INCORPORATING COMPETITOR DATA INTO CUSTOMER RELATIONSHIP MANAGEMENT

A Dissertation

Presented to

the Faculty of the C.T. Bauer College of Business

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

by

Doug Walker

May, 2008

INCORPORATING COMPETITOR DATA INTO CUSTOMER RELATIONSHIP MANAGEMENT

APPROVED:

James D. Hess, Bauer Professor of Marketing Science
Chairperson of Committee

Michael Ahearne, Associate Professor of Marketing

Niladri B. Syam, Associate Professor of Marketing

Christian J. Murray, Associate Professor of Economics

Arthur D. Warga, Dean
C. T. Bauer College of Business

ACKNOWLEDGEMENTS

My sincerest thanks to my dissertation committee chairman, Jim Hess; the members of my committee, Mike Ahearne, Niladri Syam, and Chris Murray; and the rest of the faculty, students and staff - none of this would have been possible without you.

To my wife, Kim, and my daughters, Jessica and Rachael, thank you for your love, encouragement, patience, sacrifice, and prayers. I love you with all of my heart.

INCORPORATING COMPETITOR DATA INTO CUSTOMER RELATIONSHIP MANAGEMENT

Abstract

Fueled by technological innovation, customer relationship management (CRM) research and practices have been driven primarily by the exponential growth in customer transaction data held by firms. Consideration of the competition has largely been lost in this flood of firm-focused data. The practice of CRM seems to have strayed from its market orientation roots.

Academic leaders in the field of CRM have called for research incorporating competitor data. Researchers are beginning to answer that call. Few would argue that the availability of competitor data enhances CRM decision making, including the allocation of marketing effort. However, since in most contexts competitor data is difficult and expensive to acquire, how important is it to the firm? This study shows that in a pharmaceutical context the firm's marketing effort allocation decisions would fundamentally change based on the availability of competitor data to be used in the analysis.

Specifically, when the firm does not consider the competitions' marketing efforts and customers' perceptions of the competing brands, the estimates of response to the firm's marketing efforts are biased for a sizeable minority of the firm's customers, leading to a misallocation of the firm's resources. Since this type of data is typically available for only a portion of the firm's customers, it must be imputed for the rest of the customers in the database. A data augmentation method that imputes a composite of the data collected via a survey for customers that did not participate in the survey is presented. Results using this method outperforms a model using firm data only and a model using firm data and survey data on the perceived characteristics of each brand, even if the perceived drug characteristics are known for all of the customers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Customer relationship management (CRM) has naturally evolved within firms that embrace the concept of a market orientation, where the firm is focused on generating customer-focused market intelligence, disseminating that intelligence, and responding to it (Kohli and Jaworski 1990). The emphasis firms place on analytical CRM, which utilizes customer databases, has exploded in recent years as improving technology allows firms to collect, store, and analyze customer data ever more efficiently and less expensively than ever before.

The evolution of CRM has been driven by the customer data firms have chosen to use to guide marketing activities. Early segmentation efforts focused primarily on demographic differences among the firm's customers. Firms made the implicit assumption that customers that are similar demographically will respond to a particular marketing appeal in a similar manner (see Kotler and Armstrong 1994). Next, firms began to consider transactional data in addition to customer demographics to inform marketing effort decisions. Both researchers and practitioners became interested in the recency, frequency, and monetary value (RFM) of a customer's purchase history (see Drozdenko and Drake 2002). Appreciation of an estimated lifetime value of a customer (LTV), or customer lifetime value (CLV), gained in prominence (e.g. Berger and Nasr 1998).

Interest in CRM exploded as improving technology allowed the creation of extensive customer databases, documenting not only a customer's purchases, but the marketing efforts directed at the customer as well. In fact, firms were now able to capture virtually all of the interactions between the customer and the firm, regardless of which party initiated the contact. Statistically based methods, such as latent class modeling (Wedel and Kamakura 2000) and concomitant variable methods, where segments defined by transactional variables can be described using demographic variables (e.g. Gupta and Chintagunta 1994), generated interest.

Although CRM has its roots in market orientation, up until now, a key component of market orientation, the competition, had been largely ignored (Boulding et al. 2005). Kohli and Jaworski (1990) emphasize the consideration of key exogenous factors, including the competition, during intelligence generation in a market-oriented firm. Narver and Slater (1990) concur, identifying competitor orientation as a basic component of market orientation. The primary reason the competition had been widely ignored then was the same as it is today, data availability. Firms did not have easy access to data detailing customer interaction with the competition, as they did for their own interactions with the customer. However, firms in one industry, pharmaceuticals, do enjoy access to sales data for all brands of ethical drugs at the individual physician level. (This level of data accessibility is rare outside of the U.S.) Now, the firm cannot only consider the purchase history of each customer and the marketing effort directed at each customer, but the “size” of the customer in terms of their total category demand within a particular category. Although the response models had become more comprehensive with the inclusion of competitor sales data, an important limitation still remained. The marketing effort of the competition was still being largely ignored, introducing an omitted variable problem that could potentially impact the estimates of the response parameters (Manchanda and Chintagunta 2004).

Gonul et al. (2001) and Venkataraman and Stremersch (2007) utilize datasets containing brand level data on competitor marketing effort for small panels of physicians. Venkataraman and Stremersch (2007) also consider the effectiveness and side effects of the brands in the category. However, these are composite measures based on clinical trials and labeling, not as perceived by the physicians. Moon et al. (2007) incorporate unobserved competitor marketing effort into their analysis via a hidden Markov model. We are unaware of any study to date that focuses on the magnitude of the bias in response to the firm's marketing efforts resulting from omitting competitor marketing efforts and physician perceptions of drug characteristics. Equally important, the studies that did incorporate competitor data did not contemplate that this data was available for only a portion of the firm's customers. Augmenting the database to include all of the firm's customers will also be addressed in this study.

Ultimately, physician response can more completely be portrayed as a function of customer demographics, firm sales, firm marketing effort, competitor sales, competitor marketing effort, and physician perceptions of the drugs. This study is unique in that competitor data is considered not for only a small sample of the firm's customers, but for all of the firm's customers via a survey and data augmentation. Additionally, physician perceptions of drug characteristics are examined, again via a survey and data augmentation. Bias in the estimated response to the firm's marketing efforts when this data is ignored will be carefully investigated. Specifically, the analysis will attempt to determine whether the firm's marketing allocation decisions would be fundamentally different based on the availability of competitor data.

2. LITERATURE REVIEW

In this section, we will consider several research streams that are relevant to this research. First, we will discuss the papers in the vast CRM literature that incorporate competitor data. Second, since the context of this study involves the marketing of a category of ethical drugs, we will review relevant papers in the pharmaceutical sales literature. Third, since the exclusion of competitor data when estimating response can be thought of as a missing data issue, we will look at topics in that literature relevant to this study. Finally, we will discuss applicable data augmentation methods.

Competitor Data in CRM

CRM researchers are not ambivalent to the importance of considering the competition when making CRM decisions. Numerous studies, concentrating only on firm-specific data, demonstrate CRM can enhance firm profits, at least in the short run (e.g. Cao and Gruca 2005; Ryals 2005). However, Boulding and colleagues (2005, pg. 161) state that "a failure to integrate competition into a firm's CRM activities potentially puts it at serious risk." Bell and his co-authors (2002) concur, emphasizing that the learning gained from examining a firm's own customers is incomplete without considering prospective customers. In a pharmaceutical context, Manchanda et al. (2005) consider the lack of competitor detailing data to be a "major issue". These comments seem relevant to shared customers where the firm enjoys varying shares of those customers' total category requirements.

The firm's share-of-wallet for each customer is one competitor-oriented measure that has received some attention from CRM researchers. Researchers have conceptualized that knowing the firm's share-of-wallet can be of value in segmenting a firm's customers (e.g. Reinartz and Kumar 2003). The basic premise, which is quite intuitive, is that the firm should focus on customers with substantial category demand, but of which the firm has a small share (Anderson and Narus 2003). There is some empirical support for this approach (Reinartz, Thomas, and Kumar 2005).

Share-of-wallet has commonly been conceptualized as a measure of customer loyalty and used as a proxy for competitor effort (e.g. Bowman and Narayandas 2004; Reinartz et al. 2005). Share-of-wallet has been found to positively impact customer profitability (Reinartz et al. 2005) and has been theorized to mediate the effect of customer retention on profits (Zeithaml 1985).

Several papers have taken the findings that share-of-category requirements are predictive of customer profitability as incentive to devise methods to estimate the share-of-wallet for a firm's customers. The underlying assumption, of course, is that knowing this information will result in better informed CRM decisions. Bhattacharya et al. (1996) looked at the relationship between share-of-category requirements and the marketing mix. They found a small but significant relationship, but cautioned against making causal claims. Du, Kamakura and Mela (2007) prescribe a larger investment in large category-demand, low category-share customers, and propose a database augmentation method that estimates share-of-wallet.

Pharmaceutical Sales

This study incorporates sales effort in the form of detailing, but does not investigate salespeople. In fact, the analysis focuses on the customers. In the context of ethical drug sales, the customers are the physicians. Although a review of sales research in general is not appropriate, a summary of the pharmaceutical sales literature will be of value in presenting the context for this study.

Gonul et al. (2001) utilize a physician-level database that includes prescription writing, detailing, and sampling by brand for a small panel of physicians. Their response model accommodates physician heterogeneity over three latent classes via the intercept term, but assumes the impact of detailing and sampling on prescription writing is constant across brands and physicians. The multinomial logit model does not allow for consideration of persistence in physician prescription writing behavior over time. The public-policy motivated findings suggest detailing and sampling serve primarily an informative role.

Using a fixed-effects model, physician-specific effects are considered by Mizik and Jacobson (2004). The authors also include lagged prescriptions to allow for physician preferences to persist over time. Competitor marketing effort is excluded from their database. Their analysis shows that detailing and sampling do impact prescription behavior, although the effects are small.

Using Bayesian methods, Manchanda and Chintagunta (2004) are able to investigate physicians’ response to detailing at the individual physician level. They focus on the total number of prescriptions written in a particular drug category and find that detailing does positively influence the number of prescriptions written, although, as expected, at a decreasing marginal rate. The marketing efforts of the competition are not considered. A discussion of the potential benefits of reallocating details is included in the study.

The potential endogeneity inherent in a pharmaceutical sales response model is analyzed by Manchanda, Rossi and Chintagunta (2004). They model the number of prescriptions written in the category as a function of detailing, but then make detailing dependent on the parameters of the response function. They report that accounting for reverse causality results in better model fit. Substantive findings include an apparent over-detailing of high volume physicians. The authors suggest that their results may be due to the effects of latent competitor sales efforts. In their study competitor effort is unaccounted for, although it may actually be partially controlled for implicitly, since the individual specific intercepts represent unobserved heterogeneity in Bayesian analysis.

Venkataraman and Stremersch (2007) consider the interaction between the characteristics of each drug in a category and the marketing efforts exerted for each of those drugs. Specifically, the authors incorporate a measure of each brand's effectiveness and the corresponding side effects. Effectiveness and side effects are not measured based on the perceptions of each physician, but rather they are summary statistics derived from a meta-analysis of clinical trials and drug labeling, respectively. Generally, their results suggest effective drugs with few side effects benefit more from marketing effort.

