t p_{\text{NBD}}(y_{it}|\text{Det}_{it}, \beta_i, \alpha)\, p_{\text{Poisson}}(\text{Det}_{it}|\beta_i, \gamma), \tag{11}$$

where β_i is identified from both the NBD and the Poisson portions of the model. Examination of the form of the likelihood in Equation 11 and the mean function in Equation 8 indicates some potential problems for certain data configurations. In the Poisson portion of the model, elements of the γ vector and the collection of β_i values enter multiplicatively. In terms of the Poisson likelihood, $[\gamma_1, \{\beta_{1i}\}]$ and $[-\gamma_1, -\{\beta_{1i}\}]$, for example, are observationally equivalent. What identifies the signs of the parameters is the NBD regression. In other words, if the signs of the detailing coefficients are flipped, the NBD regression fits suffer, thus lowering the posterior at that mode. This suggests that in data sets in which there is only weak evidence for the effects of detailing (or, in general, of any independent variable), there exist two modes in the posterior of comparable height. Navigation between the modes can be difficult for Metropolis–Hastings algorithms. To gauge the magnitude of this problem, we simulated several different data sets with varying degrees of information about the effect of the

[2]To achieve this approximation, we approximate the expectation of the log with the log of the expectation. Simulation results showed that in the range of the estimated parameters, this approximation is quite accurate.

independent variable. We found that for moderate amounts of information, similar to that encountered in our data, the multimodality problem was not pronounced. However, for situations with little information about the $\{\beta_i\}$, there could be two modes.

RESULTS

We implement both the conditional NBD model (Equation 9) with the hierarchy outlined in Equation 4 and what we term the "full" model, which consists of the NBD regression and the Poisson marginal model (Equations 9 and 10) coupled with the hierarchy mentioned in Equation 4. Inference is conducted by means of a blocked random-walk Metropolis algorithm (details on the algorithm are available in the Appendix). With all models, we run the Markov chain Monte Carlo (MCMC) algorithm for 50,000 iterations, and we discard the first 25,000 to ensure adequate dissipation of initial conditions, or burn-in. We have investigated the performance of our algorithms with simulated data to ensure that 50,000 iterations are enough to navigate the full posterior effectively.

We begin by discussing results for the conditional model. Our hierarchical specification allows the physician practice type and sampling to influence both the base rate of prescription writing for Drug X and the effects of detailing and the carryover effect. Recall that we specify that the means of the random-coefficient distribution or first-stage prior are linked to these variables through a matrix of coefficients: $\beta_i = \Delta z_i + \nu_i$. Table 1 provides the posterior distribution of the Δ coefficients.

Table 1 shows that physicians with a specialization in the therapeutic class (SPE) write [exp(.26 + .44)/exp(.26)], or 55% more mean prescriptions than PCPs.[3] They also write 115% more mean prescriptions than OTH. Mean sampling (SAMP) also has a large effect on prescription writing. Detailing tends to have more effect on SPEs than PCPs, and there is a pronounced negative interaction between sampling and detailing. The effect of detailing is reduced when sampling is at a higher level. Finally, there is an average carryover coefficient of approximately .3 (i.e., most effects of a detail occur within three months), with higher carryover for SPEs.

Figure 3 plots the posterior means from the conditional NBD model against the coefficients from the full model. For the intercept, the coefficients are nearly identical; however, the coefficients on detailing differ dramatically. The full model shrinks several of the larger coefficients to smaller values. Perhaps more important, the correlation

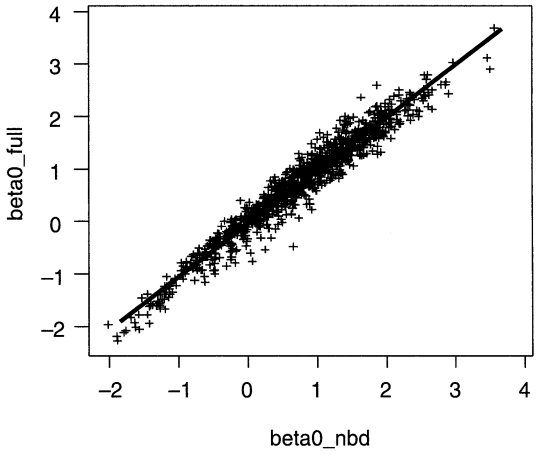
[3]We must exponentiate the predicted mean because the hierarchy is specified in the log-conditional mean (see Equation 2).

Table 1
RESPONSE (β) PARAMETERS

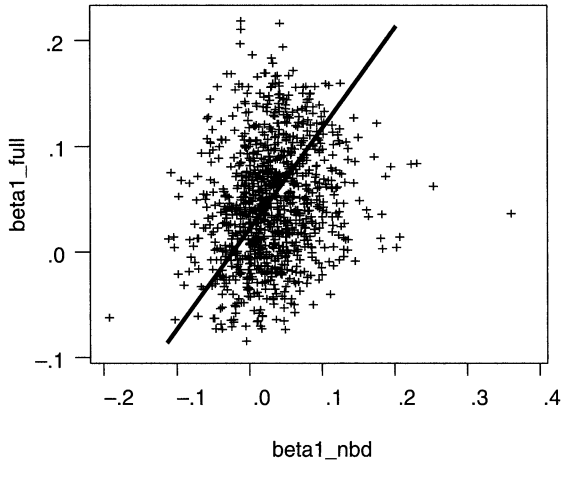
	Intercept	*SPE*	*OTH*	*SAMP*
β_{0i}	.26	.44	–.33	.72
	(.050)	(.090)	(.078)	(.052)
β_{1i}	.038	.016	.006	–.020
	(.0083)	(.011)	(.015)	(.0064)
β_{2i}	.28	.15	.037	–.01
	(.018)	(.032)	(.032)	(.019)

