Heterogeneous Learning and the Targeting of Marketing Communication for New Products

Sridhar Narayanan
Puneet Manchanda[1]

August 2005
This version: December 2007

[1] Sridhar Narayanan is Assistant Professor of Marketing at the Graduate School of Business, Stanford University and Puneet Manchanda is Associate Professor of Marketing at the Ross School of Business, University of Michigan. The authors would like to thank Pradeep Chintagunta, Peter Rossi, Guenter Hitsch, seminar participants at Cornell, Dartmouth, Erasmus, Emory, Harvard, Hong Kong University of Science and Technology, Minnesota, New York University, Northwestern, Purdue, Stanford, Stony Brook (Economics), SUNY-Buffalo, UCLA, University of Southern California, University of Texas at Dallas, Washington University at St. Louis, University of Wisconsin, University of Washington, Yale University and participants at the Advanced Research Techniques Forum (2005), Marketing Dynamics Conference (2005) and Bayesian Methods in Marketing Conference at Linz (2005). They would also like to thank ImpactRx and Pat Angelastro for providing the data. Manchanda would like to acknowledge support from the Kilts Center for Marketing and the Beatrice Foods Faculty Fund at the University of Chicago. All correspondence may be addressed to the first author at Stanford University, Graduate School of Business, 518 Memorial Way, Stanford, CA 94305 (Tel: 650-723-9675, Fax: 650-725-7979) or via email at sridhar.narayanan@stanford.edu.

Heterogeneous Learning and the Targeting of Marketing Communication for New Products

Abstract

New product launches are often accompanied by extensive marketing communication campaigns. Firms' allocation decisions for these marketing communication expenditures have two dimensions – across consumers and over time. This allocation problem is different relative to the problem of allocation of resources for existing products. This is because in the case of new products, consumers are uncertain about the quality of new products and learn about them through marketing communication. Further, different consumers may have different rates of learning about product quality, i.e. there may be heterogeneous learning. Thus, consumer responsiveness to marketing communication could vary along two dimensions. For each consumer, this responsiveness would vary over time, as she learns about product quality. Across consumers, there would be differences in responsiveness in each time period. For optimal allocation of marketing communication across both consumers and time, firms would need estimates of how responsiveness to marketing communication varies across consumers and over time.

Past studies in this area have typically studied one of these two dimensions in which responsiveness varies. They have either looked at heterogeneity in responsiveness across agents or the variation in responsiveness over time. In the context of new products, past research has looked at how consumer learning about product quality causes responsiveness to vary over time. In this study, we build a model that allows for heterogeneous learning rates and obtain individual-level learning parameters for each consumer. We use a novel and rich panel dataset that allows us to estimate these parameters of the model.

To obtain individual-level estimates of learning, we add a hierarchical Bayesian structure to the Bayesian learning model. We exploit the natural hierarchy in the Bayesian learning process to incorporate it within the hierarchical Bayesian model. We use data augmentation, coupled with the Metropolis Hastings algorithm to make inferences about individual-level parameters of learning. We conduct this analysis on a unique panel dataset of physicians, where we observe prescription decisions and detailing (sales force effort) at the individual physician-level for a new prescription drug category.

Our results show that there is significant heterogeneity across physicians in their rates of learning about the quality of new drugs. We also find that there are asymmetries in the temporal evolution of responsiveness of physicians to detailing – physicians who are more responsive to detailing in early periods are less responsive later on and vice versa. These finding have interesting implications for targeting of detailing across physicians and over time. We find that firms could increase their revenues if they took these temporal and cross-sectional differences in responsiveness into account while deciding their allocations of detailing.

Keywords: Resource Allocation, Pharmaceutical Markets, Learning Models, Markov Chain Monte Carlo Methods

1. Introduction

New products are the lifeblood of firm performance – they account for about a quarter of all sales and revenue growth. Firms also spend about half their marketing budgets promoting new products (see Urban and Hauser 1993). A major concern of firms is the allocation of marketing resources during the launch and rollout of a new product. The mechanism that governs this resource allocation over time is the response of the market to this new product. This response is driven by considerable uncertainty about the product quality of the new product. A major role of marketing activity is to disseminate information about new products in a manner such that this uncertainty is reduced. Consumers typically use this information to reduce their uncertainty about a new product's quality via a "learning" process. Besides the evolution of this learning process over time, consumers may also differ in their learning behavior i.e., there may be heterogeneous learning. For example, some consumers may learn faster than others about the quality of these new products. Firms could then use their knowledge of this heterogeneous learning to allocate their marketing resources both over time and across consumers. If the extent of this heterogeneity is "large," then firms stand to gain significantly if they can incorporate this knowledge into their product rollouts.

However, previous literature has typically calibrated this learning process for the market as a whole. In this research, we use unique data on the launch of new (ethical) pharmaceutical drugs and a novel methodology to calibrate learning processes for each individual (consumer) physician. Our approach is able to estimate the individual rate of learning for each physician after controlling for other behaviors such as risk aversion. We then use these estimated rates to examine firms' marketing resource allocation during the time of product rollout.

The pharmaceutical industry in general is particularly well suited to study this problem for three reasons. First, uncertainty about drug quality is particularly relevant in the case of prescription drugs. In spite of an extensive process of clinical trials before the launch of new drugs, there is still considerable uncertainty about their quality, their side effects and other associated risks. For instance, the Food and Drug Administration (FDA), without whose approval prescription drugs cannot be marketed, has the following to say about new drugs (CDER 2000) -

> "The practical size of pre-marketing clinical trials means that we cannot learn everything about the safety of a drug before we approve it. Therefore, a degree of uncertainty always exists about the risks of drugs."

Furthermore, while the pre-marketing clinical trials provide information at the mean level (efficacy, incidence of side effects etc.), physicians with heterogeneous patient bases are often uncertain about the quality of a new drug for their specific patient bases.

Second, pharmaceutical firms spend a large amount of money on marketing communication directed towards physicians. For example, the industry spent $ 8.5 billion on marketing communication directed at physicians in 2000 (Wittink 2002, Neslin 2001). Most of this is spent on detailing, which involves personal sales calls made by salespersons of the pharmaceutical firm on physicians. Third, since detailing is a personal interaction between a physician and the firm's representatives, it is allocated at the individual physician level. Firms need to decide how many calls to make to each individual physician and when to make these calls. These three factors - true uncertainty about a new product, large expenditure and individual level allocation make the pharmaceutical industry an excellent setting for our problem.

In this study, we use a unique panel dataset of physicians, which contains prescription and detailing information for all drugs in a category, to estimate individual physician-level effects in the presence of learning. We develop a methodology to estimate individual physician level effects by specifying a Bayesian learning model for each physician and estimating the model in a hierarchical Bayesian framework using Markov Chain Monte Carlo methods. Our results show that there is considerable heterogeneity in learning rates across physicians. We then show that this heterogeneity is economically significant for the firm and accounting for it can lead to considerable revenue gain for the firm. Specifically, heterogeneity in learning accounts for about 50-56% of the revenue gain, with the rest coming from heterogeneity in the other individual-level effects.

The rest of the paper is organized as follows. We first discuss the extant literature related to this study briefly. Then we discuss the data. We then present the model in detail and discuss our estimation strategy. Subsequently, we discuss the identification of our model parameters. We then discuss the results of our analysis and present the results from our counterfactual simulations. Finally, we conclude, discussing the model, the results and the limitations of the study.

2. Related Literature

This study is broadly related to three streams of research in the literature – pharmaceutical promotions, learning and targeted marketing. We discuss these briefly and point out the contribution of this study relative to these streams.

While there have been many studies that have looked at pharmaceutical promotions, none of these have looked at targeting of these pharmaceutical promotions in the presence of learning. . Early studies in the marketing literature (Parsons and Vanden Abeele, 1981; Lilien et al. 1981) studied the effect of sales force effort on sales using aggregate data. Recent research (Kamakura et al. 2004; Manchanda and Chintagunta 2004; Gönül et al. 2001 and Manchanda et al. 2004a) has used panel data to investigate the effect of detailing on pharmaceutical demand (see Manchanda and Honka 2005 for an extensive review of detailing studies). Other research has also investigated the role of pharmaceutical promotion and classified it as informative and persuasive (Leffler 1981; Hurwitz and Caves 1988; Rizzo 1999; Narayanan et al. 2005). The broad consensus in this literature is that detailing affects prescriptions by physicians positively and that there is heterogeneity in physician response to detailing.

The second stream of research that is relevant to this study is the literature on Bayesian learning (good examples of early studies are Stoneman 1981, Meyer and Sathi 1985 and Roberts and Urban 1988).[2] Erdem and Keane (1996) was a pioneering paper in marketing that used a model of Bayesian learning to incorporate informative effects of advertising. Since then, there has been a growing interest in problems involving learning (Crawford and Shum 2005, Coscelli and Shum 2004, Anand and Shachar 2004, Ackerberg 2003, Narayanan et al. 2005 and Byzalov and Shachar 2004). These studies find evidence that there is significant amount of learning and uncertainty reduction through advertising and product experience. Ackerberg (2003), Byzalov and Shachar (2004) and Narayanan et al. (2005) specifically address the issue of informative and non-informative roles of advertising or promotional activity. Narayanan et al. (2005) find evidence for the presence of both these roles, while the other studies find evidence for only the informative effect. Heilman et al. (2000) and Ackerberg (2001) also suggest that the role of marketing communication is different for new and familiar products.

Targeted promotions have interested researchers in marketing in recent years. Targeting promotions to segments of consumers has long been an industry practice as well as a topic for research. Numerous studies have looked at price discrimination (cf. Villas-Boas 1999; Fudenberg

[2] See Lilien (1974a, b) for a model that does not impose a specific functional form on the learning process.

and Tirole 2000), targeted coupons (Shaffer and Zhang 1995) and the targeting of advertising (cf. Iyer et al. 2005) to different segments of consumers. Rossi et al. (1996) and Manchanda et al. (2004a) show that firms can obtain significant benefits by targeting their promotions.