Missing Data

Customer databases for most firms consist primarily of firm-specific data. In other words, competitor data related to the customers in the database are missing. Imagine a rectangular customer database with customers on the rows and variables relating to those customers on the columns. The missing data literature deals primarily with situations where *some* of the values in any particular column are missing. If a firm has firm-specific data, but no competitor data, entire columns of data could be considered to be "missing", not just some of the values in the columns. Little can be done to impute the missing values when this is the case. However, since enhancement of the customer database for some customers via a survey is part of this research, the projection of values for variables collected in the survey for those customers not included in the survey involves methods used to address missing data. Fortunately, assuming the participants in the survey are randomly selected, the mechanism that produced the missing data is the easiest to address. Even so, an understanding of the key issues in missing data analysis is appropriate.

Little and Rubin (2002) discuss the importance of discovering the mechanism that leads to missing data, since the mechanism determines the appropriate methodological response. The authors list three missing data mechanisms, with the key issue being if the actual value of the missing data is the reason it is missing.

Using their notation, consider a complete rectangular data set Y, with each element in the dataset represented as y_{ij}, where i is the row and j is the column. Also, consider a matrix M of the same dimensions, where the value for element m_{ij} is 1 if the value is observed and 0 if it is missing. Data are called missing completely at random (MCAR) if the conditional distribution of M is dependent only on some unobserved parameters, ϕ, but not on the values of the data Y, expressed as

$$f\left(M \mid Y,\phi\right)= f\left(M \mid \phi\right) \text{ for all } Y,\phi . \tag{1}$$

If the observed elements in Y are labeled Y_{obs} and the missing elements are labeled Y_{mis}, data are considered to be missing at random (MAR) if

$$f(M \mid Y, \phi) = f(M \mid Y_{obs}, \phi) \text{ for all } Y_{\text{mis}}, \phi, \tag{2}$$

indicating that the reason the data are missing is dependent only on the values of the elements in Y that are observed. Finally, missing elements in Y are not missing at random (NMAR) if the distribution of the matrix M is dependent on the actual values in Y that are missing.

Missing data enters this study in two ways. First, for comparison purposes, the typical scenario where the firm has no competitive data for any of their customers will be considered. Second, survey data will be missing for some physicians due to non-response. If the mechanism behind any non-response is unrelated to the values that would have been entered on the survey if completed, but rather is due to observed values in the existing database, the missing data mechanism will be MAR. The non-response will be categorized as NMAR if the non-response is dependent on values related to the survey items. MCAR is the simplest missing data mechanism to address, while NCAR is the most difficult.

Database Augmentation

Database augmentation techniques are firmly entrenched in the missing data literature. In fact, augmentation is a special case of imputation, a common technique for handling missing data (Kamakura and Wedel 2003). Typically, a firm will conduct a survey or purchase data for a random sample of their customer database. This data, along with data already existing for customers included in the survey, will be analyzed to discover relationships between the survey data and the internal data. Predictive models based on these relationships will then be developed to estimate the values for the surveyed variables for those customers not included in the survey. The objective is to leverage the survey data in such a way that informs decision making concerning all of the customers in the database.

Du et al. (2007) demonstrate a database augmentation method in a banking context. They use survey data on share-of-wallet for a variety of banking product categories, along with internal data on customers' income and tenure, to estimate share-of-wallet for customers excluded from the survey. Their method simultaneously imputes whether the customer has an external balance in a category, and then if they do, the size of the external balance. Their method does not consider competitor brand shares.

Sub-sampling, whether in the form of surveys or test markets, creates a need for database augmentation. Methods for imputing missing values can be as simple as entering the mean level for the observed values for a variable where the value is unobserved. At the other extreme, sophisticated methods designed to account for large proportions of missing data and a variety of measurement scales for the missing values have been developed (e.g. Kamakura and Wedel 2003).

All of these methods involve imputing values based on models utilizing the information found in not only the survey data but the existing internal data as well. Little and Rubin (2002) give three criteria to guide imputations. First, the imputation should be conditioned on observed variables. Second, when possible, the procedure should be multivariate to preserve correlations between missing variables. Third, imputed values should be drawn from a predictive distribution rather than just imputing means to avoid overstating a central tendency. Additionally, the authors encourage the use of multiple imputation, as opposed to single imputation, to account for imputation uncertainty. Multiple imputation involves drawing a series of complete datasets for analysis, with parameter estimates being the mean from each of the analyses and the standard errors including imputation uncertainty.

3. OMITTED VARIABLES

The theoretical foundation of this study rests on the premise that excluding variables from a model will bias the parameter estimates related to the variables that are included in the model. However, two conditions must be met for omitted variable bias to exist. Assume the following regression model, with subscripts suppressed,

$$Y = \alpha + \beta W + \gamma X + \delta Z + \varepsilon. \tag{3}$$

If the variable Z is excluded from the model, all of the parameters in the model could be biased as long as δ is not equal to zero and at least one of the variables remaining in the model, W or X, is correlated with Z. The extent and direction of the bias cannot be determined as long as there are two or more variables remaining in the model, but the magnitude of the bias is a function of the size of δ and the degree to which Z is correlated with the variables remaining in the model.

Relevance of Excluded Variables

The first condition that must be met for bias to exist concerns the relevance of the omitted variables in the model. Both conceptual and empirical research has solidified the importance of including competitor variables in models where the impact of a firm's marketing efforts on customer response is being investigated. Researchers looking at market orientation, the foundation of CRM (e.g. Kohli and Jaworski 1990; Narver and Slater 1990), and CRM researchers in marketing (e.g. Bell et al. 2002; Boulding et al. 2005) have all emphasized the importance of considering the competition when making marketing allocation decisions. Empirical researchers analyzing physician response that have had access to panel data including competitor marketing effort, have found those competitor variables to be significant in modeling response to the firm's marketing efforts (Gonul et al. 2001; Venkataraman and Stremersch 2007).

Correlation Between Excluded Variables and Variables Remaining in Model

In addition to the omitted variables being relevant, they must also be correlated with at least one variable remaining in the model for bias to exist. In the pharmaceutical context, Manchanda and Chintagunta (2004) speculate that the addition of competitor detailing in their model may fundamentally change their findings. Mizik and Jacobson (2004) argue that the exclusion of competitor detailing in their model does not create bias in their response parameters because they speculate that the correlations between detailing among firms is very low. They back this assertion by looking at the correlations between detailing levels for all brands in a dataset and category external to their study and find them to be low. This argument could be misleading in two ways. First, the correlations among detailing for the brands in a category may vary across categories. A correlational analysis of the detailing levels among the four brands in this study show relatively high degrees of correlation, as shown in Table 1. This is consistent with commonly reported practice in the pharmaceutical industry, along with conversations with the focal firm in this study, where baseline detailing levels are set based on the total category demand of each physician.

Table 1: Correlations Among Detailing Levels Across Brands

	Brand B	Brand C	Brand D
Focal Brand	0.39	0.42	0.41
Brand B		0.52	0.56
Brand C			0.64

Second, omitted competitor detailing does not necessarily have to be correlated with the included firm detailing variable for the parameter estimate for firm detailing to be biased. *All* parameter estimates in the model can potentially be biased if *any* variable included in the model is correlated with a relevant omitted variable (Wooldridge 2002). Even if the correlation between the focal firm's detailing and competitor detailing is low, competitor detailing would be expected to be correlated with a lagged dependent variable appearing in the model.

4. MODEL

Demand for Products in a Category

Each firm, of course, is interested in maximizing profits. Obviously, this maximization applies across all of the firm's products, but even with the available data pertaining to a single product in a particular category, a consideration of the firm's profit function is worthwhile.

A physician's total category demand for a particular class of drugs, over some defined time period, can be conceptualized as follows. Each physician has a limited, and generally fixed, number of appointment slots available to see patients. This number will be represented as n. Obviously, this number will vary across physicians for a variety of reasons. For example, the time spent with each patient, on average, may depend to some extent on whether or not the physician is employed by a health maintenance organization (HMO). A certain proportion, q, of each physician's patients will be diagnosed with a condition that could be treated with a drug from the category in question. Again, this proportion would be expected to vary by physician. For instance, a cardiologist would be expected to prescribe heart medication to a higher proportion of their patients than would a family practice doctor. Of those patients diagnosed with a particular condition, a physician would treat a certain percentage of them, h, with a drug from the category being considered. This percentage would likely vary across physicians due to several reasons, for example, years in practice.

Therefore, using the indicated notation presented above, the expected number of prescriptions for a particular drug category and physician would be the product of the number of patients seen in a period, the proportion of those with a condition potentially treatable by drugs in the category, and the percentage of those with the condition where drugs in the category are the best treatment option, or $n \times q \times h$. Obviously, these values could change over time. For example, if a physician is enjoying a growing practice, more patients will be seen and n will increase. Greater specialization over time in conditions treatable by drugs in the category would increase the proportion of patients seen that will be diagnosed with the relevant condition, so q will increase. Finally, positive experience with drugs in the category or evolving best practices could result in a greater percentage of those with the condition being treated with drugs in the category, increasing h. Each firm's marketing mix could certainly impact the total category demand for a category of drugs, primarily by increasing the percentage of patients diagnosed with the condition being treated with a drug from the category, represented by h. This impact would most likely be seen relatively early in the life cycle of the category. In a mature category, firms' marketing efforts would be less likely to alter a physician's total category demand, but rather would influence each brand's share of prescriptions for the physician. In the category being studied and over the time period of the data, total category demand both in aggregate and by physician are generally constant.

With this conceptualization of total category demand as a foundation, several aspects of the ethical drug market have led to reasonable simplifications in the profit function in previous research (e.g. Manchanda et al. 2004). First, the costs of producing an ethical drug are primarily sunk. In fact, the marginal costs are so small compared to the sales price that they are typically assumed to be zero. Second, expenditures on the sales force (detailing) dominate other marketing expenditures. In a representative drug category, 80% of total marketing expenditures pertain to detailing (Manchanda and Chintagunta 2004). The response to changes in price, typically a key variable in analyzing demand, is of much less importance in the ethical drug market. Price is only indirectly salient to the patient and far less important to the physician than the appropriateness of a particular drug for each patient. Therefore, in an ethical drug context, detailing is the critical variable. Third, although the cost of a detail can certainly vary from one visit to the next and over physicians, the cost will not be nearly as variable as say, for example, different advertising campaigns. Therefore, the marginal cost of a detail is typically assumed to be constant. The resulting simplified profit function for the firm is

$$\pi_j = rS_j - cD_j, \tag{4}$$

where j = physician, r = revenue from a prescription, S = number of prescriptions (or scripts), c = marginal cost of a detail, and D = number of details.