Figure 3
COMPARISON OF PHYSICIAN-LEVEL ESTIMATES: CONDITIONAL NBD VERSUS FULL MODELS

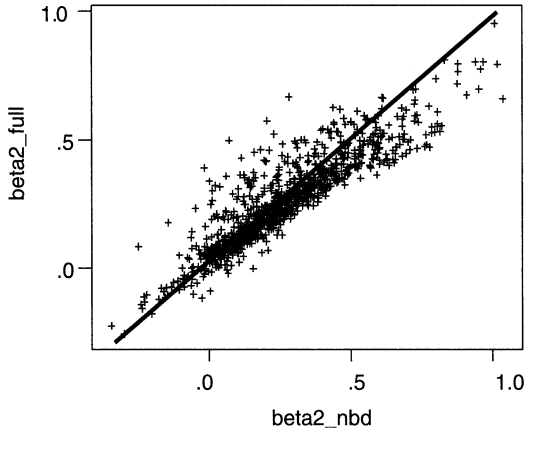
A: Beta0 Comparison



B: Beta1 Comparison



C: Beta2 Comparison



between the full and conditional model coefficients is weak. This suggests that the information obtained from the levels of detailing along with the inferred detailing allocation rule is quite different from the information available only from the response to the time-series variation in detailing used in the conditional model. The full model also reduces the size of the carryover effect.

The most striking difference between the coefficient estimates from the conditional and full models is the precision of the estimates. If we examine the posterior distribution of the long-term effects of detailing, $[\beta_{1i}/(1 - \beta_{2i})]$, we find that only 50 of 1000 physicians have posterior distributions of the long-term effect, which have more than 90% of posterior mass on positive values. This means that for most of the physicians, we cannot reliably estimate detailing effects in the conditional model. However, the full model exploits not only the information in time-series variation in detailing but also the information in the level of detailing. This information is of exceptional value in estimating detailing effects. More than 540 of the 1000 long-term effects are precisely estimated for the full model. The long-term baseline value and the carryover effect are also more precisely estimated. As we show subsequently, this improvement in precision at the individual level leads to better forecasts.

Our specification of the detailing marginal distribution allows for flexibility in the nature of the relationship between the coefficients in the conditional model and the mean value of detailing. Recall from Equation 8 that we specify $\ln[E(\text{detailing}) = \eta_i] = \gamma_0 + \gamma_1[\beta_{0i}/(1 - \beta_{2i})] + \gamma_2[\beta_{1i}/(1 - \beta_{2i})]$. Table 2 provides the information on the posterior distribution of γ. The posterior of γ reveals that, as we expected, the higher the intercept, the higher is the level of detailing. The intercept determines the level of "baseline" volume of prescriptions in the absence of promotional detailing. The larger-volume physicians for Drug X will have a higher monthly average number of details.

We believe that more surprising is the large negative coefficient on the long-term effect of detailing. This implies that, all other things being equal, physicians who are less responsive to detailing are detailed more than physicians who are more responsive. We also estimate V_β, the covariance matrix of the physician-level β coefficients. The correlation between β_{0i} and β_{1i} is estimated to be –.03, which means that the intercepts and detailing effect variables have near-zero correlation. Thus, the data are consistent with two independent dimensions in the setting of mean detailing. The first effect is the high-volume effect, and the second, and independent, effect is that lower response to detailing physicians is detailed more than higher response to detailing physicians.

To provide a sense of the relative size of the effects, Figure 4 plots the distribution of the mean level of detailing by decile of the β_{0i} and β_{1i} distributions. There are large differences in detailing; both coefficients vary. From the smallest to largest decile of the intercept, mean detailing increases fivefold from approximately .5 details per month to 2.5 details per month. The response-coefficient effect is even more dramatic, ranging from around 3.7 details per month for the least responsive physicians to approximately .5 details per month for the most responsive.

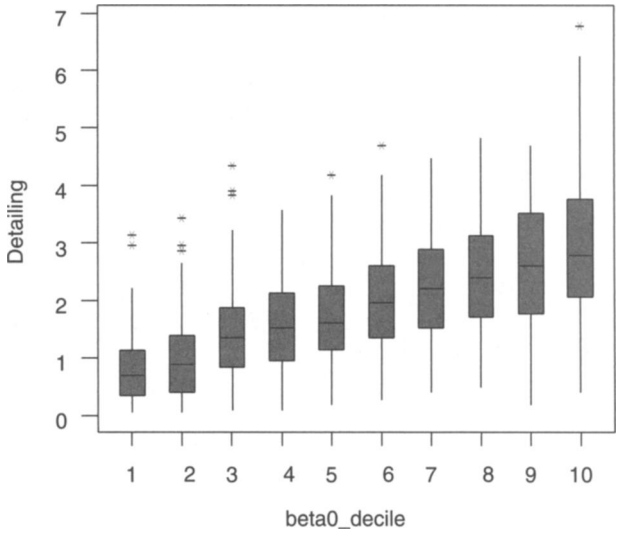
The results are difficult to reconcile with most standard models of optimal sales force allocation. Thus, it might be argued that our full model is misspecified in some important way so that the results are an artifact of the model approach and are not supported by "model-free" evidence. Given that we have 23,000 observations (we lost 1 observation per physician in creating the lagged prescription vari-