This paper builds on the literature on pharmaceutical promotion by explicitly modeling the process by which detailing affects individual physician prescription behavior from the launch of a product (drug). This is the first study to document heterogeneity in both the informative and persuasive effects of marketing communication such as detailing. In addition, this study documents, for the first time, the effect of patient influence on physicians' prescription behavior. The paper also adds to the literature on Bayesian learning models by proposing a methodology to estimate learning rates at the individual level, as opposed to the extant literature that assumes that all agents learn at the same rate. Finally, this paper contributes to the research on targeted marketing by providing insights on targeting in early stages after the launch of a product. Thus, this paper contributes both substantively and methodologically to the literature in marketing.

3. Data

The data used for the empirical analysis in this study is from the category of prescription drugs known as 'Erectile Dysfunction (ED) Drugs'. The drugs in this category are prescribed to treat ED amongst adult men – a condition that affects between 15 and 30 million men in the United States. There is only one category of oral drugs that can treat this condition and currently there are three drugs that have been approved by the FDA. The three drugs are Viagra (marketed by Pfizer and approved by FDA in March 1998), Levitra (marketed jointly by GSK Pharmaceuticals and Bayer and approved in August 2003) and Cialis (marketed by Eli Lilly and approved in November 2003). To date, no further drugs have been approved after Cialis.

The data are at the physician level and consist of a panel of 900 physicians in the United States. These data were obtained from a New Jersey based data-vending firm - ImpactRx - that has set up this panel and sells the data to pharmaceutical firms. The panel is a representative sample of the universe of physicians, balanced across geographic regions, physician specialties and prescription volume. For these physicians, we have observations of prescriptions written by them for their patients and also of detailing calls made by representatives of pharmaceutical firms (detailers). These data are collected directly from the physician, using a Personal Digital

Assistant. We have 15320 prescription observations and 16700 observations of detailing calls in the dataset.

We have specialty information (e.g., General Practitioner, Urologist) for each physician. In each prescription observation, we observe the drug that was prescribed to the patient. Unlike most existing data sources, which collect prescription data from pharmacies or insurance companies, our dataset actually captures the physician's decision. This is because pharmacy data, for instance, may not be able to capture the fact that the drug actually filled out in the pharmacy may sometimes be different from that prescribed by the physician. Thus, this captures the physician's *intended* prescription and is not susceptible to "slippage" or missing data. A unique feature of the prescription data in our dataset is that we observe if a patient requested a drug. This is recorded by the physician at the time of the consultation by the patient. In terms of detailing, the physician records the drug that is detailed in that call on every detailed visit.

In terms of temporal patterns, about 90% of physicians are detailed at least once for the new drugs within three months of launch. The mean number of calls after which Levitra and Cialis are adopted is 2.8 and 2.4 and the correlation in adoption "calls" across physicians for the two drugs is 0.48. By the end of the data, 490 (55%) of physicians have prescribed all three drugs at least once, while 159 (18%) physicians prescribed only one drug. The mean number (standard deviation) of total new prescriptions written by a physician was 8.95 (18.22), 12.26 (19.60) and 21.39 (30.10) for Cialis, Levitra and Viagra respectively resulting in overall market shares of 21.0%, 28.8% and 50.2%. By the last month of the data, however, the share pattern had changed significantly and was 38.1%, 33.4% and 28.5% for Cialis, Levitra and Viagra respectively. The mean number (standard deviation) of total details received by a physician was 5.55 (6.14), 7.65 (7.64) and 5.68 (6.12) for Cialis, Levitra and Viagra respectively. The average correlation between total prescriptions and total details across physicians was 0.38 (Cialis), 0.42 (Levitra) and 0.32 (Viagra). The number of patient requests (i.e., when a patient specifically asks for a drug to be prescribed) was relatively small – on average patients requested any one of the three drugs only 15.4% of the time. The largest number of requests in aggregate is for Viagra (51.7% of the total requests), followed by Levitra (25.5%) and Cialis (22.8%). However, in the last month in which data are available, these proportions were 32.6% (Viagra), 21.1% (Levitra) and 46.3% (Cialis).

4. Model

4.1 Model Development

The paper that is closest to ours, both in terms of the empirical setting and the model itself, is that in Crawford and Shum (2005) [henceforth CS]. We base our model on the CS model, which in turn is an extension of learning models previously proposed in the literature (Erdem and Keane 1996 and others). We shall first describe the CS model and then point out the similarities and differences of our approach with respect to that model.

CS was one of the early attempts to understand decisions made under uncertainty in the context of prescription drugs. They observe a sequence of prescription decisions for a panel of patients, for the category of anti-ulcer drugs. These drugs provide both symptomatic and curative benefits, i.e. they mitigate the symptoms of the conditions for which they are prescribed and also cure the condition, such that the patients can discontinue treatment. CS model learning about the specific patient-drug match, both with respect to the symptomatic benefit of the drug, as well as the curative benefit. They assume that patients are uncertain about these match values, but learn about them in a Bayesian manner through consumption experience. Further, patients are assumed to be forward looking, i.e. they maximize their expected stream of future utilities in choosing a drug. In their setup, risk averse patients compare their expected utilities across different drugs and choose the drug that provides them the highest expected utility. Patients are also assumed to start out with rational expectations, i.e. they know the population distribution of match values, but don't know their specific match value.

First we describe our model assumptions that are similar to those in CS. Like CS, we assume that at every prescription occasion, physicians maximize a utility function to decide which drug to prescribe. We similarly abstract away from any agency issues involved in this decision and assume that physicians represent the physician-patient pair perfectly. The physicians in our model are risk-averse agents, who are uncertain about the qualities of the new drugs and maximize expected utilities to make the drug choice at every prescription occasion. Physicians are assumed to be Bayesian updaters, i.e. they are assumed to have some prior beliefs about the drug qualities, and they are assumed to update these prior beliefs with any new information they receive using Bayes rule to obtain posterior beliefs. They are assumed to use the most recent posterior beliefs when evaluating their expected utilities for the choice alternatives. These posterior beliefs are in turn the prior beliefs for the subsequent updation occasion. Another common assumption between the two studies is that physicians know the

overall quality distribution for a drug across all patients at the initial time period, though they do not know the quality of the drug for their specific patient bases. Finally, like in CS, we also have a conditional-choice situation – we observe a patient in our dataset only if he decides to take a drug and hence, we do not model the patient's decision of whether to seek drug therapy or not.

Next, we describe some significant differences in our model assumptions relative to those in CS. First, the focus of CS is on learning by individual patients about their match values for drugs. By contrast, our focus is on learning by individual physicians about the mean quality of the drugs for their patients. Our data are at the physician-level, while those of CS are at the patient-level. Thus, we observe the sequence of prescription decisions by a specific physician, but cannot track specific patients. CS, on the other hand, have data on the sequence of visits by a particular patient, but do not observe the sequence of prescriptions by a physician. This suggests that our data are well suited to study learning about overall quality of a drug, while theirs are well suited to study learning about patient-specific match values. A second difference of our approach relative to CS is that in their case, the drug can cure the disease as well as provide symptomatic benefits, while in our case, the drug can only provide symptomatic benefits. Thus, while they model learning about curative as well as symptomatic match values, we model learning only about an overall symptomatic quality of the drug. A third difference is that CS model learning only through the patient's own experience and feedback to the physician, while we model learning both through feedback as well as marketing communication directed at the physician. We are thus able to address the problem of optimal resource allocation for physician-directed marketing communication by pharmaceutical firms, especially in the context of new drugs. Another difference is that while the CS model assumes forward-looking agents, we assume that agents are myopic.[3] The most important difference in our approach relative to CS is that they assume *that all physicians learn at the same rate*. In the Bayesian learning model used in both CS and this paper, the rate at which physicians learn is summarized by the variances of the signals through which they update their beliefs about the drugs. We allow the signal variances to be different for different physicians, thus allowing for different rates of learning for different physicians. On the other hand, in the CS approach, all physicians have the same signal variance and thus learn at the same rate, all else remaining equal. A key contribution of this paper is that we model heterogeneous learning and develop a general Bayesian hierarchical approach (using

[3] Our model without this assumption is analytically intractable. This is because the state-space for a model without this assumption is very large. For our problem, it will be 7200 – the product of 900 physicians and 8 state variables.

Markov chain Monte Carlo methods) to estimate learning rates at the lowest level of aggregation (individual physicians in our case).

4.2 Model Specification

When physician *i* has to make a decision on which drug to prescribe at occasion *t*, she chooses the alternative *j* that provides the greatest utility, with the utility function defined as

$$\tilde{U}_{ijt} = -\exp\left(-r_i \tilde{Q}_{ijt}\right) + X_{ijt}\beta_i + \varepsilon_{ijt} \quad (1)$$

where

$\tilde{Q}_{ijt}$ is physician *i*'s belief about the true quality of drug *j* at time *t* and is stochastic from the point of view of the physician

r_i is the coefficient of absolute risk aversion

X_{ijt} is a row vector (*1 x K*) of physician, drug and time (patient) specific variables[4],

β_i is a column vector (*K x 1*) of physician specific sensitivity to these variables

ε_{ijt} is an i.i.d. physician, drug and time specific shock

This utility function, which resembles the one in CS, has a sub-utility function in drug quality that has the constant absolute risk aversion (CARA) form. Thus, the underlying assumption is that physicians have a constant absolute risk aversion with respect to the uncertainty in drug quality. It is linear in drug and patient specific variables and in an additive shock that includes the match value between a drug and a specific patient. As we describe later, our dataset includes only the new prescriptions to patient and hence, the i.i.d. assumption for this shock is a reasonable one. Note that since we have an observation every time a patient walks into the office and since we do not have repeated observations for the same patient, the prescription occasion *t* is identical to patient *t* for the physician.