We assume the number of prescriptions written is some function of detailing. Since the total category demand in this category is essentially constant, we can concentrate on the impact of detailing on market share, rather than on brand demand. Ideally, once this functional relationship is specified, elasticities can be calculated, allowing for the determination of a superior allocation of details. Additional variables could certainly be added and a more sophisticated cost function could be applied, but regardless, it is evident that within this context, the main consideration is the impact of detailing on prescription share. Generally speaking, we are considering a linear model very loosely of the form

$$Sh_t = \alpha + \delta D + \gamma Sh_{t-1} + \varepsilon, \tag{5}$$

where Sh is the share of total category prescriptions, α represents the intercept, and δ is the impact of detailing on prescription share. Estimation of the parameter γ allows for persistence in prescription writing over time. In the general model, marketing effort, D in this case, and lagged share, Sh_{t-1}, will incorporate all brands in the category. Attempting to precisely specify this relationship will be the central modeling task in this research.

General Model

Previous research investigating the impact of marketing on prescription writing behavior has utilized a variety of modeling approaches. Three primary considerations guide the modeling choice in this research. First, the objective is to model market share. Linear models are immediately ruled out due to the restricted range of the dependent variable. Second, the estimation of the elasticity of market share relative to marketing variables is of primary concern. Multiplicative log-linear models with the natural log of market share as the dependent variable suffer from the problem of constant elasticity. For example, the elasticity of market share relative to a particular variable should approach zero as market share approaches one. Exponential log-linear models, again with the natural log of market share serving as the dependent variable, are similarly problematic. In addition, the exponential model implies that elasticity should increase indefinitely as the value of the variable increases. Finally, the relationship between the relevant independent variables and the resulting elasticity in market share must be consistent theoretically with how the variables are expected to impact market share. A random utility model, such as the multinomial logit, implies that the elasticity of market share increases as the independent variable increases from low levels, reaches a peak, and then declines (Cooper and Nakanishi 1988). Market share elasticity, considering the key variables in this study (specifically detailing), is expected to decline monotonically as the level of the variables increase, making the multinomial logit approach not well suited for this research.

The proposed model for the share of prescriptions written for drug i by physician j in period t begins with a general model for brand share, m_{ijt}. Kotler (1971) considers the well-known multiplicative competitive interaction (MCI) model to be the fundamental theorem of market share, represented as

$$m_{ijt} = \frac{M_{ijt}}{\sum_{k=1}^{K} M_{kjt}}, \tag{6}$$

where M_{ijt} is the marketing effort for drug i directed at physician j in period t and the denominator represents the combined marketing effort for all of the brands (Cooper and Nakanishi 1988). The MCI model is appropriate given the three considerations discussed earlier, providing a model for market share that allows for monotonically decreasing market share elasticity over the range of the independent variables. Although statistically equivalent, Cooper and Nakanishi (1988) describe how marketing effort in the MCI model can alternatively be conceptualized as the attraction consumers feel for each particular brand. In this paper, relative marketing effort is of primary concern. However, physician perceptions of each drug's characteristics will also be included in the model, along with lagged share to account for persistence in prescription writing behavior. Therefore, the attraction conceptualization is more appropriate in this research.

Typically, the MCI model is specified using a multiplicative function of relevant variables. Suppressing all but the subscript for brand, i, marketing effort can be expressed as

$$M_i = \exp(\alpha_i + \varepsilon_i) \prod_{y=1}^{Y} X_{yi}^{\beta_y} \ , \tag{7}$$

where α_i is a parameter for the constant effect of brand i, X_{yi} is the value of the y^{th} variable X_y for brand i, β_y is a parameter corresponding to variable X_y, and ε_i is an error term.

Our objective is to build upon the MCI model in equation (6) to produce a model that is linear in its parameters and that represents the share of total category prescriptions written by physician j for the focal brand in period t. To minimize notational complexity and therefore improve expositional clarity, we will demonstrate this transformation assuming two particular marketing mix variables are relevant in the model. Once the transformation is complete, we will express it in its general form.

Expanding equation (6) produces an initial brand share model,

$$m_{ijt} = \left[\frac{e^{\alpha_{ij} + \varepsilon_{ijt}} D_{ijt}^{\delta_{ij}} A_{ijt}^{\varphi_{ij}}}{\sum_{k=1}^{K} e^{\alpha_{kj} + \varepsilon_{kjt}} D_{kjt}^{\delta_{kj}} A_{kjt}^{\varphi_{kj}}} \right], \tag{8}$$

where i = brand, j = physician, t = period, D = detailing, A = ads read in journal and α_{ij} = the constant effect of brand i with respect to physician j in a category with K brands. The parameters δ and φ represent the effects of detailing and promotional activities, respectively. The parameters δ and φ can vary by brand, physician, or both, addressing heterogeneity in physician response to marketing effort. The specification in equation (8) is referred to as the differential-effects MCI model (DeSarbo et al. 2002).

The model shown in equation (8) could be estimated directly using non-linear techniques. However, the estimation will be much simpler and the derivation of the elasticities much clearer by transforming equation (8) into an equation that is linear in its parameters. First, a logarithmic transformation generates

$$\ln\left(m_{ijt}\right) = \alpha_{ij} + \varepsilon_{ijt} + \delta_{ij}\ln\left(D_{ijt}\right) + \varphi_{ij}\ln\left(A_{ijt}\right) - \ln\left(\sum_{k=1}^{K}\exp\left(\alpha_{kj} + \varepsilon_{kjt}\right)D_{kjt}^{\delta_{kj}}A_{kjt}^{\varphi_{kj}}\right). \qquad (9)$$

Next, a log-centering operation is required. The first of two steps in this process are to sum equation (9) across all brands, i = (1, …, k), then divide by the number of brands, k, producing

$$\ln\left(\tilde{m}_{jt}\right) = \overline{\alpha}_{j} + \overline{\varepsilon}_{jt} + \delta_{ij}\ln\left(\tilde{D}_{jt}\right) + \varphi_{ij}\ln\left(\tilde{A}_{jt}\right) - \ln\left(\sum_{k=1}^{K}e^{\alpha_{kj}+\varepsilon_{kjt}}D_{kjt}^{\delta_{kj}}A_{kjt}^{\varphi_{kj}}\right), \qquad (10)$$

where $\tilde{m}$, $\tilde{D}$, and $\tilde{A}$ represent the geometric means of brand share, detailing, and promotional activities, respectively. The second step in the log-centering operation requires subtracting equation (10) from equation (9), resulting in

$$\ln\left(\frac{m_{ijt}}{\tilde{m}_{jt}}\right) = \alpha_{ij} - \overline{\alpha}_{j} + \delta_{ij}\ln\left(\frac{D_{ijt}}{\tilde{D}_{jt}}\right) + \varphi_{ij}\ln\left(\frac{A_{ijt}}{\tilde{A}_{jt}}\right) + \varepsilon_{ijt} - \overline{\varepsilon}_{jt}. \qquad (11)$$

This can also be written as

$$\ln\left(\frac{m_{ijt}}{\tilde{m}_{jt}}\right) = \alpha_{1j} + \sum_{m=2}^{K} \alpha^{*}_{mj} d_{mj} + \delta_{ij} \ln\left(\frac{D_{ijt}}{\tilde{D}_{jt}}\right) + \varphi_{ij} \ln\left(\frac{A_{ijt}}{\tilde{A}_{jt}}\right) + \varepsilon_{1jt} + \sum_{m=2}^{K} \varepsilon^{*}_{mjt} d_{mj}, \tag{12}$$

where $\alpha^* = \alpha_{ij} - \alpha_{1j}$, $\varepsilon^* = \varepsilon_{ijt} - \varepsilon_{1jt}$ and d = 1, if m = i and 0 otherwise (DeSarbo et al. 2002).

Since the left hand side is now a ratio, either brand shares or the actual number of brand prescriptions can be used, resulting in

$$\ln\left(\frac{S_{ijt}}{\tilde{S}_{jt}}\right) = \alpha_{1j} + \sum_{m=2}^{K} \alpha^{*}_{mj} d_{mj} + \delta_{ij} \ln\left(\frac{D_{ijt}}{\tilde{D}_{jt}}\right) + \varphi_{ij} \ln\left(\frac{A_{ijt}}{\tilde{A}_{jt}}\right) + \varepsilon_{1jt} + \sum_{m=2}^{K} \varepsilon^{*}_{mjt} d_{mj}. \tag{13}$$

Focusing on brand 1 (the focal brand), we transform equation (13) into

$$\ln\left(\frac{S_{1jt}}{\tilde{S}_{jt}}\right) = \alpha_{1j} + \delta_{1j} \ln\left(\frac{D_{1jt}}{\tilde{D}_{jt}}\right) + \varphi_{1j} \ln\left(\frac{A_{1jt}}{\tilde{A}_{jt}}\right) + \varepsilon_{1jt}. \tag{14}$$

This baseline model, fully specified, introduces a number of challenging, but addressable, econometric issues that will be discussed in full. It also provides us a convenient platform for testing several nested models that allow investigation into the value of various types of competitor data. Applying a general notation, reorganizing terms, and allowing for a variety of variables produces

$$\ln\left(\frac{S_{1jt}}{\tilde{S}_{jt}}\right) = \alpha_{1j} + \sum_{y=1}^{Y} \delta_{y,1j} \ln\left(\frac{X_{y,1jt}}{\tilde{X}_{y,jt}}\right) + \varepsilon_{1jt}, \qquad (15)$$

where X_y is the y^{th} variable.

5. ALTERNATIVE MODELS

The firm's ability to accurately estimate customer response to their marketing efforts can be limited by data availability. When relevant variables are omitted from the model, response parameters can be biased and segment assignment can be compromised, leading to sub-optimal allocation of marketing resources. The extent of the problem depends on the size of the effects the omitted variables have on the dependent variable and the degree to which the omitted variables are correlated with those variables included in the model. The construction of alternative models that represent common omitted variable situations will allow these problems to be investigated.

We will estimate and compare various specifications of the general model shown in equation (15). The notation used in equation (15) is intentionally general to allow for a clear exposition of the separation of focal firm and competitor variables that follows. Actual variables used in the models will be presented explicitly in the presentation of each alternative model.

The term representing marketing effort and other attraction variables will be manipulated to produce the alternative specifications that will be considered in the analysis. The firm may or may not have data on the marketing effort of the competitors, by brand and by physician. Additionally, the firm may or may not be aware of physician's perceptions of the effectiveness and side effects of the brands in the category. The resulting three specifications (one of which is the general model) are shown in Table 2.

Each alternative specification presented in Table 2 is nested within the general model. However, this is not clearly evident in equation (15). To produce a representation of the general model where each alternative specification is just the general model with omitted variables, the terms in equation (15) must be expanded and manipulated algebraically, since the values for each marketing variable for brand 1, $X_{y,1jt}$, also appear in the geometric mean that makes up the denominator, $\tilde{X}_{y,jt}$.

First, equation (15) can be rewritten as

$$\ln\left(\frac{S_{1jt}}{\tilde{S}_{jt}}\right) = \alpha_{1j} + \sum_{y=1}^{Y} \delta_{y,1j} \ln\left(\frac{X_{y,1jt}}{\left(X_{y,1jt}\prod_{k=2}^{K} X_{y,kjt}\right)^{\frac{1}{K}}}\right) + \varepsilon_{1jt}. \tag{16}$$

Rearranging terms, expanding, and relaxing the restriction that the effect of focal firm variables on the focal firm's market share is equivalent to the effect of corresponding competitor variables produces the general model with the focal brand variables distinct from the competitor variables as follows[1]

[1] The exponent -1/k has been changed to 1/k so the anticipated signs of the parameters associated with the competitor variables will be negative, easing interpretation.