Table 2
γ PARAMETERS

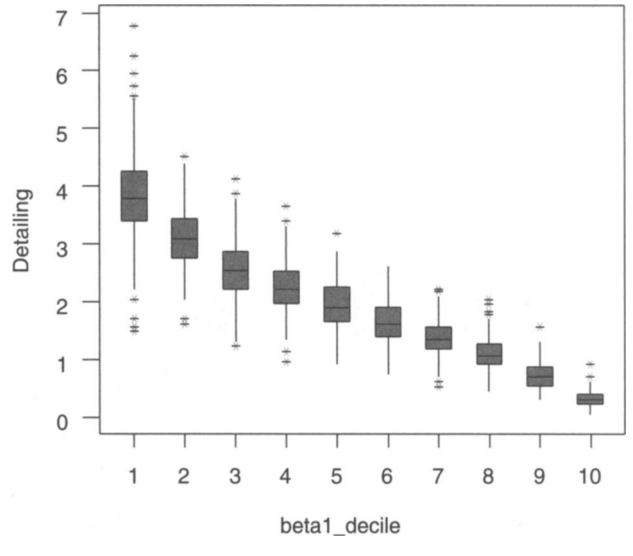
Parameter	*Mean*	*Standard Deviation*
γ_0	.73	.083
γ_1	.19	.040
γ_2	–6.1	.35

Figure 4
DISTRIBUTION OF DETAILING BY PARAMETER DECILES: FULL MODEL

A: Intercept



B: Detailing Slope Coefficient

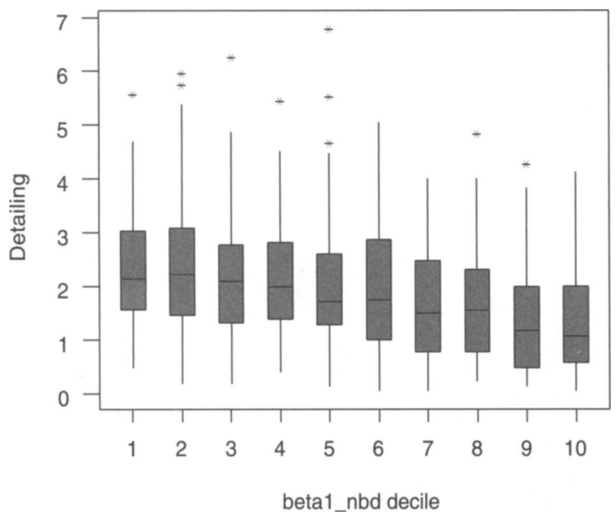


able) and that the data are discrete, it will be difficult to provide truly model-free corroboration of our findings. However, we can examine the extent to which the *conditional* NBD estimates support our findings. Figure 5 plots the distribution of detailing for each of the deciles formed from the NBD estimates of the detailing effect coefficients. We have found that these are noisy estimates, but they have the advantage of not requiring additional assumptions about the way detailing is set. Figure 5 shows similar effects as does Figure 4, but they are of smaller magnitude. According to the NBD model, the least responsive physicians have almost twice as many details as the most responsive physicians.

Our results imply that physicians with a higher baseline level of prescription writing are detailed more. This is in accord with the stated detailing policy. However, we also find that highly detailed physicians have low marginal responses to detailing, and the least detailed physicians have high responsiveness. Both effects are independent and of approximately the same magnitude. A possible explanation for this finding is that the effects are driven primarily by the absence of competitive data. It is possible that there is more competitive detailing activity for the most responsive physicians and that this lowers the effectiveness of detailing. Another explanation for the finding is that all physicians have the same concave response curve to detailing, and we are simply detecting declining derivatives along this concave curve.[4] However, an examination of the data does not provide evidence in support of this explanation. If there were a common response curve, optimal policies would dictate an equal level of detailing for all physicians. As is shown in Figure 2, we observe systematic differences in the level of detailing by decile.

[4]We thank an anonymous reviewer for bringing this explanation to our attention.

Figure 5
DISTRIBUTION OF DETAILING BY RESPONSE COEFFICIENT DECILE: CONDITIONAL MODEL



Another explanation of this finding is that there is suboptimal allocation of sales force effort. Suppose that physicians were targeted for a high level of details with some criterion that is not related to true responsiveness. For example, industry managers talk about physicians who are "opinion leaders" and emphasize the importance of targeting them. If there were diminishing returns to detailing, we would observe lower marginal effects for the highly detailed physicians. Finally, as we noted previously, our results apply to an "average" detail. Our results might be explained by the presence of systematic variation across physician responsiveness in detail attributes (e.g., the minutes spent, the number of drugs detailed). The bottom line is that in this world, allocation away from the least responsive physicians to the more responsive ones should result in higher sales. For this reason, we examine the first-order conditions for optimal allocation and the extent to which they are consistent with our data.