This utility is stochastic from the physician's perspective because the belief about the quality $\tilde{Q}_{ijt}$ is stochastic. But it must be remembered that ε_{ijt} is not stochastic from the point of view of the physician (i.e., the physician observes the ε_{ijt} on each prescription occasion). The physician is assumed to be an expected utility maximizer. This expected utility is

[4] The set of variables included in this vector are described later in this section (equation 12).

$$U_{ijt} = E\left[\tilde{U}_{ijt}\right] = E\left[-\exp\left(-r_i\tilde{Q}_{ijt}\right)\right] + X_{ijt}\beta_i + \varepsilon_{ijt} \tag{2}$$

We allow the true mean qualities of the drugs to be heterogeneous across physicians. This is analogous to the heterogeneous true match values in the CS model. In our case, these mean qualities represent the average qualities of the drugs for the specific patient base of each physician and are heterogeneous because patient bases are different across physicians.

Physicians are assumed to update their quality belief in each period based on signals they receive through detailing and through patient feedback (via past prescriptions). We make the assumption that prescriptions written by a physician generates a feedback from the patient after a period of 30 days. This is based on the fact that about 98% of all prescriptions in the ED category are for 30 days or less (we obtained this number from the Pharmtrends database maintained by IPSOS North America). The detailing and feedback signals are assumed to be normally distributed around the physician-specific true mean quality of the drug.

Assume that there are nd_{ijt} detailing signals at time *t* and the m^{th} signal is assumed to be given as

$$\tilde{D}_{ijtm} \sim N\left(Q_{ij}, \sigma^2_{D_i}\right) \tag{3}$$

and that there are nf_{jt} patient feedback signals, and the m^{th} signal is given by

$$\tilde{F}_{ijtm} \sim N\left(Q_{ij}, \sigma^2_{F_i}\right) \tag{4}$$

A series of unobserved signals that are normally distributed can be summarized by their sample mean, which is also normally distributed. We define these sample means as follows

$$\tilde{D}_{ijt} = \frac{\sum_m \tilde{D}_{ijtm}}{nd_{ijt}} \sim N\left(Q_{ij}, \frac{\sigma^2_{D_i}}{nd_{ijt}}\right) \tag{5}$$

$$\tilde{F}_{ijt} = \frac{\sum_m \tilde{F}_{ijtm}}{nf_{ijt}} \sim N\left(Q_{ij}, \frac{\sigma^2_{F_i}}{nf_{ijt}}\right) \tag{6}$$

The quality belief at time t=0 is assumed to be a normal distribution whose mean is the mean of the population distribution of the true quality and whose variance is the variance of this

population distribution. Thus, the assumption is that physicians know the distribution of the true quality across all physicians but are uncertain about the true quality for their own patient base.

As described earlier, and similar to CS, the physician is assumed to update her beliefs in a Bayesian manner, i.e. at any given period of time, she combines her prior belief about the quality of the drug with the information obtained through both detailing and feedback signals and applies Bayes Rule to form her posterior belief. Since the prior belief at time t=0 and all signals are assumed to be normally distributed, it turns out that the posterior belief at every time period is also a normal distribution. This posterior belief is given by

$$\tilde{Q}_{ijt} \sim N\left(Q_{ijt}, \sigma^2_{Q_{ijt}}\right) \tag{7}$$

where

$$Q_{ijt} = \frac{\sigma^2_{Qijt}}{\sigma^2_{Qij(t-1)}} Q_{ij(t-1)} + nf_{ijt} \frac{\sigma^2_{Qijt}}{\sigma^2_{F_i}} \tilde{F}_{ijt} + nd_{ijt} \frac{\sigma^2_{Qijt}}{\sigma^2_{D_i}} \tilde{D}_{ijt} \tag{8}$$

and

$$\sigma^2_{Q_{ijt}} = \frac{1}{\frac{1}{\sigma^2_{Q_{ij(t-1)}}} + \frac{nd_{ijt}}{\sigma^2_{D_i}} + \frac{nf_{jt}}{\sigma^2_{F_i}}} \tag{9}$$

It is important at this stage to point out how heterogeneity in learning manifests itself in the model. Note that the variances of the detailing and feedback signals ($\sigma^2_{D_i}$ and $\sigma^2_{F_i}$ respectively), are physician specific. It can be seen from equation (8) that for a physician for whom $\sigma^2_{D_i}$ is a large value, relatively lower weight is placed on the detailing signal $\tilde{D}_{ijt}$ than for a physician with a low value of $\sigma^2_{D_i}$, everything else remaining the same. The former physician is a slow learner from detailing than the latter. Similarly, a physician with a higher value of $\sigma^2_{F_i}$ would be a slower learner from feedback than a physician with a lower value of this parameter. Thus, the variances of the detailing and feedback signals summarize the heterogeneity in learning across physicians. By contrast, in traditional learning models, these variances are assumed to be homogenous across agents. Allowing for these variances to be individual specific is a key contribution of our approach.

Given that the quality belief in any period is a normal distribution with mean Q_{ijt} and variance $\sigma^2_{Q_{ijt}}$ and r_i is the coefficient of absolute risk aversion, the expected utility of the physician is

$$U_{ijt} = E\left[\tilde{U}_{ijt}\right] = -\exp\left(-r_i Q_{ijt} + \frac{r_i^2 \sigma^2_{Q_{ijt}}}{2}\right) + X_{ijt}\beta_i + \varepsilon_{ijt} \quad (10)$$

We next describe what variables are included in the linear subutility function ($X_{ijt}\beta_i$).

In our model, we have already described how detailing affects the prescription decision of the physician, through its effect on learning about drug quality. This effect is referred to as the informative effect of detailing or sometimes as its indirect effect (Ackerberg 2004; Narayanan, Manchanda and Chintagunta 2005). It can be seen from equation (8) that the effect of detailing on the physician's quality belief is highest initially and reduces with every subsequent updation. This can be seen from the fact that the variance of the detailing signal $\sigma^2_{Q_{ijt}}$ converges towards zero with every updation and thus, the coefficient of the detailing signal in equation (8) also converges to zero. Thus, the informative effect of detailing is highest at the introductory phase of a drug and after the physician has learnt about the drug and reduced her uncertainty, this effect is negligible.

However, it is well documented that detailing has an effect on physicians' prescription behavior even in the case of mature products (cf. Gönül et al 2001; Manchanda and Chintagunta 2004). It has been suggested that this effect in the case of mature products may be because of an image or prestige role of detailing, or perhaps due to reminder effects. We shall refer to all effects of detailing except the informative effect defined in the previous section as the persuasive effect of detailing. As noted earlier, these terms – *informative effect* and *persuasive effect* – are labels to differentiate between the two effects and are not meant to convey that one effect exclusively involves information and the other persuasion. The inclusion of these effects of detailing is a significant difference from CS, in which marketing communication is not part of the model.

The persuasive effect of detailing is captured in a reduced form manner by including a stock of detailing counts for the 30-day period preceding the prescription occasion in the linear

X_{ijt} variable in the utility function (equation 1). The coefficient of this variable measures the persuasive effect of detailing. This effect would capture any role of detailing that remains unchanged over the product life cycle of the drug. This approach is similar to that adopted in the literature (Ackerberg 2003; Anand and Shachar 2004; Narayanan, Manchanda and Chintagunta 2005).

Our dataset allows us to account for patient influence on the prescription decision of the physician. We observe in the data if the patient requested a specific drug or not. We allow a dummy variable indicating whether the patient requested the drug or not in the linear X_{ijt} variable in the utility function in equation (1). The coefficient of this variable captures the influence of the patient's request on the prescription decision of the physician. In an indirect way, this also captures the effect of direct to consumer advertising (DTC), since DTC often asks a patient to talk to the doctor about the drug. In summary, X_{ijt} is thus given by

$$X_{ijt} = \left(\text{DetailingStock}_{ijt} \quad \text{PatientRequest}_{ijt} \right) \tag{11}$$

The objective of our empirical analysis is to suggest revenue-enhancing allocation plans for detailing. Since detailing has to be allocated at the individual physician level, the estimates of the effect of detailing also have to be obtained at the individual physician level. Past studies using Bayesian learning models, including CS have used frequentist methods for estimation, making the estimation of individual physician level parameters infeasible. A Bayesian approach is a natural way to estimate individual physician-level parameters. We provide a new modeling approach by estimating a Bayesian learning model under a Hierarchical Bayesian framework. We use Markov Chain Monte Carlo (MCMC) methods to estimate the individual-level parameters.

However, the challenge in specifying the model as a Hierarchical Bayesian Model is that the quality beliefs are unobserved. In a standard frequentist estimation of learning models (Erdem and Keane 1996), one could integrate out these unobserved quality beliefs by simulation methods. However, for using MCMC methods, we make use of a simple observation. From equation (8), it is clear that not just is the quality belief $\tilde{Q}_{ijt}$ a stochastic variable, but also its mean Q_{ijt} is stochastic. This is because, from equation (8), Q_{ijt} is a function of two stochastic variables – the realizations of the detailing and feedback signals, $\tilde{D}_{ijt}$ and $\tilde{F}_{ijt}$ respectively. Further, since these

two variables are assumed to have normal distributions, Q_{ijt} is a linear combination of normal variables, and therefore is also a normal variable.

In particular, we can derive the distribution of Q_{ijt}, conditional on $Q_{ij(t-1)}$ as

$$Q_{ijt} \mid Q_{ij(t-1)} \sim N\left(\bar{Q}_{ijt}, v_{ijt}^2\right) \tag{12}$$

where

$$\bar{Q}_{ijt} = \frac{\sigma_{Qijt}^2}{\sigma_{Qij(t-1)}^2} Q_{ij(t-1)} + \left(nf_{ijt} \frac{\sigma_{Qijt}^2}{\sigma_{F_i}^2} + nd_{ijt} \frac{\sigma_{Qijt}^2}{\sigma_{D_i}^2} \right) Q_{ij} \tag{13}$$

and

$$v_{ijt}^2 = nf_{ijt} \frac{\sigma_{Q_{ijt}}^4}{\sigma_{F_i}^2} + nd_{ijt} \frac{\sigma_{Q_{ijt}}^4}{\sigma_{D_i}^2} \tag{14}$$

Note that from equation (9), the variance of the quality belief $\sigma_{Q_{ijt}}^2$ is not a stochastic variable, conditional on the parameters of the model. Given the parameters of the model, it is known deterministically. The main difference between the mean Q_{ijt} and variance of the quality belief $\sigma_{Q_{ijt}}^2$ is that while the former depends on the (unobserved) *realizations* of the detailing signal $\tilde{D}_{ijt}$ and the feedback signal $\tilde{F}_{ijt}$, the latter does not. Hence, in a frequentist estimation of a learning model, we integrate out the mean but not the variance of the quality belief. Analogously, in a hierarchical model, we specify the mean of the quality belief as a level of the hierarchy, but not the variance.