Table 2: Variables Included in Alternative Models Based on Data Availability

Variable	(1) Firm Model	(2) Effectiveness Model	(3) Competitor Effort and Effectiveness Model
Lagged Share	yes	yes	yes
Focal Brand Detailing	yes	yes	yes
Competitor Detailing	no	no	yes
Focal Brand Journal Ads Read	no	no	yes
Competitor Journal Ads Read	no	no	yes
Perceived Effectiveness of Focal Brand	no	yes	yes
Perceived Effectiveness of Competitor Brands	no	yes	yes
Perceived Profile of Focal Brand Side Effects	no	yes	yes
Perceived Profile of Competitor Brand Side Effects	no	yes	yes
Years in Practice	yes	yes	yes
Sex	yes	yes	yes
Specialty	yes	yes	yes

$$
\begin{aligned}
\ln\left(\frac{S_{1jt}}{\tilde{S}_{jt}}\right) &= \alpha_{1j} && \text{intercept,} \\
&+ \sum_{y=1}^{Y} \delta_{y,1j} \ln\left(\left(X_{y,1jt}\right)^{\frac{K-1}{K}}\right) && \text{attraction for focal brand,} \\
&+ \sum_{y=1}^{Y} \delta_{y,cj} \ln\left(\left(\prod_{k=2}^{K} X_{y,kjt}\right)^{\frac{1}{K}}\right) && \text{attraction for competitor brands,} \\
&+ \varepsilon_{1jt} && \text{and error term.} \qquad (17)
\end{aligned}
$$

The presentation of the general model as shown in equation (17) is beneficial. Producing the alternative specifications in Table 2 now requires only the omission of specific terms from equation (17). Therefore, it will be clearly evident that each alternative specification is nested within the general model. Terms can then be recombined, as appropriate, for each alternative model. The actual variables included in the models, rather than the general notation shown in equation (17), will be presented explicitly in the following presentation of the alternative models.

$\langle 1 \rangle$ *Firm Model*

The Firm Model is the sparsest of the alternatives, with the assumption that the firm has no knowledge of competitor marketing effort or physician perceptions of brand characteristics.[2] This specification is consistent with the data typically available in a firm's database. Dropping the unobserved terms from equation (17) and inserting variable names and covariates produces

[2] We will assume the firm always knows the number of brands in the category, K, which seems reasonable, particularly in this context.

$$\ln\left(\frac{S_{1jt}}{\tilde{S}_{jt}}\right) = \alpha_{1j}^{\langle 1\rangle} + \gamma_{1,1j}^{\langle 1\rangle} Years + \gamma_{2,1j}^{\langle 1\rangle} Female + \gamma_{3,1j}^{\langle 1\rangle} OBG + \gamma_{4,1j}^{\langle 1\rangle} Urol$$

$$+ \delta_{1,1j}^{\langle 1\rangle} \ln\left(\frac{S_{1jt-1}}{\tilde{S}_{jt-1}}\right) + \delta_{2,1j}^{\langle 1\rangle} \ln\left(\left(Det_{1jt}\right)^{\frac{3}{4}}\right) + \varepsilon_{1jt}^{\langle 1\rangle}, \tag{18}$$

where Years = years in practice,

Female = sex, with 1 being female,

OBG = dummy, with 1 being OB/GYN specialist,

Urol = dummy, with 1 being urology specialist, and

Det = detailing.[3]

The terms remaining in the firm-focused model are virtually assured of being correlated with the omitted terms, resulting in biased parameter estimates. The omitted variables in this model include competitor detailing and physician perceptions of effectiveness and side effects for each brand. This anticipated bias is indicated by the superscript associated with the parameters.[4]

$\langle 2\rangle$ *Effectiveness Model*

The Effectiveness Model assumes no knowledge of competitor effort, but does contend the firm is aware of the physician perceptions of the effectiveness and side effects for each brand. This specification builds upon the data typically available in a firm's database consistent with the firm model presented above. However, it also assumes managers and salespeople are aware of physicians' perceptions of the characteristics of

[3] Operationally, detailing and prescriptions will always have 1 added to their values as in commonly done in this context to avoid the undefined ln(0).
[4] The superscripts correspond to the model numbers in Table 2.

each drug in the category. This assumption may or may not be reasonable; however, it presents the opportunity to examine the value of a model that excludes competitor marketing effort while considering physician perceptions of the competing brands. In contrast to the firm model, the specification of this model requires adding both focal firm and competitor variables for perceived brand characteristics to the firm model in equation (18), while still excluding competitor detailing, producing

$$\ln\left(\frac{S_{1jt}}{\tilde{S}_{jt}}\right) = \alpha_{1j}^{\langle 2\rangle} + \gamma_{1,1j}^{\langle 2\rangle} Years + \gamma_{2,1j}^{\langle 2\rangle} Female + \gamma_{3,1j}^{\langle 2\rangle} OBG + \gamma_{4,1j}^{\langle 2\rangle} Urol$$

$$+ \delta_{1,1j}^{\langle 2\rangle} \ln\left(\frac{S_{1jt-1}}{\tilde{S}_{jt-1}}\right) + \delta_{2,1j}^{\langle 2\rangle} \ln\left(\left(Det_{1jt}\right)^{\frac{3}{4}}\right)$$

$$+ \delta_{3,1j}^{\langle 2\rangle} \ln\left(\left(Eff_{1jt}\right)^{\frac{3}{4}}\right) + \delta_{3,cj}^{\langle 2\rangle} \ln\left(\left(\prod_{k=2}^{4} Eff_{kjt}\right)^{\frac{1}{4}}\right)$$

$$+ \delta_{4,1j}^{\langle 2\rangle} \ln\left(\left(SE_{1jt}\right)^{\frac{3}{4}}\right) + \delta_{4,cj}^{\langle 2\rangle} \ln\left(\left(\prod_{k=2}^{4} SE_{kjt}\right)^{\frac{1}{4}}\right) + \varepsilon_{1jt}^{\langle 2\rangle}, \tag{19}$$

where Eff = perceived effectiveness of the brand, and

SE = quality of the side effect profile for the brand.

$\langle 3\rangle$ *Competitor Effort and Effectiveness Model*

The Competitor Effort and Effectiveness Model is the fully specified model presented in equation (17). The assumption for the model is that the firm has augmented their database to include not only physician perceptions of brand characteristics, but also competitor detailing and the frequency with which physicians view journal ads for each of the brands in the category. The explicit presentation of the general model is

$$\ln\left(\frac{S_{1jt}}{\tilde{S}_{jt}}\right) = \alpha_{1j}^{\langle 3\rangle} + \gamma_{1,1j}^{\langle 3\rangle} Years + \gamma_{2,1j}^{\langle 3\rangle} Female + \gamma_{3,1j}^{\langle 3\rangle} OBG + \gamma_{4,1j}^{\langle 3\rangle} Urol$$

$$+ \delta_{1,1j}^{\langle 3\rangle} \ln\left(\frac{S_{1jt-1}}{\tilde{S}_{jt-1}}\right) + \delta_{2,1j}^{\langle 3\rangle} \ln\left(\left(Det_{1jt}\right)^{\frac{3}{4}}\right) + \delta_{2,cj}^{\langle 3\rangle} \ln\left(\left(\prod_{k=2}^{4} Det_{kjt}\right)^{\frac{1}{4}}\right)$$

$$+ \delta_{3,1j}^{\langle 3\rangle} \ln\left(\left(Eff_{1jt}\right)^{\frac{3}{4}}\right) + \delta_{3,cj}^{\langle 3\rangle} \ln\left(\left(\prod_{k=2}^{4} Eff_{kjt}\right)^{\frac{1}{4}}\right)$$

$$+ \delta_{4,1j}^{\langle 3\rangle} \ln\left(\left(SE_{1jt}\right)^{\frac{3}{4}}\right) + \delta_{4,cj}^{\langle 3\rangle} \ln\left(\left(\prod_{k=2}^{4} SE_{kjt}\right)^{\frac{1}{4}}\right)$$

$$+ \delta_{5,1j}^{3} \ln\left(\left(JA_{1jt}\right)^{\frac{3}{4}}\right) + \delta_{5,cj}^{\langle 3\rangle} \ln\left(\left(\prod_{k=2}^{4} JA_{kjt}\right)^{\frac{1}{4}}\right) + \varepsilon_{1jt}^{\langle 3\rangle}, \tag{20}$$

where JA = journal ads read by the physician.

6. DATA

Firm records and survey data were combined to produce the dataset used in this study. The firm records are at the physician level and come from the internal database of a pharmaceutical firm competing in a large ethical drug category. The survey data consists of the responses from a subset of the physicians in the firm's database. The four brands included in the study comprise about 90% of the market.

Primary and Secondary Data

The initial data provided by the firm consists of the names, contact information, and internal rankings for each physician assigned to the category by the firm. The firm's internal rankings place each physician in one of eleven categories roughly based on their total category demand. Since the bottom two categories had very limited or no activity in the category year-to-date, they were dropped from consideration, resulting in a population of physicians numbering 7,101. Fax numbers were secured for these physicians.

The survey was completed by a subset of these 7,101 physicians.[5] The survey data provides variables related to brand-specific detailing by the competition and physicians' perceptions of the characteristics of each brand (effectiveness in treating each of three common symptoms, as well as side effects). Detailing by the focal firm is also included. By surveying variables included in the firm's data, the accuracy of the physicians' responses can be considered.

All 7,101 physicians were invited to participate in the survey in exchange for a small honorarium. The survey was delivered via fax to each physician's primary practice

[5] A reproduction of the survey, with the category details removed to protect the identity of the focal firm, appears in the Appendix.