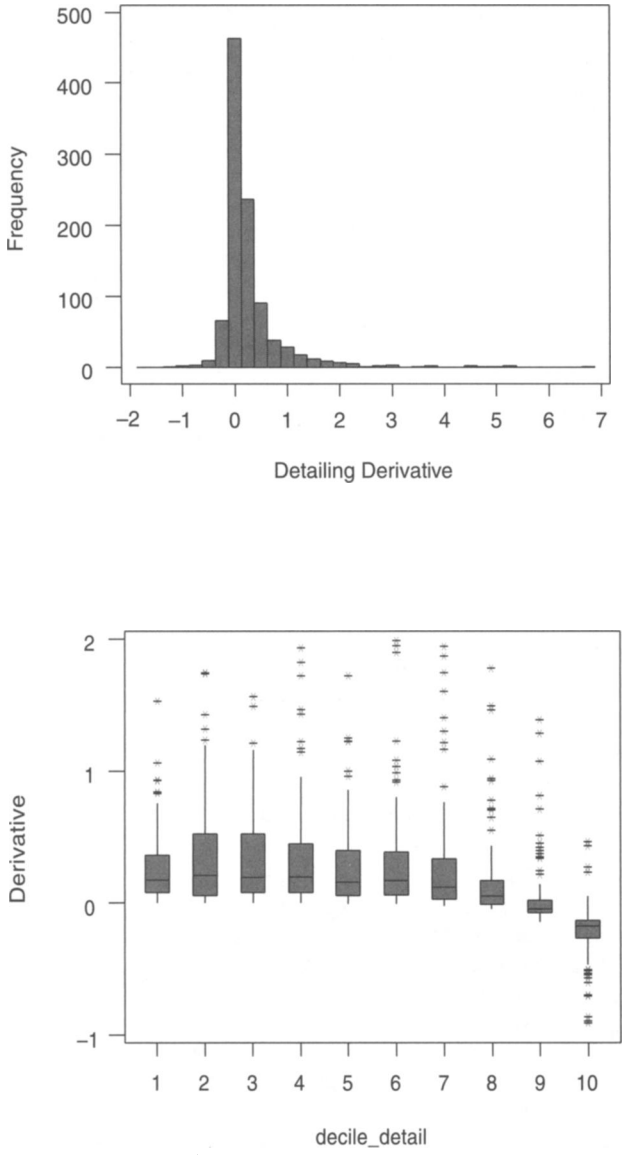
Information about the optimality of the current allocation of detailing levels is provided by derivative of the expected long-term number of prescriptions with respect to detailing (Equations 5–6):

$$(12)\qquad \frac{\partial}{\partial \text{Det}} e^{\ln(\mu_i^*)} = e^{\ln(\mu_i^*)}\Big|_{\text{Det}\,=\,D^*}\left[\frac{\beta_{1i}}{(1-\beta_{2i})}\right].$$

Our model provides a local approximation to this derivative in the region of detailing observed for each physician. Evaluating Equation 12 at the mean level of detailing for each physician, we obtain the histogram in the top panel of Figure 6. A value of 1 means that we expect that for this physician, an additional detail will generate one more new prescription in the long run. It should be emphasized that one more new prescription generates an expected flow of refill prescriptions so that ultimate profitability of the detail call cannot be calculated without some assumptions about the expected number of future refills. The histogram in Figure 6 shows a great deal of variation across physicians in the level of the derivative. However, most of the physicians have derivatives that appear to be close to zero.

To obtain a better appreciation for the size of the derivative, we consider the optimality condition for the setting of detailing levels. Under optimality, the marginal profits generated by an additional detail must equal the cost of the detail. As we mentioned previously, industry sources place the cost of a detail at approximately $80. The derivative computed in Equation 12 is the expected change in the number of new prescriptions created by the marginal detail. In some sense, this is similar to the expected change in the number of patients who use Drug X. Each new patient generates a flow of prescriptions and profits (note that the marginal cost of a prescription is virtually zero, so the flow is price × expected number of prescriptions over the lifetime of the new patient). Although we do not have the patient-level data required to build a lifetime model of revenues, we can gauge the plausibility of our derivative numbers by computing the revenue stream that is required to equate expected profits to the marginal cost of a detail. If change in profits = change in new prescriptions × total revenue generated by a new prescription, we can use our derivative measure to solve for total revenue that is required to equate change in profits with the marginal cost of $80 (divide $80 by the derivative value for each physician).

Figure 6
DISTRIBUTION OF DETAILING DERIVATIVES



$$\text{Tot Rev}_i = \frac{\text{MC}}{\frac{\partial}{\partial \text{Det}} e^{\ln(\mu_i \mid \text{Detailing} = \bar{D}_i)}}. \tag{13}$$

The median of this distribution is $296 with an interquartile range of $50 to $917. Using aggregate data from IMS Health, we find that the margin on a prescription for Drug X is approximately $100. This implies that for the derivatives to imply optimal setting of detailing, we would require each new prescription to generate approximately two future refills at the median of the distribution. In our view, a lifetime of three prescriptions (296/100) seems plausible. Thus, it appears that the average physician is detailed at close to an optimal level. However, the dispersion in the distribution of the derivative is equally important. A full 25% of all physicians require fewer than .5 total lifetime prescriptions (50/100) to justify an additional detail, and the lowest 25% of physicians require at least 9.2 (917/100) lifetime prescriptions. This means that at least 50% of physicians are not detailed at optimal levels.