Given that we can write the unobserved mean of the quality belief in any period as a random variable, conditional on the mean of the quality in the previous period, we thus have a natural hierarchy of quality beliefs

$$\begin{aligned} &Q_{ijt} \mid Q_{ij(t-1)} \sim N\left(\bar{Q}_{ijt}, v_{ijt}^2\right) \\ &Q_{ij(t-1)} \mid Q_{ij(t-2)} \sim N\left(\bar{Q}_{ij(t-1)}, v_{ij(t-1)}^2\right) \\ &.. \\ &Q_{ij1} \mid Q_{j0} \sim N\left(\bar{Q}_{ij1}, v_{ij1}^2\right) \end{aligned} \tag{15}$$

We can then specify the rest of the hierarchical model as follows. We assume that the prescription choice follows a probit process. Thus, the random errors ε_{ijt} of the utility function in equation (1) follow a multivariate normal distribution.

$$\begin{bmatrix} \varepsilon_{i1t} \\ . \\ \varepsilon_{iJt} \end{bmatrix} \sim MVN\left(0, \Sigma\right) \tag{16}$$

The alternative that provides the greatest utility is chosen. Hence, U_{ijt} follows a truncated multivariate distribution, conditional on choice. If choice is given by the indicator variable I_{ijt} (which is 1 if brand *j* is chosen and 0 otherwise), the truncation is such that

$$U_{ijt} > U_{ikt},\ I_{ijt} = 1, I_{ikt} = 0\ \forall k \neq j \tag{17}$$

Let the vector γ_i (dimension K x 1) denote the individual level parameters of the model. These parameters are specified as a function of physician-level characteristics, as follows

$$\gamma_i = \left[\beta_i' \quad \ln\left(\sigma_{D_i}^2\right) \quad \ln\left(\sigma_{F_i}^2\right) \quad \ln\left(r_i\right) \quad \left(Q_{i1}, \quad ... \quad , Q_{iJ}\right) \right]' \sim MVN\left(\Lambda Z_i, V_\gamma\right) \tag{18}$$

where Z_i is a M x 1 column vector of physician characteristics, including a first element which has the value 1; and Λ (K x M matrix) and V_γ (K x K matrix) are parameters.

Thus, the hierarchical model can be specified as follows

$$U_{ijt} \mid I_{ijt}, X_{ijt}, Q_{ijt}, \sigma_{Q_{ijt}}^2, r_i, \beta_i, \Sigma$$

$$Q_{ijt} \mid Q_{ij(t-1)}, nd_{ijt}, I_{ij(t-1)}, \sigma_{D_i}^2, \sigma_{F_i}^2, Q_{ij}, r_i$$

$$\gamma_i \mid \Lambda, Z_i, V_\gamma$$

In order to complete the model, the priors for the parameters are specified as follows:

$$Q_{j0} \sim N\left(\overline{Q}_0, \theta_0^2\right), \Sigma = \begin{pmatrix} \sigma_1^2 & . & . & 0 \\ . & \sigma_2^2 & . & . \\ . & . & . & . \\ 0 & . & . & 1 \end{pmatrix},\ \sigma_j^2 \sim Inverse\,Gamma\left(s_{1j}, s_{2j}\right),$$

$$\lambda = vec\left(\Lambda'\right) \sim N\left(\overline{\lambda}, V_\lambda\right),\ V_\gamma \sim Wishart(g, G)$$

In order to make draws from the joint posterior distribution of parameters, we make use of Gibbs Sampling, and make draws from the full conditional distributions of subvectors of the parameter vector. Further, we use data augmentation to draw utilities and qualities. Finally, the Metropolis Hastings algorithm is used to make draws for parameters for which the full conditional distributions are not easy to directly draw from. The full details of the likelihood, the full conditional distributions and the sampling algorithm are given in the Appendix.

5. Identification

Given the structure of our model, it is important that we provide some intuition of the identification of the individual-specific parameters of learning and the separation of the informative and persuasive effects of detailing. Additionally, our simulation results (available from the authors on request) showed that our model could recover the parameters of the model with a reasonable degree of accuracy.

The individual level parameters in the model are – the variances of the detailing and feedback signals ($\sigma^2_{D_i}$ and $\sigma^2_{F_i}$ respectively), the risk aversion parameter r_i, the true mean quality of each drug Q_{ij} and the linear coefficients β_i. It must also be noted that we need observations for multiple new drugs in order to infer that the learning rates are systematic to the physician. It would otherwise be hard to separate from random events like specific realizations of the random error ε_{ijt}, which would give us similar prescription patterns. However, if we observe the patterns of evolution of prescription behavior for the same physician for multiple drugs, we would be able to make inferences about the learning rate and thus about the parameters that summarize this learning rate ($\sigma^2_{D_i}$ and $\sigma^2_{F_i}$).

The true quality Q_{ij} of the drug is identified by the physician's steady state prescription behavior. We have already seen that as the physician learns, the quality belief Q_{ijt} evolves to the true quality of the drug Q_{ij}. At the extreme, at steady state, the quality belief is indeed the true quality. Thus, we need to observe a long enough time series of prescriptions for the physicians to be able to correctly identify the true mean quality. In the data we use for our empirical analysis, the data is observed for about 9 months after launch. Hence, the prescription behavior of the

physician towards the end of the dataset tells us about the true mean quality of the drug. One caveat to this is that we cannot test for whether the steady state has been reached or not and hence it is an assumption that 9 months is adequate to reach steady state. To see the extent to which results are affected, we estimated our model after dropping one observation per physician (a loss of 6% of data per physician on average). We find that our results (available from the authors on request) are largely unchanged.

Coming to the identification of the risk aversion parameter, we first note from equation (11) that the expected utility of the physician depends not just on the mean of the quality belief Q_{ijt} in any time period, but also on its variance $\sigma^2_{Q_{ijt}}$. We have seen earlier that this variance declines monotonically as the physician learns about the drug. It is easy to see from equation (11) that the expected utility of the physician is inversely related to this variance. Additionally, this variance is interacted with the risk aversion parameter – the higher the risk aversion, the lower will be the expected utility for a given value of this variance. The lower the expected utility, the lower would be the probability of prescribing a drug. Since all physicians are assumed to start with the same quality belief and this variance is the same for all physicians, systematic differences in shares of prescribing the new drug in early periods after introduction would be explained only by differences in risk aversion, all else being equal. Since we observe two new drugs being introduced in our data, the identification is even stronger. Physicians who are more risk averse would have lower initial probability of prescription for every new drug compared to the average physician. Thus, correlation in such behavior across initial periods after introduction of at least two new drugs is helpful in identifying an individual specific risk aversion coefficient, while it can, in principle be identified even without such observations of multiple drug introduction. In our dataset, we observe two new introductions, thus strengthening the identification of this parameter.

The variances of the detailing and feedback signals are identified from the evolution patterns of physician prescription behavior and how they are related with detailing and feedback signals. We have seen earlier that with every signal, the physician's quality belief is updated. As a result, the mean of this quality belief evolves from the mean of the initial quality belief towards the mean of the final quality belief. Equation (9) summarizes the process of updating of the mean of the quality belief. As the quality belief for a drug evolves, the probability of prescribing that drug also evolves over time. In the discussion on model specification, we have explained how a greater variance of the detailing signal implies a slower learning rate. Thus, the rate of learning

helps identify the variance parameter for the detailing signal. A similar argument holds in the case of the variance of the feedback signal.

The linear coefficients β_i are identified from the covariance of the prescription behavior of the physician with the linear variables (detailing stock variable and dummy variable for patient requests). An important concern could be about the separate identification of the informative and persuasive effects of detailing. The identification of the persuasive effect, which is the coefficient of the detailing stock variable in the utility equation (equation 1), is aided by the fact that in our dataset, the incumbent drug – Viagra – has existed for about 5 years before the first observation in the data. Hence, we make the *a priori* assumption that physicians have learnt fully about Viagra and hence the informative effect for this drug is zero. Hence, any effect of detailing of Viagra on the prescription behavior of physicians is entirely due to the persuasive effect. For the new entrants – Levitra and Cialis - both the informative and persuasive effects are present. Hence, we are able to identify both these effects.

A potential concern could be about endogeneity in the detailing allocation. Specifically, if firms optimally decide on their detailing levels to individual physicians based on a complete knowledge of their behavior, then the detailing would be endogenous and hence parameter estimates would be biased. There is an institutional feature of the pharmaceutical industry because of which this may not be a significant concern for the purpose of our study. The typical rule that firms use to decide on their detailing levels to physicians is a volumetric decile-based rule. Physicians are typically classified into deciles based on their total category volumes and detailing allocations are made at the decile levels (see Manchanda et al. (2004a) for a discussion of these allocation rules). We find that this is consistent with the patterns of detailing in our data. Figure 6 illustrates the allocation of detailing for Levitra across deciles of physicians over the first seven months after launch of the drug. It is clear that the allocation of detailing for each decile is relatively unchanged over this period. Manchanda et al. (2004b) also find similar decile-based detailing patterns by firms during a launch of a new product.

6. Results

We first present the parameter estimates of the model. In Table 1, we report the parameter estimates for the individual-level parameters for the model. The individual-level parameters are the detailing signal variance ($\sigma^2_{D_i}$), feedback signal variance ($\sigma^2_{F_i}$), the absolute risk aversion

(r_i), the coefficients for the detailing stock (β_{1i}) and patient requests (β_{2i}); and the true mean qualities for Cialis (Q_{1i}) and Levitra (Q_{2i}).In our Bayesian inference, we obtain a distribution for each individual-level parameter for each physician. We compute the mean parameter value for each physician and then report the mean and standard deviation across physicians of this individual-level mean parameter value in Table 1.