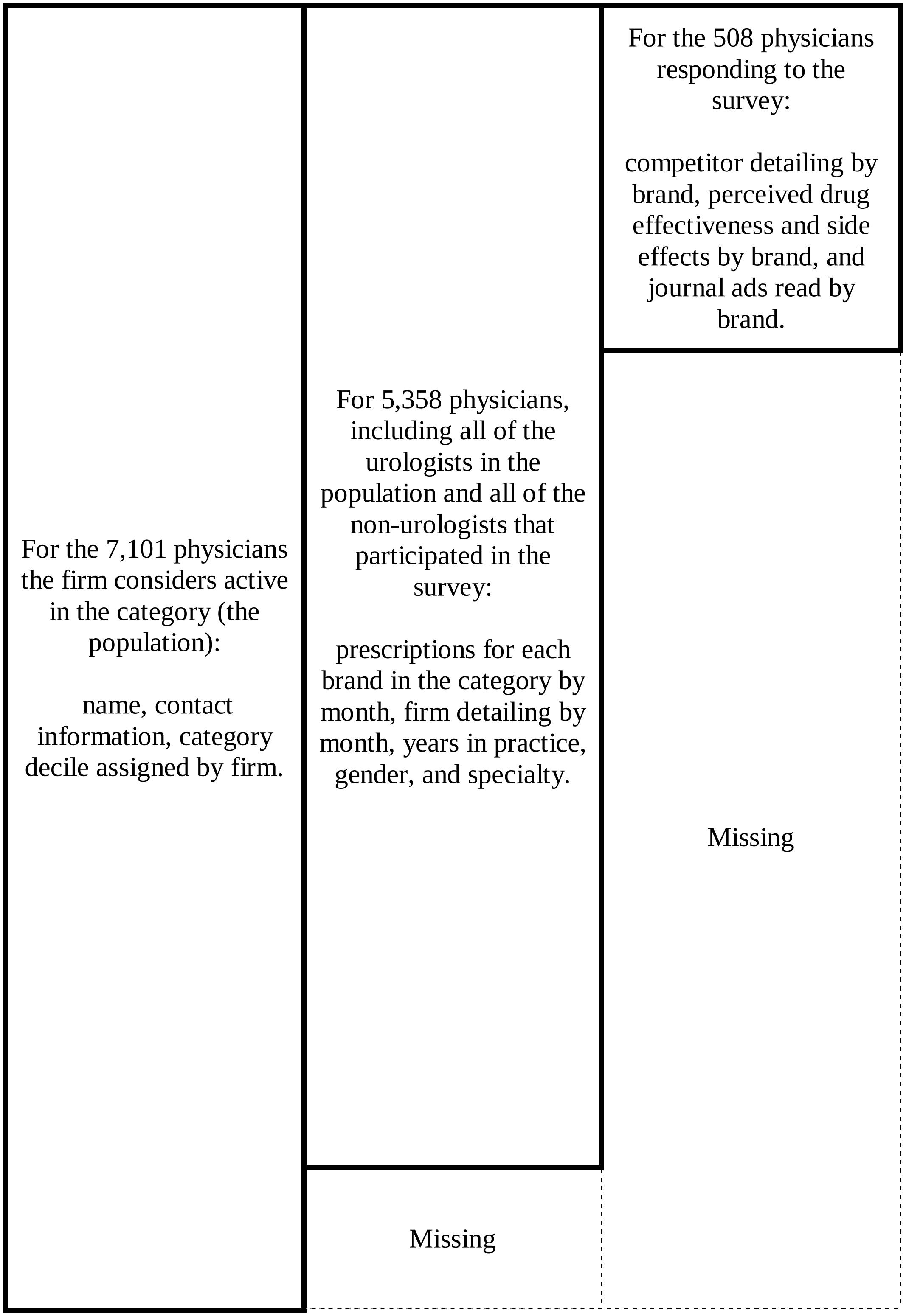
location. The initial solicitation yielded 393 useable responses. Non-respondents were offered an additional opportunity to participate. One hundred-fifteen useable responses resulted from the second offer, resulting in a total of 508 respondents and a response rate of 7.2%.

The firm then provided additional data for 5,358 of the 7,101 physicians, including all of the physicians practicing in the dominant specialty in this category, as well as for any physician in a different specialty that responded to the survey. This includes a monthly panel (January 2007 – October 2007) of brand-specific data on prescriptions written, along with data on detailing and sampling for the focal firm. The specialty, number of years in practice, and sex of each physician was also provided. The dataset resulting from combining the firm and survey data is graphically represented in Figure 1.

The firm roughly categorizes physicians based on their total category demand, with higher prescribing physicians belonging to a higher numbered category. Some physicians in the data are unclassified. Table 3 shows the percentage of respondents coming from each category and the corresponding breakdown of non-respondents by category.[6] Generally speaking, the sample appears to have been drawn proportionally from each category, with the possible exceptions of under-sampling in the lowest category demand groups (categories 3 and 4) and over-sampling from category 9. This suggests that the sample may overemphasize high prescribing physicians unless the unclassified group is made up of higher than average prescribers.

[6] Since firm data was available for non-respondents only in the dominant specialty, the comparison of respondents to non-respondents is restricted to that specialty. Results are assumed to be similar for other specialties.

Figure 1: Graphical Representation of the Databases Used in the Study



Rows represent physicians. Columns consist of variables in the indicated categories.

Table 3: Comparison of Category Representation for
Respondents and Non-Respondents

Firm's Categorization Based on Total Category Demand	% of Respondents	% of Non-Respondents
3	0.7	1.7
4	1.5	2.4
5	4.4	3.0
6	6.6	5.0
7	7.7	9.2
8	19.9	22.0
9	36.4	31.3
10	0.7	0.5
Unclassified	22.1	25.0
Total	100.0	100.0

Since detailing and prescription writing data is available for some physicians in each category, the average figures can be multiplied by the percentage representation in each category to produce weighted averages that can be used for comparison. These results appear in Table 4. Since the over-sampled unclassified group is made up of physicians that are detailed more and prescribe more than physicians in category 9, the net effect is a similar set of weighted averages, indicating the sample is a reasonable representation of the population.

Incorporation of Survey Competitor Detailing Data

Although the firm provided panel data for brand prescription writing and firm detailing over a number of months, the data collected via the survey was taken at a particular point in time. The analysis could be conducted as a cross-sectional analysis, incorporating the survey data and considering only one period of prescription and firm

Table 4: Comparison Between Survey Respondents and Non-Respondents

	Respondents	Non-Respondents
Weighted Average Number of Details	1.436	1.435
Weighted Average Number of Total Prescriptions for the Focal Brand	1.707	1.723
Weighted Average Number of New Prescriptions for the Focal Brand	0.602	0.608

detailing data. This approach would effectively throw away the majority of the available data. A second approach would be to use the survey data to get some idea of the extent of competitor detailing and apply that knowledge over time, allowing all of the prescription and firm detailing data to be used. The latter approach will be used in this study.

The survey intentionally did not identify the focal firm. Firm data was collected along with competitor data, allowing the survey data to be compared to the firm's records to assess the physicians' ability to accurately report detailing activity for the focal firm. The necessary assumption to incorporate the survey data on competitor detailing into the dataset over time is that the accuracy with which each physician is able to report the detailing efforts of the focal firm will be similar to the accuracy in reporting competitor detailing. The objective is not to predict the number of competitor details for each physician during each time period. A survey taken at a single point in time is unable to provide the insight for that task. However, using the survey data for focal firm detailing along with the firm's record of detailing over time, a mean level of detailing for each physician coupled with a distribution of those details over time allows for model estimation using a multiple imputation approach. Multiple imputation requires using several randomly drawn datasets from the representative distribution of the missing data in order to account for what is essentially measurement error in the imputed data. In this case, competitor detailing over time is missing data.

The objective is to produce a dataset generator that can produce randomly drawn datasets using physician reported competitor detailing levels, along with the model derived from the firm detailing records and the survey data for focal firm detailing. The proposed relationship between survey reported detailing and actual details is

$$Det_{ijt} = e^{\alpha_{ij} + \sum_{m=1}^{5} \beta_m d_m + \varepsilon_{ijt}} SurvDet_{ij}^{\gamma}, \tag{21}$$

where d is a dummy variable accounting for seasonal variations in detailing levels, assuming a value of 1 if m = t and 0 otherwise. A transformation produces a function that is linear in its parameters and allows a convenient link in count data functions,

$$\ln\left(Det_{ijt}\right) = \alpha_{ij} + \sum_{m=1}^{5} \beta_m d_m + \gamma \ln\left(SurvDet_{ij}\right) + \varepsilon_{ijt}. \tag{22}$$

In effect, this mean of competitor detailing is a function of the survey reported number of details per period with an adjustment for month. Note the notation indicating the intercept will vary over physicians.

Detailing was initially assumed to be distributed Poisson. Estimating the model using Poisson regression, however, indicated the data was overdispersed, with

$g(\mu_i) \sim \chi_1^2 = 5.97 > 3.84$ (Cameron and Trivedi 1998). Typically, overdispersed count data is estimated using a negative binomial model. However, since this model is to be used for making random draws, a simpler alternative exists. With panel data, a Poisson model allowing for random effects essentially is similar to the negative binomial with the overdispersion parameter varying across groups (Greene 2007). Conducting random draws using the parameter estimates from the random effects Poisson model is convenient. The results from estimating a Poisson regression model with random effects as shown in equation (22) are presented in Table 5. These results, coupled with the survey responses concerning competitive detailing, are used to produce the set of datasets to be used in the multiple imputation.

Table 5: Competitor Detailing Model Based on Survey Data

Log Likelihood	-3956.516	
Variable	Est.	SE
Intercept	0.14*	0.07
Reported Monthly Detailing	0.49*	0.08
Time 1	-0.32*	0.07
Time 2	-0.54*	0.07
Time 3	-0.24*	0.07
Time 4	-0.17*	0.07
Time 5	-0.41*	0.08
Alpha	0.48*	0.04

* = significant at 0.1

Incorporation of Other Survey Variables

In addition to competitive detailing, the survey was also used to collect data on physicians' perception of the effectiveness of each brand in the category, along with the side effect profile of each brand. Brand effectiveness was measured using a set of three five-point scales, focusing on the three common symptoms associated with the condition leading to the prescribing of drugs in the category. To account for measurement error, one of the three ratings is randomly drawn for each physician for each dataset constructed for the multiple imputation. Side effects are reported to be minimal in this category. Therefore, a single item, five-point scale was considered adequate to measure the physicians' perceptions of the side effect profile. Likewise, the number of times per period the physician sees a journal ad for a particular brand of drug in the category was measured with a single item. Each brand regularly has ads appearing in the major journals. Finally, years-in-practice was not available for all physicians. The existing data on this variable fit a lognormal distribution. Missing values were randomly drawn from this distribution for each dataset created.

7. ESTIMATION

A number of econometric issues need to be accounted for to accurately estimate each of the alternative models shown in equations (18), (19), and (20). First, econometric concerns common in this type of data, endogeneity and heterogeneity, must be addressed. Second, multiple imputation require the estimation of the model using a set of datasets. The results can then be used to calculate parameter estimates and standard errors.

Endogeneity

Two endogenous variables appear in the models. First, the lagged dependent variable is by definition correlated with the error term in the model and therefore must be treated as an endogenous variable. Second, detailing cannot be assumed to be exogenous. Detailing levels are set by the firm, at least partly based on market share results observed in the past. Instruments will be used to account for endogeneity in this study, specifically lagged values of share and detailing.

Effective instruments must be shown to be valid and exogenous. Validity pertains to whether the instrument is correlated with the endogenous variable after accounting for all of the other exogenous variables in the model. For all three alternative models (see Table 2), the instruments appear to be valid. Lagged share was regressed on the instrument, share lagged two periods, along with all of the exogenous variables in the model. For each of the three alternative models, the t-statistic for the parameter attached to the instrument was significant (respectively, $t = 27.4$, $t = 26.3$, and $t = 26.1$).

Similarly, lagged detailing provides valid instruments for firm detailing. Detailing lagged one, two, and three periods is used in each model. For the Effectiveness and Competitor Effort Model, the t-statistics for the three lagged instruments are $t = 14.4$, 12.6, and 8.8, respectively. The results are similar for the two other alternative models.