Further evidence on the optimality of detailing can be obtained by examining the distribution of the derivative by detailing decile. The bottom panel of Figure 6 shows that virtually all physicians in the highest decile of detailing have negative derivatives. We interpret this as evidence of overdetailing for this group of physicians.

MODEL VALIDATION

Modeling how the marketing-mix or independent variables depend on market response parameters provides two basic benefits: (1) removal of possible endogeneity biases and (2) more precise estimates of response parameters by exploiting the information in the marginal distribution of the marketing variable. However, some people would argue that the ultimate test of the full model is predictive validation. In this section, we compare the conditional NBD hierarchical model with the full hierarchical model in both time-series and cross-sectional validation experiments.

The most straightforward validation exercise is use of our estimated physician-level coefficients to predict the future values of new prescriptions (y). To implement the time-series validation, the last 2 months are reserved for each of the 1000 physicians. We reestimated the conditional model (Equation 9) and full model (Equations 9 and 10) on the first 21 months of data. We then used the posterior means to predict the values of y for the last 2 months of data, using the given level of detailing observed. The full model has a mean square error of 38.89, and the conditional model has a 12% greater value of 43.41.

A more strenuous test of the model performance is use of the model to predict prescription behavior in a new sample of physicians that we have not previously studied. We drew our estimation sample of 1000 physicians randomly from a much larger population of physicians. Drawing a new sample without replacement enables us to test the model on data never used in model specification searches or estimation. However, the cross-sectional validation exercise presents some notable methodological challenges. The cross-sectional validation problem is to predict values of y_{jt}, t = 1, ..., T, for physician j in this holdout sample with the information in x_{jt}, t = 1, …, T, and z_j. For the standard conditional or regression style model, this is straightforward. We simply predict the values of the response coefficients using the z vector and use these to make predictions. We use the information in x to form predicted values. However, in our full model, the level or mean of x has information about the response parameters. We must develop a way to incorporate this information in the estimates of the response parameters used in predictive validation.

Given the mean of x for the holdout physician j and knowledge of the gamma parameters, this equation specifies that the distribution of β is not the usual normal distribution given by the hierarchical model, because it is restricted to lie in a linear subspace. To determine this, we rearranged the terms in Equation 8 to obtain the following:

$$\gamma_1\beta_0 + \gamma_2\beta_1 + [\ln(\eta_j) - \gamma_0]\beta_2 = \ln(\eta_j) - \gamma_0. \tag{14}$$

Equation 14 is of the form $k'\beta = c$, where $k' = [\gamma_0, \gamma_2, \ln(\eta_j) - \gamma_0]$, and $c = \ln(\eta_j) - \gamma_0$.

With no information on the level of x_j, the distribution of β_j is given by the hierarchical model $\beta_j \sim N(\Delta z_j, V_\beta)$. Information on the mean or level of x_j implies that the restriction $k'\beta = c$ holds. Therefore, we must compute $\beta_j|E[x_j]$ or $\beta_j|k'\beta = c$, where the k vector and c are as we provided previously. To compute this distribution, we consider the conditional distribution of (β_1, β_2) given $k'\beta = c$ by using the appropriate linear transformation and normal distribution theory.

We define $\beta' = (\beta_a', \beta_b')$ and $\theta = T\beta$, where T is constructed as follows:

$$T = \begin{bmatrix} k' & 0' \\ 0 & I_{\dim(\beta)-1} \end{bmatrix}. \tag{15}$$

In addition, $\theta' = (\theta_a', \theta_b')$, $\theta_a = k'\beta$, and $\theta_b = \beta_b$, where $\beta_a = \beta_0$ and $\beta_b' = (\beta_1, \beta_2)$. The solution to the problem of computing $\beta_j|k'\beta = c$ is to find $\theta_b|\theta_a = c$, which can be computed from standard normal theory.

$$\theta_b|\theta_a = c \sim N\left[\mu_{\theta_b} - V_{ba}V_{aa}^{-1}\left(c - \mu_{\theta_a}\right), V_{bb}\right], \tag{16}$$

where $\mu_\theta = T\mu_\beta$, and $V_\theta = TV_\beta T'$.

To implement this approach, we use the posterior mean of γ and the sample mean of x for each of the holdout physicians. We then compute the expectation of (β_{1j}, β_{2j}) using Equation 16. We compare the estimates with estimates from the conditional model that only use the information in z_j. We find an improvement in the mean square error of 4% (29.69 versus 30.78).

Both the time-series and the cross-sectional validation exercises provide evidence that our full modeling approach improves prediction. Given the mature and competitive nature of this category, the relatively weak detailing effects in the data, and the tremendous variation in new prescription counts, we find that the magnitude of the improvements in prediction is significant.

A GENERAL FRAMEWORK

In the model we have presented, we use distributional assumptions that are appropriate for the count data. This does not limit the applicability of our approach. Our approach is a general one that can be applied to many settings in which managers strategically choose the marketing-mix or x variables. The basic contribution is to provide a framework for situations in which the marketing-mix variables are chosen with some knowledge of the response parameters of the sales response equation. This applies generically to many marketing-mix situations (see Gönül, Kim, and Shi 2000). In this section, we develop a general framework and discuss how other contributions in the literature can be perceived as special cases of this general framework.