The estimate for detailing signal variance is 1.0567. This parameter value implies that it takes about 9.5103 detailing calls for the uncertainty of a physician to reduce to one-tenth of its initial value. This is an estimate of the *informative* effect of detailing. Similarly, the parameter estimate of the feedback signal variance is 1.0986, and this corresponds to 9.8874 feedback signals to reduce the uncertainty to one-tenth of its initial value. This also suggests that an average detailing call is somewhat more informative than an average past prescription since it requires a smaller number of detailing calls than feedback signals to reduce the physician's uncertainty about drug quality.[5] This is consistent with the findings of prior research (cf. Narayanan, Manchanda and Chintagunta 2005). The standard deviations for these estimates suggest that there is considerable heterogeneity across physicians in these parameters. We shall return to a discussion on heterogeneity in these parameters later in this section.

The parameter estimates for the coefficient of the linear detailing stock (β_{1i}) suggest that there is a positive *persuasive* effect of detailing. Thus, even after a physician has reduced his uncertainty about the drug, there is still a positive effect of detailing. There is also a strong positive effect of patient feedback, as indicated by the parameter estimates for the coefficient for patient requests (β_{2i}). To the best of our knowledge, this is the first time that the effect of patient requests has been quantified. Given the data description earlier, it is clear that the incidence of patient requests is very low at about 15%, but if a patient makes a request, it has a strong effect on the physician's prescription behavior.

The parameter estimates for the mean true qualities for Cialis and Levitra are both close to one (the true quality of Viagra, which we fix for identification purposes), suggesting that the qualities of the three drugs are similar, on average, albeit at the individual physician level, they may differ significantly. Cialis is on average marginally of higher quality than Viagra and

[5] Note that this conclusion is not invariant to the number of patient feedback signals per past prescription. Hence we compare the degree of information in a detailing call and in a past prescription, as opposed to that in a detailing call and in a patient feedback.

Levitra is of marginally lower quality than Viagra. This is consistent with industry reports that Cialis is a higher quality drug. Note that in steady state, the share of prescriptions of the drug with the highest true quality would be largest, all else being equal.

In Table 2, we report the parameter estimates for the pooled parameters. The pooled parameters in this model are the variances of the normal errors in the utility function (σ_1^2 and σ_2^2). The additional pooled parameters include the parameters that relate the individual-level parameters to physician demographics, i.e. λ and these are reported in Table 4. The main conclusion from this table is that while observed demographics have significant effects on the detailing stock coefficient, they are mostly not significant in the case of the other parameters.

Next, we move to the heterogeneity in the individual-level parameter estimates. Figure 1 depicts the histograms of the individual–level parameters across physicians. Figures 1(a) and 1(b) respectively show the histograms of the detailing and feedback signal variances. These figures suggest that there is considerable heterogeneity in these parameters across physicians. There is much greater heterogeneity in the detailing signal variance than the feedback signal variance. Figure 1(c) shows the distribution of the risk aversion parameter. Physicians are fairly heterogeneous in their risk aversion levels as well. We shall assess the economic significance of heterogeneity in learning as well as risk aversion in the next section.

We find from observing Figure 1(d) that for all physicians, the mean of the detailing stock coefficient is positive. Furthermore, for an overwhelming majority of these physicians, the 95% credible intervals of the individual level parameter does not include zero. Thus, it can be concluded that the detailing stock has a positive effect on their choice probabilities of the detailed brand. Figure 1(e) shows the distribution across physicians of the patient request coefficient. It is positive for all physicians and the 95% credible interval does not cross zero for any physician.

In Figures 1(f) and 1(g), we show the distributions across physicians of the true mean qualities of Cialis and Levitra respectively. Once again, on average, Cialis has a higher quality than Viagra and Levitra a lower quality than Viagra, but there is considerable heterogeneity in the true quality levels. Therefore, for a particular physician, it is not necessary that this rank ordering of drugs be maintained.

When talking about heterogeneity, a valid concern could be whether the individual-level parameters are significantly different from each other. One could compare the individual-level parameters for each pair of physicians to see if they differ from each other in statistical terms.

We present a more informal analysis of heterogeneity of these individual parameters across physicians. In Table 3, we compare the across-physician and within-physician standard deviations of these parameters. If the across-physician standard deviation for a parameter is smaller than or similar to the within-physician standard deviation, it would suggest that the 95% credible intervals for the physicians overlap and hence they are not significantly different from each other in terms of that parameter. On the other hand, if the across-physician standard deviation is larger than the within physician standard deviation, it would provide greater support to the claim that the individual-level parameters are significantly different for different physicians.

We find that the across-physician variation is much higher than the within-physician variance in the case of the two signal variances. Since the heterogeneity in the detailing signal variance summarizes the heterogeneity in learning as discussed in the model specification section, we can conclude that there is significant amount of heterogeneity in learning across physicians.

The coefficient of the detailing stock also has a higher across-physician standard deviation than the within-physician standard deviation. Thus, there is a significant amount of heterogeneity in this parameter as well. The same is the case for the true mean qualities of Cialis and Levitra. In the case of the risk aversion parameter, however, and for the coefficient for patient requests, the within-physician standard deviation is larger than the across-physician standard deviation, suggesting that physicians do not significantly differ in these parameters.

In order to explore the heterogeneity in learning across physicians even further, we plot a histogram of the number of detailing calls required to reduce the physician's uncertainty about a new drug to one-tenth its initial value. This is computed using the parameter estimates of the detailing signal variance for each physician. Figure 2, which shows this plot, suggests that heterogeneity in the detailing signal variance parameter indeed manifests itself in significant heterogeneity in the number of calls required to reduce the physicians' uncertainty.

An interesting pattern in the parameter estimates is the negative correlation between the informative and persuasive effects of detailing for physicians. This means that physicians who have a high informative effect of detailing are likely to have a low persuasive effect and vice versa. The informative effect of detailing is summarized by the detailing signal variance - a high informative effect is equivalent to saying that the learning rate for the physician is high. This manifests itself in the parameter estimates as a low value of the detailing signal variance. Similarly, a low informative effect manifests itself in a high detailing variance. The persuasive effect is summarized by the coefficient of the detailing stock variable. Thus, a physician for

whom the value of this coefficient is low would have a low persuasive effect and vice versa. Therefore, a negative correlation between the informative and persuasive effects of detailing implies a positive correlation between the detailing signal variance and the detailing stock coefficient. In Figure 3, we plot the detailing signal variance for each physician against the detailing stock coefficient to show this positive correlation between the parameters and consequently the negative correlation between the informative and persuasive effects. In Figure 4, we plot the detailing signal variance against the risk aversion parameter, to assess whether these parameters are correlated. As can be seen in the figure (and verified by a regression), the risk aversion coefficient is not significantly correlated with the detailing signal variance.

Finally, we also carry out model comparison with a series of null models – (a) a random-coefficient logit model with state-dependence, (b) the full model with the persuasive effect, but no learning, (c) the full model with learning, but no persuasive effect, (d) the full model with a common learning rate across physicians for detailing and feedback, (e) the full model with heterogeneous learning rates on feedback but a common learning rate for detailing and (f) the full model with heterogeneous learning rates on detailing but a common learning rate for feedback. As can be seen from Table 7, our model performs better on all five criteria – likelihood, Bayes factor, out-of-sample likelihood, within sample hit rates and out-of-sample hit rates. Since learning plays a particularly significant role in early periods after introduction of a new product, we should expect to see even better hit rates for our model in such early periods, relative to models that do not have learning. We report the hit rates for the first three months of the data and indeed find that our model performs considerably better than the null models without learning.

We can also gain some insights by comparing the fit statistics of the various null models. We see that the in-sample hit rates for the first three months in the sample jump significantly between null model 2 and null model 4. This is intuitive, since model 4 has learning while model 2 does not and learning helps explain the early evolution of prescription patterns. However, the fit statistics of null models 4 through 6 are comparable, while that for the full model are much higher. The main insight is that heterogeneity in learning, both through detailing and feedback gives a big improvement in fit. However, this is not achieved through adding heterogeneity in learning through either detailing or feedback. To understand this better, we look at the correlation between the detailing and feedback signal variances at the individual level and find that they are negatively correlated, with a correlation coefficient of -0.63). This suggests that gains in fit are likely to be achieved only when both individual level parameters are allowed to be

heterogeneous. This is because otherwise, the model is trying to summarize two cross-sectional heterogeneity distributions that are negatively correlated via just one distribution.[6]

7. Managerial Implications

The negative correlation between the informative and persuasive effects has implications for the allocation of detailing efforts across physicians and over time. To illustrate this, let us consider a situation with two physicians, one with a high informative effect and a low persuasive effect (a "fast" learner), and the other with the opposite (a "slow" learner). For both these physicians, the informative and persuasive effects are present in the introductory phases of the new drug's life cycle. In later stages, only the persuasive effect plays a role. For the fast learner, the total effect of detailing starts off at a very high level, but reduces rapidly till it converges to the persuasive effect and then remains constant at that level. For the slow learner, everything remaining the same, the total effect again starts at a high level (perhaps not as high as for the fast learner) but falls more slowly. It converges to a persuasive effect that is higher than for the fast learner. Since this total effect denotes the responsiveness of the physician to detailing and the optimal detailing level depends on responsiveness, it would be optimal for firms to allocate higher amounts of detailing initially to the fast learner but then rapidly reduce this allocation. In contrast, for the slow learner, we might expect to see a different rate at which optimal detailing allocation reduces.