Not only must instruments be valid, but also exogenous. Two over-identifying restrictions allow the exogeneity of the instruments in the model to be tested using a Sargan test (Sargan 1958). First, the model is estimated using the instruments. The residuals from that estimation are then regressed on all of the exogenous variables. The number of observations times the resulting R^2, NR^2, is distributed χ^2 with two degrees of freedom, equivalent to the number of over-identifying restrictions. The null hypothesis is that all instruments are exogenous, so a failure to reject indicates exogenous instruments. For each of the three models, the Firm Model, the Effectiveness Model, and the Effectiveness and Competitor Effort Model, the result is a failure to reject the null. NR^2 is 1.65, 1.63, and 1.34 respectively for the three models, all below the value of χ^2 with two degrees of freedom, 5.99 ($\alpha = 0.05$). Therefore, the instruments for all three alternative models are effective in addressing the endogeneity inherent in the models.

Heterogeneity

Physician response to the various variables in the models can certainly be expected to vary across physicians. Heterogeneity must be addressed. The nature of the data allow a number of reasonable alternatives.

A fixed effects approach is a consideration whenever panel data is being analyzed. In this case, the impact of unobserved, time invariant effects can be estimated for each physician. One strength of the fixed effects approach is that the unobserved effects can be correlated with the time varying variables. A key disadvantage is parameter estimates are possible only for time varying variables. In this research, there are variables of interest that do not vary over time, or are assumed not to vary over the time periods analyzed in this study. Therefore, a fixed effect approach is problematic.

Estimating random effects is another alternative with panel data. The key distinction between fixed and random effects is with random effects, the unobserved effects must be assumed to be orthogonal to the variables included in the model. Unlike with fixed effects, parameter estimates for time invariant variables can be estimated. Random effects is a reasonable alternative if time invariant variables that are correlated with time variant regressors can be observed and included in the model. Unfortunately, with both fixed and random effects, only heterogeneity in the unobserved effects across physicians can be addressed.

Given distributional assumptions, Bayesian approaches potentially allow all parameter estimates to vary across individuals. Likewise, a latent class approach allows heterogeneous response for all parameters estimated, not across individuals, but across groups of individuals. A latent class approach requires no distributional assumptions. Given a specified number of points of support, or latent classes, a latent class estimation produces a discrete, finite sample distribution of the parameters. Latent class analysis can also be informative managerially, producing often managerially relevant segments based on response. Considering the objectives of this study, comparing parameter estimates to investigate bias, reallocation of marketing resources based on segmentation, and incorporating survey data, a latent class approach is particularly attractive.

As discussed earlier in the section describing the data to be used in the study, multiple datasets were constructed to be used in the multiple imputation. Six datasets were analyzed in this study, an adequate number to produce relevant inferences using complete data methods (Little and Rubin 2002). Using instruments for endogenous variables and latent class analysis to address heterogeneity of response, each of the six datasets were used to estimate each of the three alternative models. Information criteria were used to determine the appropriate number of latent classes. The Bayesian (or Schwartz) information criterion (BIC) indicates a three latent class model for all six datasets. The Akaike information criterion (AIC), which penalizes additional parameterization less than does the BIC, is minimized with five latent classes in all cases. All things equal, parsimony suggests the three class model.

8. RESULTS

The results from the estimation of each of the three alternative models appear in Tables 6, 7, and 8. The Effectiveness Model fits significantly better than the Firm Model ($\Delta LL = 25.11 > \chi^2_{12} = 18.55$), The Effectiveness and Competitor Effort Model fits significantly better than the Effectiveness Model ($\Delta LL = 16.19 > \chi^2_{9} = 14.68$), and the Effectiveness and Competitor Effort Model fits significantly better than the Firm Model ($\Delta LL = 41.30 > \chi^2_{21} = 29.62$).

Table 6: Firm Model Results

Model	Class 1		Class 2		Class 3	
Log likelihood	-1326.55					
Avg. Membership %	38.0		30.3		31.7	
Variable	Est.	SE	Est.	SE	Est.	SE
Intercept	-0.30*	0.16	-0.05	0.22	-0.06	0.06
Lagged Share	0.10	0.28	1.03*	0.34	0.53*	0.14
Firm Detailing	0.22*	0.06	-0.01	0.07	-0.17*	0.05
Years in Practice	0.00	0.00	0.00	0.00	0.00^{o}	0.00
Sex (female)	-0.04	0.05	-0.11*	0.05	0.07*	0.03
OBG/GYN	0.09	0.10	-0.03	0.07	-0.05*	0.03
Urologist	-0.17*	0.08	-0.08	0.13	-0.20*	0.10
Sigma	0.42*	0.06	0.43*	0.11	0.21*	0.03

* = significant at 0.1

Interestingly, there are similarities in the latent classes across the three models. Most noticeably, each model assigns a positive coefficient for firm detailing for latent

Table 7: Effectiveness Model Results

Model	Class 1		Class 2		Class 3	
Log likelihood	-1301.44					
Avg. Membership %	38.3		32.6		29.7	
	Est.	SE	Est.	SE	Est.	SE
Intercept	-0.25	0.23	-0.02	0.26	-0.07	0.12
Lagged Share	0.74*	0.32	0.24^{o}	0.42	0.44*	0.29
Firm Detailing	0.20*	0.11	-0.01	0.09	-0.18*	0.07
Focal Brand Effectiveness	0.24^{o}	0.22	0.22^{o}	0.36	0.01	0.16
Competitor Brand Effectiveness	-0.11	0.30	-0.48^{o}	0.53	-0.06	0.20
Focal Brand Side Effects	0.04	0.08	-0.03	0.07	-0.01	0.06
Competitor Brand Side Effects	-0.02	0.12	0.05	0.17	0.03	0.07
Years in Practice	0.00	0.00	0.00	0.00	0.00^{o}	0.00
Sex (female)	-0.11^{o}	0.07	-0.02	0.12	0.08*	0.04
OBG/GYN	0.02	0.07	0.05	0.10	-0.04	0.05
Urologist	-0.13^{o}	0.08	-0.26^{o}	0.23	-0.12*	0.16
Sigma	0.50*	0.02	0.30*	0.05	0.21*	0.03

* = significant at 0.1 ^{o} = significant considering only variance within datasets

Class 1, indicating there are a group of physicians that are currently being under-detailed. Similarly, the detailing levels appear to be adequate for physicians in Class 2. Class 3 physicians appear to be over-detailed; however, the negative coefficient is not significant for the Effectiveness and Competitor Effort Model. This result for the Firm Model is consistent with previous research that did not account for drug effectiveness and competitor detailing. It also confirms the speculation in Manchanda and Chintagunta (2004) that the finding of an over-detailed group may be the result of omitting competitor effort from the analysis.

Table 8: Effectiveness and Competitor Effort Model Results

Model	Class 1		Class 2		Class 3	
Log likelihood	-1285.25					
Avg. Membership %	44.4		28.9		26.7	
	Est.	SE	Est.	SE	Est.	SE
Intercept	-0.09	0.17	0.14	0.14	-0.27*	0.21
Lagged Share	0.83*	0.06	0.27*	0.13	-0.00	0.12
Firm Detailing	0.18*	0.07	-0.03	0.05	-0.11^{o}	0.07
Competitor Detailing	-0.11*	0.07	-0.05	0.06	-0.04	0.06
Focal Brand Effectiveness	0.22*	0.09	-0.01	0.09	0.05	0.09
Competitor Brand Effectiveness	-0.24*	0.13	-0.10	0.19	-0.13	0.20
Focal Brand Side Effects	0.01	0.06	0.06^{o}	0.05	0.04	0.06
Competitor Brand Side Effects	0.10	0.09	-0.26*	0.11	-0.05	0.15
Focal Brand Journal Ads	0.10*	0.04	0.06^{o}	0.04	0.20*	0.05
Competitor Brand Focal Ads	-0.09*	0.05	0.00	0.06	-0.16*	0.05
Years in Practice	0.00	0.00	0.00^{o}	0.00	0.00^{o}	0.00
Sex (female)	-0.14*	0.04	0.07^{o}	0.05	0.16*	0.04
OBG/GYN	0.03	0.06	-0.05	0.04	0.05	0.05
Urologist	-0.07	0.06	-0.61*	0.10	0.00	0.05
Sigma	0.51*	0.01	0.23*	0.02	0.24*	0.02

* = significant at 0.1 ^{o} = significant considering only variance within datasets

There are some interesting differences in parameters within classes across the models, particularly the effects of lagged share and drug effectiveness. These differences, along with competitor detailing, which appears only in the Effectiveness and Competitor Effort Model, suggest the underlying biases found in the Firm Model and the Effectiveness Model. Although the parameter estimates for firm detailing are similar within classes but across models, the assignment of physicians to the various classes differ significantly across models.

Table 9 and Table 10 show the extent of the misclassification of physicians, with Table 9 comparing the Firm Model to the Effectiveness and Competitor Effort Model and Table 10 comparing the Effectiveness Model to the Effectiveness and Competitor Effort Model. The bias in parameter estimates for the misclassified physicians are now evident. The Firm Model produces significantly biased parameter estimates for firm detailing for 35.6% of the physicians in the sample. The Effectiveness Model fares better, producing significantly biased firm detailing parameters for 9.1% of the physicians in the sample.

Table 9: Physician Class Assignments Comparison: Firm vs. Effectiveness and Competitor Effort Model

	Class Assignment Using Firm Model		
Class Assignment Using Effectiveness and Competitor Effort Model	Class 1 Parameter for firm detailing = 0.22	Class 2 Parameter for firm detailing = -0.01	Class 3 Parameter for firm detailing = -0.17
Class 1 Parameter for firm detailing = 0.18	148 physicians 29.1% of sample t_{diff} = -0.46	63 physicians 12.4% of sample t_{diff} = 1.89*	0 physicians 0.00% of sample t_{diff} = 4.15*
Class 2 Parameter for firm detailing = -0.03	6 physicians 1.2% of sample t_{diff} = -3.13*	67 physicians 13.2% of sample t_{diff} = -0.21	79 physicians 15.6% of sample t_{diff} = 2.03*
Class 3 Parameter for firm detailing = -0.11	33 physicians 6.5% of sample t_{diff} = -3.37*	48 physicians 9.4% of sample t_{diff} = -0.95	64 physicians 12.6% of sample t_{diff} = 0.67

Description of the Latent Classes

A comparison of the average levels of key variables shown in Table 11 gives insight into the variance in the classifications between the Firm Model and Effectiveness and Competitor Effort Model (ECE Model). Both the Firm Model and the ECE Model assign physicians with high detailing levels, high firm brand prescriptions, and favorable

Table 10: Physician Class Assignments Comparison: Effectiveness vs. Effectiveness and Competitor Effort Model

	Class Assignment Using Effectiveness Model		
Class Assignment Using Effectiveness and Competitor Effort Model	Class 1 Parameter for firm detailing = 0.20	Class 2 Parameter for firm detailing = -0.01	Class 3 Parameter for firm detailing = -0.18
Class 1 Parameter for firm detailing = 0.18	196 physicians 38.6% of sample t_{diff} = -0.21	6 physicians 1.2% of sample t_{diff} = 1.69*	9 physicians 1.8% of sample t_{diff} = 3.84*
Class 2 Parameter for firm detailing = -0.03	7 physicians 1.4% of sample t_{diff} = -2.01*	132 physicians 26.0% of sample t_{diff} = -0.23	13 physicians 2.6% of sample t_{diff} = 1.87*
Class 3 Parameter for firm detailing = -0.11	11 physicians 2.2% of sample t_{diff} = -2.43*	10 physicians 2.0% of sample t_{diff} = -0.91	124 physicians 24.4% of sample t_{diff} = 0.70

perceptions of the effectiveness of the firm's brand to Class 1. However, the ECE Model assigns higher category prescribers to Class 1 than does the Firm Model. There are more distinctions between the models for Class 2. The Firm Model indicates firm detailing should remain constant for very high category prescribers where the firm has a low share, in spite of a high level of competitive detailing and low perceptions of the effectiveness of the brand. In contrast, the ECE Model indicates maintaining the level of detailing for low share physicians, but at a relatively lower level of detailing and for physicians that have a favorable perception of the brand. Finally, the Firm Model suggests lowering the level of detailing for low category prescribers where the share is low, even though the perception of the brand is high. The ECE Model indicates physicians that hold a low perception of brand effectiveness should be detailed less. These differences are consistent with the differences in class assignment between the two models.