Sales response models can be viewed as particular specifications of the conditional distribution of sales (y) given the marketing-mix x.

$$y_{it}|x_{it}, \beta_{it}, \tag{17}$$

where i represents the individual customer/account and t represents the time index. For example, a standard model would be to use the log of sales or the logit of market share and specify a linear regression model: $\ln(y_{it}) = x_{it}'\beta_i + \varepsilon_{it}$, $\varepsilon_{it} \sim$ Normal. Here the transform of y is specified as conditionally normal with sales response parameters β_i. Analysis of Equation 17 is usually conducted under the assumption that the marginal distribution of x is independent of the conditional distribution in Equation 17. In this case, the marginal distribution of x provides no information about β_i and the likelihood factors. If $x_{it}|\theta$ is the marginal distribution of x, the likelihood factors are as follows:

$$\begin{aligned} \ell(\{\beta_i\}, \theta) &= \prod_{i,t} p(y_{it}|x_{it}, \beta_i)p(x_{it}|\theta) \\ &= \prod_{i,t} p(y_{it}|x_{it}, \beta_i)\prod_{i,t} p(x_{it}|\theta). \end{aligned} \tag{18}$$

This likelihood factorization does not occur when the model is changed to build dependence between the marginal distribution of x and the conditional distribution. There are many possible forms of dependence, but in the context of sales response modeling with marketing-mix variables, a particularly useful form is to make the marginal distribution of x depend on the response parameters in the conditional model. Thus, we summarize our general approach in Equation 19:

$$y_{it}|x_{it}, \beta_i, \text{ and } x_{it}|\beta_i, \tau. \tag{19}$$

Equation 19 is a generalization of the models that Chamberlain (1980, 1984) developed and Bronnenberg and Mahajan (2001) applied in a marketing context. Chamberlain considers situations in which the x variables are correlated to random intercepts in a variety of standard linear and logit/probit models. Our random effects apply to all the response model parameters and we can handle nonstandard and nonlinear models. However, the basic results of Chamberlain's model with respect to consistency of the conditional modeling approach apply. Unless T increases, any likelihood-based estimator for the *conditional* model will be inconsistent. The severity of this asymptotic bias depends on the model, data, and T. For a small T, the biases have been documented to be large. What is not well appreciated is that the additional structure introduced by the model for the marginal distribution of x provides more information about the response parameters than does the conditional approach. That is, the levels of x are useful in making inferences about the β_i parameters.

The general data-augmentation and Metropolis–Hastings MCMC approach is ideally suited to exploit the conditional structure of Equation 19. That is, we can alternate between draws of $\beta_i|\tau$ (we recognize that the $\{\beta_i\}$ are independent conditional on τ and on $\tau|\{\beta_i\}$). With some care in the choice of the proposal density, the MCMC approach can handle a wide range of specific distributional models for both the conditional and the marginal distributions in Equation 19.

To specify the model in Equation 19 further, it is useful to consider the interpretation of the parameters in the β vector. We might postulate that in the marketing-mix application, the important quantities are the level of sales given some "normal" settings of x (e.g., baseline sales) and the derivative of sales with respect to various marketing-mix variables. In many situations, decision makers set marketing-mix variables proportional to the baseline level of sales. More sophisticated decision makers might recog-

nize that the effectiveness of the marketing mix is important in allocation. This means that the specification of the marginal distribution of x should make the level of x a function of the baseline level of sales and the derivatives of sales with respect to the elements of x.

The notion of dependence between the marginal distribution of x and the conditional distribution of y|x is more general than the specific model proposed in Equation 19. This dependence is sometimes described in the econometrics literature as endogeneity. In the linear model, a standard assumption is the independence between the regression error and the independent variables. Some of the price endogeneity literature (e.g., Villas-Boas and Winer 1999) suggests that a common demand shock that retailers use to set price creates a form of dependence between x and the error term. Other streams in the literature provide an omitted-variables interpretation of the dependence problem (e.g., Berry, Levinsohn, and Pakes 1995). These forms of dependence are different from that postulated in Equation 19. In addition, the methods of inference we have used are fully Bayesian and do not rely on the existence of valid instruments and asymptotic approximation. A valid instrument for our problem would need to be physician specific and correlated with detailing but not correlated with the response parameters. Moreover, we do not believe that it is appropriate to base physician-level estimates on procedures that use asymptotic approximations in a situation with only 24 observations per physician.

SUMMARY AND CONCLUSION

Both academics and practitioners build sales response models in which sales measures are related to marketing-mix variables. It is common to assume that the marginal distribution of the marketing-mix variables is independent of parameters of the sales response model. This assumption is likely to be violated in situations in which marketing managers set the levels of marketing-mix variables with at least some partial knowledge of sales response parameters. In this case, conditional modeling can produce biased response coefficients, and it fails to exploit a potentially valuable source of information in the levels of the marketing-mix variables. We develop an approach that jointly models the distribution of both sales response and marketing-mix variables. We apply this approach to physician-level data on prescription writing (sales) as a function of the level of sales force effort (as measured by sales calls, or details). We recognize the count or discrete aspect of the panel data and the model prescriptions as an NBD regression that is conditional on detailing. The marginal distribution of detailing is Poisson with the mean dependent on the response parameters in the conditional NBD regression. We estimate both the conditional NBD regression and the Poisson model of detailing in a joint hierarchical Bayes procedure.