We have also seen that the informative and persuasive effects are related to the decile of the physician. Specifically, physicians in higher deciles have a lower persuasive effect and those in lower deciles have a higher persuasive effect. Thus, if firms were to optimally allocate their resources to physicians, one would expect to see them allocating a high proportion of resources to physicians in high deciles in the early stages after the introduction of the drug and then reducing this proportion over time. However, discussions with practitioners in the industry reveal that firms do not allocate their detailing efforts in this manner. The most commonly used rule for allocating resources to physicians is a decile-based rule, which remains constant over time. We find the same patterns in our data as well, with detailing for every decile remaining approximately constant over the entire period in our dataset.

We conduct three counterfactual simulations to see how firms could increase their revenues if they took into account learning by physicians and heterogeneity in this learning. In

[6] We thank an anonymous reviewer for bring this to our notice.

each of these simulations, we conduct separate simulations for each of the two new drugs – Levitra and Cialis – one by one. We keep the total amount of detailing by firms in the first three months after the launch of the respective drug as fixed. We only alter the allocation of detailing across physicians or over time and then compute the predicted expected revenues. In all these simulations, detailing calls are rounded off, i.e., a physician can only get an integer number of calls. All these calls are assumed to be made in the beginning of the month, so as to abstract away from the problem of timing of detailing calls within a month. Competitors' detailing is kept unchanged in all these simulations. The 'optimal' allocations are obtained using a numerical optimization routine. We conduct all these simulations for a subset of 100 physicians randomly chosen from the sample. This is done in order to keep the optimization feasible. We compare the predicted revenues in the counterfactual case with those using the actual allocation plans.

In Simulation 1, we change only the temporal allocation of detailing, but keep the cross-sectional allocation unchanged, i.e. we vary the allocation of detailing to each of the three months in the launch quarter, but do not vary the proportions of detailing within that month to individual physicians. Thus, this is a two-dimensional optimization exercise, with the unknown variables being the allocation for the first two months (and the allocation of the third month automatically known from these since the proportions add up to 1). We find that by varying the month-to-month allocations of detailing, Levitra can get revenue gains of 5.9% on average compared to the current allocation plan, while Cialis could obtain an even higher gain of 8.3%. on average. This reflects the fact that firms could gain by frontloading their detailing to the period immediately after launch. This is because in this early period, all physicians have both an informative effect and a persuasive effect and are hence most responsive to detailing. As they learn about the drug, the informative effect asymptotes down to zero and hence only the persuasive effect is present. Thus, for all physicians, the responsiveness to detailing reduces over time. This result is similar to that reported in Narayanan, Manchanda and Chintagunta (2005). Note that the structure imposed by Bayesian learning causes the informative effect to decrease and it may therefore seem as if the frontloading of detailing is really driven by the structure alone. However, the empirical question that our estimates help answer is the rate at which the informative effect reduces over time and therefore the extent of frontloading.

In Simulation 2, we change the cross-sectional allocation of detailing, but keep the temporal allocation unchanged. Thus, we keep the total amount of detailing within each month fixed, but vary how much of that is allocated to each of the physicians. In each month, we have

an independent optimization, with the dimension of the unknown vector being one less than the number of physicians (i.e. 99 since the simulation was conducted for 100 physicians). We find that there are relatively modest revenue increases that can result through this exercise, 4.7% on average for Levitra and 5.8% on average for Cialis. The reason for revenue increases being lower in this case is perhaps that the firms are using some decile-based rule for allocation and while this is not optimal, it does inherently take into account heterogeneity across physicians (since the informative and persuasive effects of detailing are explained to some extent by decile). Finally, in Simulation 3, we allow both the temporal and cross-sectional allocation of detailing to change, i.e. we find the detailing level for each physician for each month that maximizes revenues. We find that this gives us a revenue increase of 10.6% on average for Levitra and 14.1% on average for Cialis.

The results of these three counterfactual simulations are summarized in Table 5. To sum up, it appears that firms could significantly increase their revenues during the launch period by simply reallocating their existing expenditure on detailing.

While these counterfactual simulations give us a sense of the revenue gains that firms could make by taking heterogeneity into account, the extent to which these gains arise because of heterogeneity in learning, in risk aversion and in the persuasive effect is not yet obvious. In order to assess the contribution of heterogeneity in these three parameters on revenue gains, we sequentially repeat simulation 3 taking into account heterogeneity for each parameter at a time while fixing the other two at the population means. We then compute the revenue gains in the three simulations described earlier. The results of this exercise (Table 6) suggest that 50-56% of the revenue gains in simulation 3 (where detailing is reallocated both temporally and cross-sectionally) come through accounting for the heterogeneity in learning. About 9-10% of the gains come from accounting for heterogeneity in the risk aversion parameter while the remaining 35-41% comes from accounting for heterogeneity in the persuasive effect.

8. Conclusion

We started off with a problem of optimal allocation of marketing communication for new products, with the specific focus on allocation of these resources over time and across consumers. Heterogeneous learning was recognized as a factor that would affect the temporal as well as cross-sectional allocation of these resources. We specified a structural model of heterogeneous learning and developed a methodology to estimate such a model at the individual physician level.

We estimated this model using a unique panel dataset consisting of physician prescriptions and detailing calls. We allowed for detailing to have both an informative and a persuasive effect and estimated both these effects at the individual physician level. We then conducted a set of counterfactual simulations to find the implications of heterogeneous learning on optimal allocations of detailing over time and across agents (physicians in our case).

Our parameter estimates indicate that there is considerable heterogeneity across physicians in terms of their learning rates. Some physicians require only a few detailing calls to substantially reduce their uncertainty about a new drug, others require a large number of repeated detailing calls in order to reduce their uncertainty to the same extent. Physicians also differ significantly in their persuasive effect of detailing, which is the only effect of detailing after they have learnt about the drug. Because of both these effects, there is a significant amount of variation across physicians in terms of how their responsiveness to detailing varies over time.

We also find that volume based deciles explain the variation in the persuasive effects of detailing to some extent. Specifically, physicians who are heavy prescribers in the category are more likely to have a low persuasive effect. Thus, their responsiveness to detailing reduces rapidly to a low level. The responsiveness of light prescribers reduces more slowly and settles down at a relatively high responsiveness after they have learnt about the drug.

We conducted three counterfactual simulations to see if firms could increase their revenues in this category by changing their detailing allocation patterns. We find that if they change just their temporal allocation without changing the cross-sectional allocation of detailing, they get a 5.9 to 8.3% increase in revenues in the first three months after launch. This reflects gains in revenues by front-loading their detailing to early periods after the launch of the drug. If they change their cross-sectional allocation without altering the temporal allocation, firms obtain a more modest 4.7% to 5.8% increase in revenues. If they change both their temporal and cross-sectional allocations, they get a substantial 10.6% to 14.1% increase in revenues. Furthermore, accounting for heterogeneity in learning accounts for 50-56% of these gains, while accounting for heterogeneity in risk aversion accounts for about 9-10% and the remaining 35-41% comes from accounting for heterogeneity in the persuasive effect.

We finally list some of the limitations of this study. In specifying this model of learning about new drugs, we have assumed away other potentially important sources of learning, for instance, learning from other physicians. This assumption of no learning through other sources is due to the absence of appropriate data and might overstate the degree of learning through

detailing that we infer. We do not directly include the effect of free samples, which could be another source of learning, although our learning through patient feedback would include the effect of samples since we include occasions when the physician gives only a free sample to the patient (and does not write a prescription) when we count the number of feedback signals.

In ignoring the effect of other marketing instruments, we might be overstating the effect of detailing if these other instruments are positively correlated with detailing. However, it must be kept in mind that detailing is by far the primary form of marketing in prescription drug categories. This is especially true for the category we study during the period of the data. Furthermore, it is the primary marketing instrument that can be allocated differentially across physicians (unlike other prominent marketing instruments like direct to consumer advertising (DTC) and journal advertising).

As mentioned earlier, another potential concern could be about endogeneity in detailing allocation. While there are reasons to believe that this may not be a very big concern for the specific category we study, this could be addressed by including a detailing supply equation in the model and jointly estimating parameters of demand and supply. A complication that would arise in this case is that the detailing supply equation cannot be static due to the presence of learning. Learning causes detailing to have persistent effects over time and hence firms are likely to take into account the effect of their detailing on future prescription behavior of physicians. This would complicate the problem substantially. We also assume away any forward-looking behavior of physicians. Physicians, if they are aware that they learn through detailing, may have incentives to be more willing to see detailers early on in order to learn about the drug quickly. Finally, firms may also be interested in determining the optimal level and the allocation of detailing (our counterfactual analysis focuses only on allocation keeping the amount of detailing fixed). These are challenging problems and the methodology to account for these phenomena are not fully developed yet. Future research could potentially address these questions.

References

Ackerberg, Daniel (2001), "Empirically Distinguishing Informative and Persuasive Effects of Advertising," *RAND Journal of Economics*, 32, pp. 316-333.

Ackerberg, Daniel (2003), "Advertising, Learning and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review*, 44, pp. 1007-1040.

Anand, Bharat and Roni Shachar (2004), "Advertising, the Matchmaker," Working Paper, Harvard University.

Becker, Gary S. and Kevin M. Murphy (1993), "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics*, 108, pp. 942-964.

Byzalov, Dmitri and Roni Shachar (2004), "The Risk Reduction Role of Advertising," *Quantitative Marketing and Economics*, 2, pp. 283-320.

Chib, Siddhartha and Edward Greenberg (1995), "Understanding the Metropolis Hastings Algorithm," *American Statistician*, 49, 327-335.

Coscelli, Andrea and Matthew Shum (2004), "An Empirical Model of Learning and Spillovers in New Drug Entry," *Journal of Econometrics*, 122, pp. 213-246.

Crawford, Gregory S. and Matthew Shum (2005), "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73, pp. 1137-1174.

Erdem, Tulin and Micheal P. Keane (1996), "Decision Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15, pp. 1-20.

CDER (2000), "Center for Drug Evaluation and Research Report to the Nation: 2000," available on the internet at http://www.fda.gov/cder/reports/RTN2000/RTN2000-3.HTM (accessed on August 21, 2004).

Fudenberg, Drew and Jean Tirole (2000), "Customer Poaching and Brand Switching," *Rand Journal of Economics*, 31, 634-657.