Table 11: Comparison of Class Profiles Between Models

		Firm Model	Effectiveness and Competitor Effort Model
Class 1	Firm details	1.44	1.40
	Average competitor details	1.18	1.22
Detail More			
	Firm brand prescriptions	0.76	0.72
	Total category prescriptions	6.18	8.71
	Perceived effectiveness of focal brand	3.65	3.64
	Perceived effectiveness of competitor brands	3.78	3.81
Class 2	Firm details	1.12	0.89
	Average competitor details	1.28	1.12
Detail Same			
	Firm brand prescriptions	0.22	0.09
	Total category prescriptions	10.17	5.65
	Perceived effectiveness of focal brand	3.38	3.57
	Perceived effectiveness of competitor brands	3.88	3.93
Class 3	Firm details	0.84	1.09
	Average competitor details	1.05	1.18
Detail Less			
	Firm brand prescriptions	0.06	0.17
	Total category prescriptions	2.51	4.33
	Perceived effectiveness of focal brand	3.54	3.32
	Perceived effectiveness of competitor brands	3.86	3.78

Reallocation

The implications of reallocating details based on the results of the three models can now be investigated. The parameters in the models are essentially elasticities. Since the key actionable variable, detailing, is a non-negative integer and the estimates are valid at the current level of detailing, reallocation based on adding or subtracting a single detail is appropriate.

The expected increase in prescriptions for the focal brand due to an increase or decrease of a single detail can be derived using any of the three alternative models, since only the firm detailing term common to all models figures into the partial derivative. The equations for all three alternative models can be written as

$$\ln\left(\frac{S_{1jt}+1}{\tilde{S}_{jt}}\right)=\delta_{1j}\ln\left(\left(Det_{1jt}+1\right)^{\frac{3}{4}}\right)+C\,,^{7} \tag{23}$$

where C consists of everything else on the right hand side for any of the three alternative models. Manipulating equation (23) algebraically produces

$$\ln\left(S_{1jt}+1\right)=\frac{4\delta_{1j}}{3}\ln\left(\left(Det_{1jt}+1\right)^{\frac{3}{4}}\right)+C'\,, \tag{24}$$

where the notation C' indicates the change in additional terms on the right hand side, so

$$\frac{\partial\ln\left(S_{1jt}+1\right)}{\partial Det_{1jt}}=\frac{\delta_{1j}}{Det_{1jt}+1}\,. \tag{25}$$

Therefore, with a single additional detail,

$$\ln\left(S'_{1jt}+1\right)-\ln\left(S_{1jt}+1\right)=\frac{\delta_{1j}}{Det_{1jt}+1}\,, \tag{26}$$

---

[7] As mentioned along with the presentation of the general model, for practical reasons "1" is added to prescriptions and detailing when estimating the models. This addition has been suppressed in previous equations, but must now be expressed explicitly for the forthcoming calculations.

where S'_{1jt} is the number of prescriptions after adding the detail. Solving for the new number of prescriptions produces

$$S'_{1jt} = (S_{1jt} + 1)\exp\left(\frac{\delta_{1j}}{Det_{1jt} + 1}\right) - 1. \tag{27}$$

Therefore, the change in prescriptions resulting from the additional detail is

$$S'_{1jt} - S_{1jt} = (S_{1jt} + 1)\left(\exp\left(\frac{\delta_{1j}}{Det_{1jt} + 1}\right) - 1\right). \tag{28}$$

A similar result for removing a detail produces the following change in prescriptions,

$$S'_{1jt} - S_{1jt} = (S_{1jt} + 1)\left(\exp\left(\frac{-\delta_{1j}}{Det_{1jt} + 1}\right) - 1\right). \tag{29}$$

Applying a reallocation of detailing plan to a dataset based on each of the three alternative models allows a comparison of the relative value of the models to the firm. The results presented above in Tables 6, 7, and 8 indicate a class of over-detailed physicians and a class of under-detailed physicians for each of the three models, respectively. The simple reallocation plan considers the relative sizes of the under-detailed segment, Class 1, and the over-detailed segment, Class 3, for each model. If Class 1 consists of more physicians than does Class 3, one detail is assumed to be taken from each physician in Class 3 and reallocated to a randomly selected group of Class 1 physicians. If Class 1 consists of fewer physicians than does Class 3, one detail is assumed to be reallocated to each physician in Class 1 from a randomly selected group of Class 3 physicians. The results of this exercise are presented in Table 12.

Table 12: Reallocation Results for Each Alternative Model

	Firm Model	Effectiveness Model	ECEM
Details Reallocated (% of Total)	416 (11.8%)	496 (14.0%)	518 (14.7%)
Additional Prescriptions for 508 Physicians in Sample (% Increase)	64 (5.4%)	99 (8.3%)	105 (8.8%)
Prescription Increase Per Physician	0.13	0.19	0.21
Projected Increase in Prescriptions for 7101 Physicians in Population	895	1384	1468

With the net revenue for each prescription filled at over $100 and a conservative estimate of resulting monthly refills of six months, the discounted value of each incremental new prescription is around $600. A reallocation plan based on the ECE Model could be expected to net the firm additional revenue of around $880,000, nearly $350,000 more than a plan based on the Firm Model results. Biased response estimates are costly to the firm.

9. DATA AUGMENTATION

Previous pharmaceutical research that has incorporated competitor detailing (and even drug effectiveness) have analyzed panel data for a portion of the physicians active in the category (Gonul et al. 2001; Venkataraman and Stremersch 2007). Neither study attempted to apply their findings to physicians outside of the panel. Consistent with these studies, firms cannot expect to acquire competitor detailing data or perceptions of brand effectiveness for more than a portion of the physicians active in the category. When a survey is used to collect the additional data, non-response virtually assures data will be missing for some (likely a majority) of the physicians.

The firm is not only interested in understanding how segments of physicians respond to their marketing efforts based on the analysis of a sample of their customers. Ultimately, they want to identify the relevant segments, and then determine segment membership for all of the physicians active in the category. Ideally, an analysis of a sample of the physicians can determine the segments that exist in the category. Then, data augmentation can be used to assign physicians outside of the sample to the segments.

Typically, data augmentation is done to predict particular missing values. For example, in Du et al. (2007), bank customers were surveyed to determine share-of-wallet in various product categories. Data augmentation was then used to predict share-of-wallet for a holdout sample based on the analysis of the calibration sample. Similarly, Kamakura and Wedel (2003) developed an augmentation method to predict particular variables, such as share-of-wallet and customer satisfaction. Here, the ultimate objective is not to predict the value of particular variables collected via a survey for those customers not included in the survey, but rather to assign physicians not included in the survey to segments.

The augmentation approach used in this study is unique is three ways. First, the missing data relates directly to competitor data, commonly considered to be a critical and common type of missing data (e.g. Boulding et al. 2005). Second, since the ultimate objective is to assign physicians to segments rather than impute individual values for missing data, the method involves a multivariate draw of missing values based on a multiplicative composite of all of the key missing survey variables, consistent with the modeling approach. Third, segment membership probabilities are not modeled as a function of observed variables, but instead are the result of using a complete data analysis technique incorporating both physicians with survey data and those with imputed data.

A dataset was first drawn and analyzed as discussed in the estimation section above and segment assignments were noted for all 508 physicians in the sample. Next, parameter estimates to be used for drawing competitor detailing datasets as shown in equation (22) were re-estimated using a randomly selected 458 of the 508 physicians in the sample. Competitor detailing was drawn for each of the 458 physicians based on these estimates. All variables associated with the survey were assumed to be unknown for the remaining 50 physicians, including competitor detailing, drug effectiveness measures, side effect measures, and journal ads read.

The product of the competitor detailing, firm brand effectiveness, and competitor brand effectiveness variables was then calculated for each of the 458 physicians in the analysis sample. This composite was then regressed on the variables in the dataset that were observed for all physicians, specifically the four demographic variables, firm detailing, and lagged share. These estimates were used to calculate the predicted composite variable of survey related variables for the 50 physicians in the holdout sample. Missing data for the 50 physicians in the holdout sample was then imputed to be the same as the physician in the calibration sample with the composite variable value closest to that of each physician in the holdout sample.

The now complete dataset was analyzed for the ECE Model as before. Segment assignments based on the data with imputed values could then be compared with segment assignments based on an analysis of the full sample to determine the accuracy of the augmentation system. The results are presented in Table 13.

Table 13: Accuracy of Segment Assignment Using Augmented Data for 50 Physicians in Holdout Sample

	Class Assignment Using Data Augmentation		
Class Assignment Using Full Sample	Class 1	Class 2	Class 3
Class 1	10 of 10 physicians 100% accuracy for Class 1	0 physicians	0 physicians
Class 2	0 physicians	19 of 19 physicians 100% accuracy for Class 2	0 physicians
Class 3	2 of 21 physicians 9.5% missed for Class 3	0 physicians	19 of 21 physicians 90.5% accuracy for Class 3

The augmentation approach appears to be quite effective for this particular dataset and holdout sample. Of the 50 physicians in the sample, only two were misclassified when augmented data was used. Two physicians assigned to the under-detailed segment (Class 1) should have been assigned to the over-detailed segment (Class 3). The 96% success rate, as compared to the results shown in Tables 9 and 10, indicate surveying a portion of physicians to acquire data on competitor detailing and drug brand effectiveness then augmenting the database to make segment assignments outperforms both the Firm Model which uses data already in the firm’s database and the Effectiveness Model which requires collecting brand effectiveness data from all of the firm’s customers.

10. FUTURE RESEARCH

Generalizability

Ethical drug manufacturers typically possess data on prescriptions written for all of the competing brands at the physician level. Access to this type of competitor data is extremely rare in other industries, and even within the pharmaceutical industry outside of the U. S. In this study, brand sales are never considered to be omitted variables due to lack of data availability. However, in most industries, firm databases typically do not include this data.

Future research could focus on generalizing the approach used in this study, allowing it to be applied in industries where competitor sales are unknown. The primary modification would be to move from an analysis of market share to an investigation of brand sales. In effect, this would require a model for total category demand. Combining the market share model with a total category demand model would indirectly produce a brand sales model (Leeflang et al. 2000). With that change, the sparsest model would include marketing effort and sales data only for the focal firm. The collection and value of competitor brand sales data could then be considered similarly to brand effectiveness and competitor effort data in this study.