Our joint procedure produces individual estimates of the effectiveness of detailing that are markedly different from those obtained from a more standard conditional approach. We can dramatically improve the precision of physician-level estimates through the additional model structure that exploits the fact that the level of detailing also has information about the magnitude of response parameters. The improvements result in an improvement in predictive power. This 4%–12% improvement in predictive validation is specific to this product category and data set. In drug categories in which detailing is more important (e.g., a growth category, in contrast to our category, in which the leading drugs are near the end of their patent life), we expect that there is much better predictive performance in the conditional model than in the full model. Another advantage of our approach is that we do not assume that the firm is optimally allocating detailing resources.[5] Although physician-level data are available to most pharmaceutical firms, the modeling technology and skill sets for individual physician level estimation are not generally accessible, because these developments have appeared only recently in the academic literature.

We infer from the data how the firm sets the level of detailing. We find that the firm details physicians with higher baseline volume more than physicians with lower baseline volume, which is consistent with our discussions with firm managers and consultants. However, we also find that there is an independent effect in which physicians who are more responsive to detailing are actually detailed less, on average, than are less responsive physicians. In particular, physicians who experience the highest level of detailing have negative estimated effects of marginal details. An interpretation of this finding is that if all details are identical, there is considerable overdetailing, at least among the top 20% of detailed physicians. Another possible interpretation is that the highly detailed doctors are also detailed intensively by other competing drug companies, thus lowering the responsiveness to the average detail. Our data do not contain information on competitive detailing, so it is not possible to make a definite conclusion. We note that managers in the pharmaceutical industry do not have access to data on competitive detailing data in making sales force allocation decisions. Because our goal was to investigate the extent to which managers fully exploit the information content in the data, it is appropriate in our analysis to use the same data to which managers have access.

The general problem of nonrandomly chosen marketing-mix variables applies in many contexts. We hope that our approach and results encourage further research in this area.

APPENDIX: INFERENCE

We derive the full-conditional distributions of the unknowns using the joint density and the specified prior distributions. To obtain the posterior distribution of the unknowns, we then draw sequentially from this series of full-conditional distributions until we achieve convergence.

Draw 1

The $\{\beta_i\}$ are distributed $N(\Delta^*Z_i, V_\beta)$ (Equation 6). Thus, the full-conditional distribution for β_i is given as follows:

$$(A1)\qquad p(\beta_i|\Delta, V_\beta, \gamma, \alpha, y_{it}, DET_{it}, Z_i) \propto \ell(\beta_i) \times \exp\left[(\beta_i - \Delta \times Z_i) \times V_\beta^{-1} \times (\beta_i - \Delta \times Z_i)'\right],$$

where $\ell(\beta_i) = \Pi_t p_{NBD}(y_{it}|Det_{it}, \beta_i, \alpha) p_{Poisson}(Det_{it}|\beta_i, \gamma)$. The full-conditional distribution for β_i is known only up to a

[5]Industry evidence points to inefficiencies in current industry practices with respect to sales force allocation (see Elling et al. 2002).

proportionality constant. We use the random-walk Metropolis–Hastings algorithm to generate a candidate on iteration n as $\beta_i^c = \beta_i^{(n-1)} + \tilde{\beta}_i$, where $\tilde{\beta}_i$ is a draw from a multivariate normal proposal density, N(0, kΨ). We set Ψ equal to the asymptotic variance–covariance matrix of the β parameters estimated on pooled data (i.e., assuming that there are no physician-level differences) using maximum likelihood estimation; k is a scalar that we chose to achieve a reasonable acceptance rate. The acceptance probability is given as follows:

$$\text{(A2)} \qquad \min\left\{\frac{p(\beta_i^c|\Delta, V_\beta, \gamma, \alpha, y_{it}, DET_{it})}{p(\beta_i^{(n-1)}|\Delta, V_\beta, \gamma, \alpha, y_{it}, DET_{it})}, 1\right\},$$

where $p(\cdot|\cdot)$ is as given previously.

Draw 2

The full-conditional distribution for Δ is given as follows:

$$\text{(A3)} \qquad p(\Delta|\{\beta_i\}, V_\beta, \Delta_0, V_\Delta, Z_i) = N(\hat{\Delta}, \hat{V}_\Delta),$$

where $\hat{\Delta} = \hat{V}_\Delta(V_\Delta^{-1} \times \Delta_0 + \Sigma_{i=1}^{I} Z_i' \times V_\beta^{-1} \times \beta_i)$, and $\hat{V}_\Delta = (V_\Delta^{-1} + \Sigma_{i=1}^{I} Z_i' \times V_\beta^{-1} \times Z_i)$. We set $\Delta_0 = 0$, and $V_\Delta = \text{diag}(1000)$.

Draw 3

The full-conditional distribution for V_β^{-1} is given as follows:

$$\text{(A4)} \qquad p\left(V_\beta^{-1}|\{\beta_i\}, \Delta_0, \rho, R, Z_i\right)$$

$$= \text{Wishart}\left\{\left[\rho \times R + \sum_{i=1}^{I}(\beta_i - \Delta \times Z_i)(\beta_i - \Delta \times Z_i)'\right]^{-1}, \rho + I\right\},$$

where I is the number of physicians. We set the prior mean of $V_\beta = (\rho R)^{-1} = \text{diag}(10)$, and the prior degrees of freedom are ρ = NPAR + 3, where NPAR is the dimension of the β vector.