Gelfand, Alan and Adrian Smith (1990), "Sampling-based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, pp. 398-409.

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

Iyer, Ganesh, David A. Soberman and J. Miguel Villas-Boas (2005), "The Targeting of Advertising," *Marketing Science*, 24, pp. 461-476.

Heilman, Carrie, Douglas Bowman and Gordon P. Wright (2000), "The Evolution of Brand Preferences and Choice Behavior of Consumers New to a Market," *Journal of Marketing Research*, 37, pp. 139-155.

Hurwitz, Mark and Richard Caves (1988), "Persuasion or Information? Promotion and the Shares of Brand Name and Generic Pharmaceuticals," *Journal of Law and Economics*, 31, pp. 299-320.

Kamakura, Wagner A., Bruce S. Kossar and Michel Wedel (2004), "Identifying innovators for the cross-selling of new products," *Management Science*, 50, pp.1120-1133.

Lilien, Gary L., (1974a), "Application of a Modified Linear Learning Model of Buyer Behavior," *Journal of Marketing Research*, 11, pp. 279-285.

Lilien, Gary L., (1974b), "A Modified Linear Learning Model of Buyer Behavior," *Management Science*, 20, pp. 1027-1036.

Lilien, Gary L., Ambar Rao and Shlomo Kalish (1981), "Bayesian Estimation and Control of Detailing Effort in a Repeat Purchase Environment," *Management Science*, 27, 493-507.

Leffler, Keith (1981), "Persuasion or Information? The Economics of Prescription Drug Advertising," *Journal of Law and Economics*, 24, pp. 45-74.

Manchanda, Puneet and Pradeep K. Chintagunta (2004), "Responsiveness of Physician Prescription Behavior to Sales force Effort: An Individual Level Analysis," *Marketing Letters*, 15, pp. 129-145.

Manchanda, Puneet, Peter E. Rossi and Pradeep K. Chintagunta (2004a), "Response Modeling with Non-Random Marketing Mix Variables," *Journal of Marketing Research*, 41, pp. 467-478.

Manchanda, Puneet, Ying Xie and Nara Youn (2004b), "The Role of Targeted Communication and Contagion in New Product Adoption," *Marketing Science*, forthcoming.

Manchanda, Puneet and Elisabeth Honka (2005), "The Effects and Role of Direct-to-Physician Marketing in the Pharmaceutical Industry: An Integrative Review," *Yale Journal of Health Policy, Law and Ethics*, 5, 785-822.

Meyer, Robert J. and Arvind Sathi (1985), "A Multiattribute Model of Consumer Choice during Product Learning," *Marketing Science*, 4, pp. 41-61.

Narayanan, Sridhar, Puneet Manchanda and Pradeep Chintagunta (2005), "Temporal Differences in the Role of Marketing Communication for New Products," *Journal of Marketing Research*, 42, pp. 278-291.

Neslin, Scott (2001), "ROI Analysis of Pharmaceutical Promotion," unpublished study conducted for the Association of Medical Publications available on the internet at http://www.rappstudy.org (accessed on August 21, 2004).

Parsons, Leonard J. and Piet Vanden Abeele (1981), "Analysis of Sales Call Effectiveness," *Journal of Marketing Research*, 18, pp. 107-113.

Rizzo, John A. (1999), "Advertising and Competition in the Ethical Pharmaceutical Industry: The Case of Antihypertensive Drugs," *Journal of Law and Economics*, 42, pp. 89-116.

Roberts, John H. and Glen L. Urban (1988), "Modeling Multiattribute Utility, Risk, and Belief Dynamics for New Consumer Durable Brand Choice," *Management Science*, 34, pp. 167-185.

Rossi, Peter, Robert McCulloch and Greg Allenby (1996), "The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15, 321-340.

Shaffer, Greg and John Zhang (1995), "Competitive Coupon Targeting," *Marketing Science*, 14, 95-116.

Stoneman, P. (1981), "Intra-Firm Diffusion, Bayesian Learning and Profitability," *Economic Journal*, 91, pp. 375-388.

Tanner, Martin and W. Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 85, pp. 528-550.

Urban, Robert and John Hauser (1993), Design and Marketing of New Products (2nd ed.), Prentice-Hall.

Villas-Boas, J. Miguel (1999), "Dynamic Price Competition with Customer Recognition," *Rand Journal of Economics*, 30, 604-631.

Wittink, Dick R. (2002), "Analysis of ROI for Pharmaceutical Promotions," unpublished study conducted for the Association of Medical Publications, available online at http://www.rappstudy.org (accessed on August 21, 2004).

Table 1: Individual Level Parameter Estimates

Parameter		*Mean*	*Std. Deviation*
Detailing signal variance	$\sigma^2_{D_i}$	1.0567	0.1519
Feedback signal variance	$\sigma^2_{F_i}$	1.0986	0.1358
Absolute risk aversion	r_i	0.0045	0.0347
Coefficient – detailing stock	β_{1i}	0.6113	0.1040
Coefficient – patient request	β_{2i}	1.5695	0.0611
True Mean Quality – Cialis	Q_{i1}	1.0099	0.0901
True Mean Quality – Levitra	Q_{i2}	0.9999	0.1009

Notes:

1. Since these parameters are at the individual level, for each individual physician, the parameter has a mean and a standard deviation. The reported parameters are the mean and standard deviation of the physician-specific parameter means.

Table 2: Pooled Parameter Estimates

Parameter		*Mean*	*Std. Deviation*
Utility Error Variance	σ^2_1	5.7057	0.9012
Utility Error Variance	σ^2_2	5.4010	0.5695

Notes:

1. Other pooled parameters include the elements of the matrix V_γ and the vector λ. These are not reported here for the sake of brevity.

Table 3: Across-physician vs. within-physician standard deviations of the individual-level parameters

Parameter		*Across-physician standard deviation*	*Within-physician standard deviation*
Detailing signal variance	$\sigma^2_{D_i}$	0.1519	0.1044
Feedback signal variance	$\sigma^2_{F_i}$	0.1358	0.1286
Absolute risk aversion	r_i	0.0347	0.1284
Coefficient – detailing stock	β_{1i}	0.1040	0.0313
Coefficient – patient request	β_{2i}	0.0611	0.0839
True Mean Quality – Cialis	Q_{i1}	0.0901	0.0529
True Mean Quality – Levitra	Q_{i2}	0.1009	0.0432

Notes:

1. Across-physician standard deviation : the mean parameter value for each physician is first computed and then the standard deviation of these mean values is reported in this table.
2. Within-physician standard deviation: the within-physician standard deviation of the parameter is computed for each physician and then the mean of these standard deviations is reported in this table.

Table 4: Demographic Heterogeneity Parameters (λ)

Parameter	*Detailing Signal Variance*	*Feedback Signal Variance*	*Risk Aversion*	*True Quality Cialis*	*True Quality Levitra*	*Detailing Stock Coefficient*	*Patient Request Coefficient*
Intercept	1.0492 (0.0268)	1.1062 (0.2399)	0.0532 (0.0061)	1.0220 (0.0160)	0.9944 (0.0179)	0.5887 (0.0183)	1.5630 (0.0108)
Specialty – GP	-0.0050 (0.0217)	-0.0172 (0.0194)	-0.0050 (0.0050)	0.0003 (0.0129)	0.0007 (0.0145)	-0.0174 (0.0148)	-0.0009 (0.0087)
Specialty – Urologist	-0.0393 (0.0306)	-0.0384 (0.0274)	-0.0018 (0.0070)	-0.0152 (0.0182)	-0.0030 (0.0205)	-0.0145 (0.0209)	-0.0180 (0.0123)
Decile 1	0.0120 (0.0252)	0.0129 (0.0226)	-0.0027 (0.0058)	-0.0091 (0.0150)	0.0044 (0.0169)	0.0524 (0.0172)	0.0110 (0.0101)
Decile 2	-0.0212 (0.0242)	-0.0203 (0.0217)	-0.0038 (0.0055)	-0.0098 (0.0145)	0.0062 (0.0162)	0.0562 (0.0165)	0.0032 (0.0098)
Decile 3	-0.0068 (0.0252)	-0.0075 (0.0226)	-0.0032 (0.0058)	-0.0337 (0.0150)	-0.0082 (0.0169)	0.0470 (0.0172)	0.0031 (0.0101)
Decile 4	0.0179 (0.0214)	0.0112 (0.0192)	-0.0072 (0.0049)	-0.0090 (0.0128)	0.0045 (0.0144)	0.0538 (0.0146)	0.0052 (0.0086)
Decile 5	0.0102 (0.0255)	0.0080 (0.0229)	-0.0116 (0.0058)	-0.0082 (0.0152)	0.0030 (0.0171)	0.0535 (0.0174)	0.0181 (0.0103)
Decile 6	0.0165 (0.0229)	0.0059 (0.0205)	-0.0009 (0.0052)	-0.0093 (0.0137)	0.0039 (0.0154)	0.0421 (0.0156)	0.0064 (0.0092)
Decile 7	0.0137 (0.0213)	0.0090 (0.0191)	-0.0054 (0.0049)	-0.0235 (0.0127)	0.0120 (0.0143)	0.0429 (0.0145)	0.0082 (0.0086)
Decile 8	0.0318 (0.0243)	0.0179 (0.0218)	-0.0135 (0.0056)	-0.0152 (0.0145)	-0.0006 (0.0163)	0.0315 (0.0166)	0.0091 (0.0098)
Decile 9	0.0536 (0.0210)	0.0474 (0.0188)	0.0026 (0.0048)	-0.0009 (0.0125)	0.0162 (0.0140)	0.0272 (0.0143)	0.0225 (0.0084)

Table 5: Counterfactual Simulations: Revenue gains through reallocation of detailing

		Temporal Allocation	
		No Change	*Change*
Cross – Sectional Allocation	*No Change*	Current Situation	Simulation 1 Cialis: 8.3% (3.20) Levitra: 5.9% (2.74)
	Change	Simulation 2 Cialis: 5.8% (1.36) Levitra: 4.7% (2.19)	Simulation 3 Cialis: 14.1% (2.78) Levitra: 10.6% (2.51)

* The reported estimates are the posterior mean revenue gains and the figures in parentheses are the posterior standard deviations of these revenue estimates, obtained by conducting these simulations for 30 draws from the joint posterior distribution of all parameters.