Endogeneity

In this study, endogenous variables were not modeled explicitly, but rather were addressed using instrumental variables. The investigation of the antecedents of these variables, as well as potential causal relationships among these variables, could lead to valuable future research. The relationship between detailing and physician perceptions of the relative effectiveness of the brands is of particular interest.

Only one study explicitly includes brand effectiveness in the model (Venkataraman and Stremersch 2007), although the assumption is that all physicians have identical perceptions. This assumption suggests perceived effectiveness is solely the result of experience with the drug. However, detailing has been found to have a mostly informative effect in some studies (e.g. Gonul et al. 2001), possibly suggesting detailing may be an antecedent to physician perceptions of brand effectiveness. If perceived effectiveness proved to fully mediate the effects of detailing on prescription writing, understanding the relative effectiveness of the brands as perceived by the physicians would make competitor detailing data redundant. Alternatively, perceptions of drug effectiveness could drive detailing levels if the manufacturer actively targeted physicians with low perceptions of the effectiveness of their brand.

Although not directly investigated in this study, the data suggest that both detailing and perceived effectiveness impact prescription writing. If perceived effectiveness fully mediated the effect of detailing, the correlation between detailing and perceived effectiveness would be expected to be very high. In fact, the correlation between physician-reported effectiveness and detailing is only 0.18, especially low considering common source bias likely exists. The correlation between detailing and market share, as well as effectiveness and market share, are 0.33 and 0.35, respectively. Coupled with the low correlation between detailing and effectiveness, the data indicate both detailing and effectiveness are valuable in explaining prescription writing. Consistent with the findings in this study, and since a survey would be required to acquire either physician perception or competitor detailing data, firms would be well advised to collect both types of competitor data when possible.

11. CONTRIBUTIONS

CRM research has been built primarily by considering customer databases consisting of transactional data between the firm and its customers. Interactions between the firm's customers and competing firms have largely been ignored. Our research answers the call for incorporating competitor data into CRM. When the competition is ignored, estimates of the impact of marketing efforts on firm sales can be biased, leading to poor marketing allocation decisions.

In our sample, ignoring the physicians' perceptions of the effectiveness of the various brands and the level of competitor detailing results in statistically significant bias in the estimated response to the firm's detailing efforts for 36% of the physicians. A reallocation of detailing based on the results of a model that includes physician brand effectiveness perceptions and competitor detailing indicate a 9% increase in prescriptions written. Since competitor data can typically be acquired only for a subset of the firm's customers, a data augmentation method is presented and shown to outperform analyses utilizing only firm data, accurately segmenting 96% of the physicians where competitor data is unavailable.

Obviously, this research investigates a single drug category in a single firm. Future research needs to replicate these results across categories and firms. Applying this approach to firms in data poor industries will provide challenging and potentially valuable research opportunities.

12. APPENDIX: PHYSICIAN SURVEY

UNIVERSITY OF HOUSTON – STUDY OF "CATEGORY" MEDICATIONS

- THIS ONE-PAGE SURVEY DEALS WITH SEVERAL DRUGS USED IN THE TREATMENT OF "CATEGORY".
- PLEASE FILL OUT THIS SHORT SURVEY IN ITS ENTIRETY TO RECEIVE YOUR HONORARIUM OF $X.
- WE INSURE THAT YOUR PERSONAL INFORMATION WILL BE HELD COMPLETELY CONFIDENTIAL.

Product Usage: Please provide the number of prescriptions that you write for each of the following drugs in an average (typical) month.

(Please enter a number in each box)

Product Usage (Rx)	**Brand A**	**Brand B**	**Brand C**	**Brand D**
Number of Rx's for this drug in an average month.				

Product Perceptions and Usage: Based on your knowledge and experience with "Category" medications, rate each of the following drugs in each of the following areas on a scale from 1 to 5, where 1 = poor, 2 = fair, 3 = good, 4 = very good, and 5 = excellent.

(Please circle one number in each box)

Reputation Dimensions:	**Brand A**	**Brand B**	**Brand C**	**Brand D**
Effective in treating "Symptom #1"	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Effective in treating "Symptom #2"	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Effective in treating "Symptom #3"	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Low incidence of side effects	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
Value for the money	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5

Promotional/Marketing Activity: Please provide the indicated promotional activity for each of the drugs.

Activity	**Brand A**	**Brand B**	**Brand C**	**Brand D**
Were you detailed on this drug in October? (please circle)	Yes No	Yes No	Yes No	Yes No
If yes above, how many details did you receive for this drug in October?				
What is the average (typical) number of details you receive for this drug in a month?				
What is the average (typical) number of journal advertisements you see for this drug in a month?				

- PLEASE ENTER YOUR NAME AND ADDRESS BELOW SO WE CAN DELIVER YOUR $X HONORARIUM.
- PLEASE RETURN THE COMPLETED SURVEY VIA FAX TO 555-555-5555.
- BACKUP FAX NUMBER IS 555-555-5555. THANK YOU FOR COMPLETING THE SURVEY!

Your Name: ______________________________ (Will appear on check)

Address: ______________________________

City: ______________ State: __________ Zip: __________

Your participation in this study is voluntary.

13. REFERENCES

Anderson, James C. and James A. Narus (2003), "Selectively Pursuing More of Your Customer's Business," *MIT Sloan Management Review*, 44 (3), 42-49.

Bell, David, John Deighton, Werner J. Reinartz, Roland T. Rust, and Gordon Swartz (2002), "Seven Barriers to Customer Equity Management," *Journal of Service Research*, 5 (1), 77-85.

Berger, Paul D. and Nada I. Nasr (1998), "Customer Lifetime Value: Marketing Models and Applications," *Journal of Interactive Marketing*, 12 (1), 17-30.

Bhattacharya, C. B., Peter S. Fader, Leonard M. Lodish, and Wayne S. DeSarbo (1996), "The Relationship Between the Marketing and Share of Category Requirements," *Marketing Letters*, 7 (1), 5-18.

Boulding, W., R. Staelin, M. Ehret, and W. J. Johnston (2005), "A Customer Relationship Management Roadmap: What Is Known, Potential Pitfalls, and Where to Go," *Journal of Marketing*, 69 (4), 155-66.

Bowman, Douglas and Das Narayandas (2004), "Linking Customer Management Effort to Customer Profitability in Business Markets," *Journal of Marketing Research*, 41 (4), 433-47.

Cameron, A. Colin and Pravin K. Trivedi (1998), *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Cao, Yong and Thomas S. Gruca (2005), "Reducing Adverse Selection Through Customer Relationship Management," *Journal of Marketing*, 69 (4), 219-29.

Cooper, Lee G. and Masao Nakanishi (1988), *Market-Share Analysis: Evaluating Competitive Marketing Effectiveness*. Boston: Kluwer Academic Publishers.

DeSarbo, Wayne S., Alexandru M. Degeratu, Michael J. Ahearne, and M. Kim Saxton (2002), "Disaggregate Market Share Response Models," *International Journal of Research in Marketing*, 19 (3), 253-66.

Drozdenko, R. G. and P. D. Drake (2002), *Optimal Database Marketing: Strategy, Development, and Data Mining*. Thousand Oaks, California: Sage Publications Inc.

Du, Rex Yuxing, Wagner A. Kamakura, and Carl F. Mela (2007), "Size and Share of Customer Wallet," *Journal of Marketing*, 71 (2), 94-113.

Gonul, Fusun F., Franklin Carter, Elina Petrova, and Kannan Srinivasan (2001), "Promotion of Prescription Drugs and Its Impact on Physicians' Choice Behavior," *Journal of Marketing*, 65 (3), 79-90.

Greene, William H. (2007), *LIMDEP Version 9.0 Econometric Modeling Guide*: Econometric Software, Plainview, NY.

Gupta, Sachin and Pradeep K. Chintagunta (1994), "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models," *Journal of Marketing Research*, 31 (1), 128-36.

Kamakura, Wagner A. and Michel Wedel (2003), "List Augmentation with Model Based Multiple Imputation: A Case Study Using a Mixed-Outcome Factor Model," *Statistica Neerlandica*, 57 (1), 46-57.

Kohli, Ajay K. and Bernard J. Jaworski (1990), "Market Orientation: The Construct, Research Propositions, and Managerial Implications," *Journal of Marketing*, 54 (2), 1-18.

Kotler, Philip (1971), *Marketing Decision Making: A Model Building Approach*. New York: Holt, Rinehart and Winston.

Kotler, Philip and Gary Armstrong (1994), *Principles of Marketing*, 6th ed. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Leeflang, Peter S. H., Dick R. Wittink, Michel Wedel, and Philippe A. Naert (2000), *Building Models for Marketing Decisions*. Boston: Kluwer Academic Publishers.

Little, Roderick J. A. and Donald B. Rubin (2002), *Statistical Analysis With Missing Data*, 2nd ed. Hoboken, New Jersey: John Wiley & Sons, Inc.

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.

Manchanda, Puneet, Peter E. Rossi, and Pradeep K. Chintagunta (2004), "Response Modeling with Nonrandom Marketing-Mix Variables," *Journal of Marketing Research*, 41 (4), 467-78.

Manchanda, Puneet, Dick Wittink, Andrew Ching, Paris Cleanthous, Min Ding, Xiaojing Dong, Peter Leeflang, Sanjog Misra, Natalie Mizik, Sridhar Narayanan, Thomas Steenburgh, Jaap Wieringa, Marta Wosinska, and Ying Xie (2005), "Understanding Firm, Physician and Consumer Choice Behavior in the Pharmaceutical Industry," *Marketing Letters*, 16 (3/4), 293-308.

Mizik, Natalie and Robert Jacobson (2004), "Are Physicians "Easy Marks"? Quantifying the Effects of Detailing and Sampling on New Prescriptions," *Management Science*, 50 (12), 1704-15.

Moon, Sangkil, Wagner A. Kamakura, and Johannes Ledolter (2007), "Estimating Promotion Response When Competitive Promotions Are Unobservable," *Journal of Marketing Research*, 44 (3), 503-15.

Narver, John C. and Stanley F. Slater (1990), "The effect of a market orientation on business profitability," *Journal of Marketing*, 54 (4), 20-35.

Reinartz, Werner J. and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (1), 77-99.

Reinartz, Werner, Jacquelyn S. Thomas, and V. Kumar (2005), "Balancing Acquisition and Retention Resources to Maximize Customer Profitability," *Journal of Marketing*, 69 (1), 63-79.

Ryals, Lynette (2005), "Making Customer Relationship Management Work: The Measurement and Profitable Management of Customer Relationships," *Journal of Marketing*, 69 (4), 252-61.

Sargan, J. D. (1958), "The Estimation of Economic Relationships using Instrumental Variables," *Econometrica*, 26 (3), 393-415.

Venkataraman, Sriram and Stefan Stremersch (2007), "The Debate on Influencing Doctors' Decisions: Are Drug Characteristics the Missing Link?," *Management Science*, 53 (11), 1688-701.

Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, 2nd ed. Boston: Kluwer Academic Publishers.

Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

Zeithaml, Valarie A. (1985), "The New Demographics and Market Fragmentation," *Journal of Marketing*, 49 (3).