Draw 4

The prior distribution of α is specified as Gamma(a, b) with mean (a/b) and variance (a/b^2). Thus, the full-conditional distribution for α is given as follows:

$$\text{(A5)} \qquad p(\alpha|\{\beta_i\}, y_{it}, DET_{it}) \propto \ell(\alpha) = \left\{\prod_i \prod_t p_{NBD}(y_{it}|Det_{it}, \beta_i, \alpha)\right\} \times \alpha^{(a-1)} \times \exp(-b\alpha).$$

The full-conditional distribution for α is known only up to a proportionality constant. We use the random-walk Metropolis–Hastings algorithm to generate a candidate on iteration n as $\ln(\alpha^c) = \ln(\alpha^{(n-1)}) + \tilde{\alpha}$, where α is a draw from a univariate normal proposal density, N(0, σ^2). We choose a value of σ^2 such that the acceptance rates are reasonable. The acceptance probability is given by

$$\text{(A6)} \qquad \min\left\{\frac{p(\alpha^c|\{\beta_i\}, y_{it}, DET_{it})}{p(\alpha^{(n-1)}|\{\beta_i\}, y_{it}, DET_{it})}, 1\right\},$$

where $p(\cdot|\cdot)$ is as given previously. We set a = .5 and b = .1.

Draw 5

The full-conditional distribution for γ is given as follows:

$$\text{(A7)} \qquad p(\gamma|\{\beta_i\}, \gamma_0, V_\gamma, DET_{it}) \propto \ell(\gamma) \times \exp\left[(\gamma - \gamma_0) \times V_\gamma^{-1} \times (\gamma - \gamma_0)'\right],$$

where $\ell(\gamma) = \Pi_i \Pi_t [\exp(\beta_i \times \gamma)]^{DET_{it}} \times \exp[-\exp(\beta_i \times \gamma)] \times (1/DET_{it}!)$. The full-conditional distribution for γ is known only up to a proportionality constant. We use the random-walk Metropolis–Hastings algorithm to generate a candidate on iteration n as $\gamma^c = \gamma^{(n-1)} + \tilde{\gamma}$, where $\tilde{\gamma}$ is a draw from a multivariate normal proposal density N(0, mΦ). We set Φ equal to the asymptotic variance–covariance matrix of the γ parameters estimated using a two-stage model with maximum likelihood estimation (i.e., we fit a Poisson regression model using DET_{it} as the dependent variables and the β_i from the conditional NBD model as the regressors); m is a scalar that we chose to achieve a reasonable acceptance rate. The acceptance probability is given by

$$\text{(A8)} \qquad \min\left\{\frac{p(\gamma^c|\{\beta_i\}, \gamma_0, V_\gamma, DET_{it})}{p(\gamma^{(n-1)}|\{\beta_i\}, \gamma_0, V_\gamma, DET_{it})}, 1\right\},$$

where $p(\cdot|\cdot)$ is as given previously. We set $\gamma_0 = (.63, 0, 0)$, and $V_\gamma = \text{diag}(25, 25, 100)$.

REFERENCES

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841–90.

Bronnenberg, Bart and Vijay Mahajan (2001), "Unobserved Retailer Behavior in Multimarket Data: Joint Spatial Dependence in Market Shares and Promotion Variables," *Marketing Science*, 20 (3), 284–99.

Chamberlain, Gary (1980), "Analysis of Covariance with Qualitative Data," *The Review of Economics Studies*, 47 (1), 225–38.

——— (1984), "Panel Data," in *Handbook of Econometrics*, Vol. 2, Z. Griliches and M.D. Intrilligator, eds. Amsterdam: North Holland, 1247–1318.

Clarke, Darral G. (1976), "Econometric Measurement of the Duration of Advertising Effect on Sales," *Journal of Marketing Research*, 13 (4), 345–57.

Elling, Martin E., Holly Fogle, Charles S. McKhann, and Chris Simon (2002), "Making More of Pharma's Sales Force," *The McKinsey Quarterly*, 3, 87–95.

Gönül, Füsun, Franklin Carter, Elina Petrova, and K. Srinivasan (2001), "Promotion of Prescription Drugs and Its Impact on Physicians' Choice Behavior," *Journal of Marketing*, 65 (July), 79–90.

———, B. Kim, and M. Shi (2000), "Mailing Smarter to Catalog Customers," *Journal of Interactive Marketing*, 14 (3), 2–16.

Manchanda, Puneet and Pradeep K. Chintagunta (2000), "Responsiveness of Physician Prescription Behavior to Salesforce Effort: An Individual Level Analysis," working paper, Graduate School of Business, University of Chicago.

Neslin, Scott (2001), "Analysis of ROI for Pharmaceutical Promotions," unpublished study conducted for the Association of Medical Publications, [available at http://www.rxpromoroi.org/rapp/index.html].

Nevo, Aviv (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, 69 (2), 307–342.

Villas-Boas, M. and Russell Winer (1999), "Endogeneity in Brand Choice Models," *Management Science*, 45 (10), 1324–38.

Wittink, Dick R. (2002), "Analysis of ROI for Pharmaceutical Promotions," unpublished study conducted for the Association of Medical Publications, [available at http://www.rxpromoroi.org/rapp/index.html].