Table 6: Revenue gains from reallocation of detailing: relative contributions of heterogeneity in learning, risk aversion and persuasive effect

Drug	*% Contribution of revenue gain by heterogeneity in*		
	Learning	Risk Aversion	Persuasive Effect
Cialis	49.78% (11.36)	9.60% (7.74)	40.62% (8.49)
Levitra	55.93% (13.11)	9.27% (7.61)	34.80% (8.37)

* The reported numbers are the posterior mean relative contributions and the figures in parentheses are the posterior standard deviations of these relative contribution. They are obtained by taking 30 draws from the joint posterior distribution of all parameters, fixing the non-focal parameters (e.g. for the contribution from heterogeneity in learning, the non-focal parameters are risk aversion and the persuasive effect) at their posterior means and conducting the simulations for each draw. The simulation estimates are reported as a percentage of the revenue gains for the simulations without any constraints on the parameters.

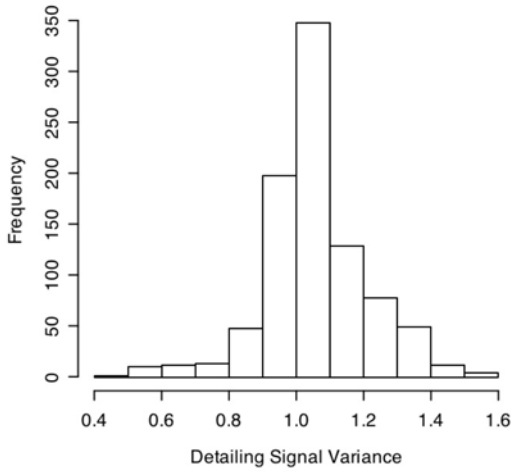
Table 7: Comparison with Null Models

Measure	Model						
	Full Model	**Null Model 1**	**Null Model 2**	**Null Model 3**	**Null Model 4**	**Null Model 5**	**Null Model 6**
Log marginal likelihood	-5186.7	-	-5807.8	-5792.1	-5517.2	-5370.0	-5382.1
Bayes Factor (wrt. Full Model)	-	-	447.1	547.6	79.9	120.6	132.7
Log marginal likelihood – out of sample	-398.7		-476.12	-421.8	-418.4	-408.1	-409.8
Hit rate – in sample	46%	29%	37%	39%	40%	40%	41%
Hit rate – out of sample	38%	-	35%	36%	36%	36%	37%
Hit rate – in sample (3 months)	52%	31%	31%	42%	42%	43%	44%

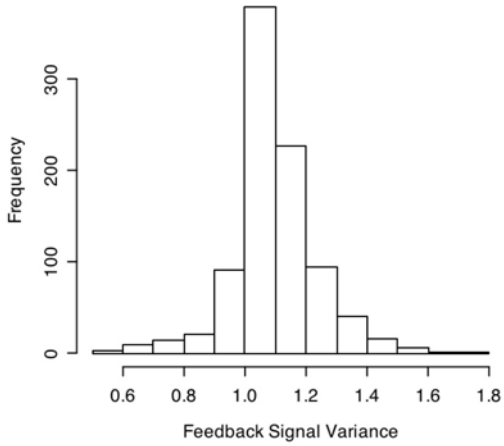
Null Model 1: Random Coefficient Logit model with state dependence.
Null Model 2: The full model with the persuasive effect of detailing, but the learning process removed.
Null Model 3: The full model with learning, but with no persuasive effect of detailing.
Null Model 4: The full model with homogenous learning, i.e. no heterogeneity on either the detailing signal variance or the feedback signal variance.
Null Model 5: The full model with homogenous detailing signal variance, but heterogeneous feedback signal variance.
Null Model 6: The full model with homogenous feedback signal variance, but heterogeneous detailing signal variance.
* A Bayes Factor greater than 1 favors the full model.
** Bayes Factor is relative to the re-computed full model with one observation per physician dropped.
*** The out of sample hit rate for the random coefficients model with state dependence is not computed as it involves the loss of two observations per physician, making it hard to compare with other models. Also, we have not reported the likelihood and Bayes factors for the same reason.

Figure 1: Histograms of means of individual level parameters

1(a): detailing signal variance $\sigma^2_{D_i}$



1(b): feedback signal variance $\sigma^2_{F_i}$



1(c): coefficient of absolute risk aversion r_i

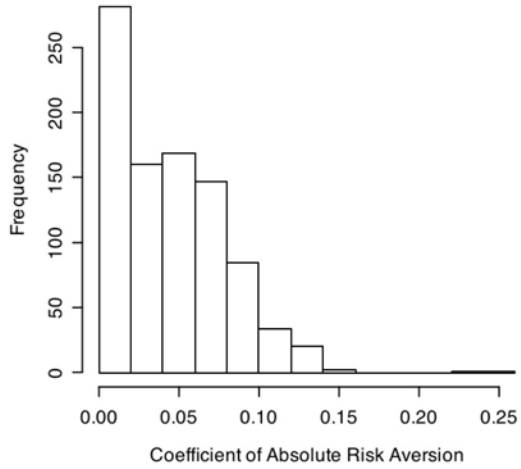
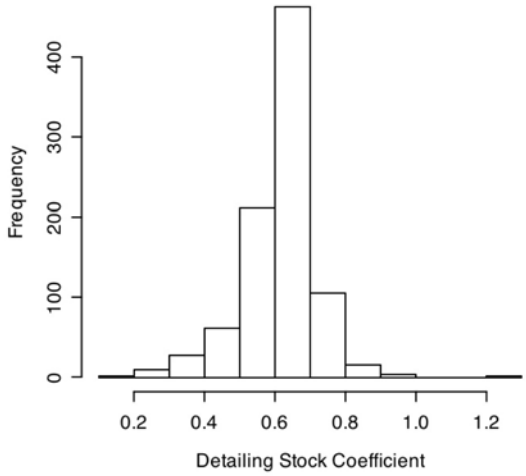
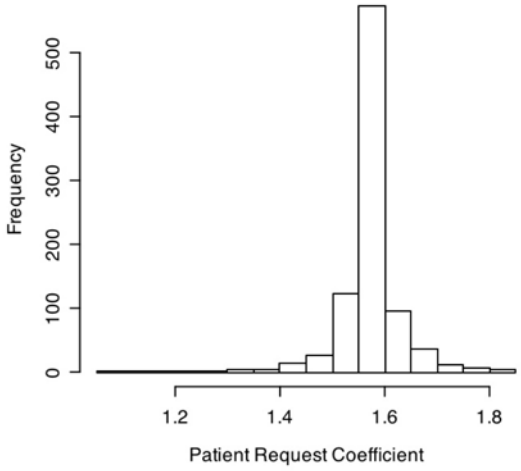


Figure 1: Histograms of Means of Individual Level Parameters

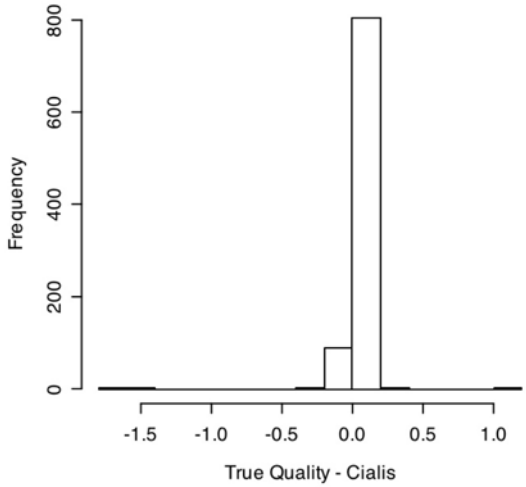
1(d): detailing stock coefficient β_{1i}



1(e): patient request coefficient β_{2i}



1(f): true mean quality for Cialis Q_{i1}



1(g): true mean quality for Levitra Q_{i2}

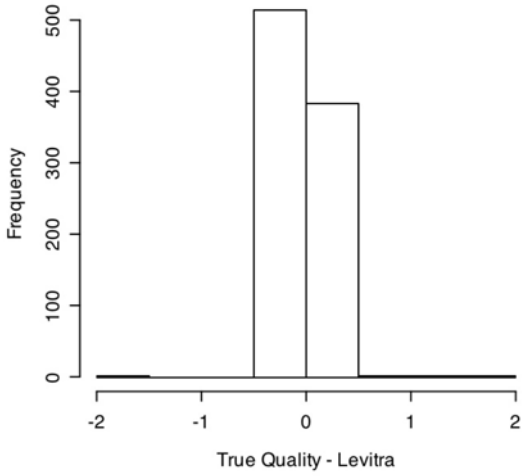


Figure 2: Histogram of the number of detailing calls required to reduce uncertainty (variance of quality belief) about a new drug to one-tenth of the initial value

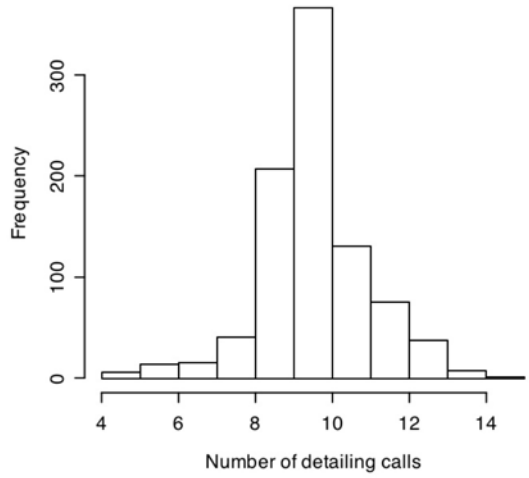


Figure 3: Plot of Informative Effect vs. Persuasive Effect

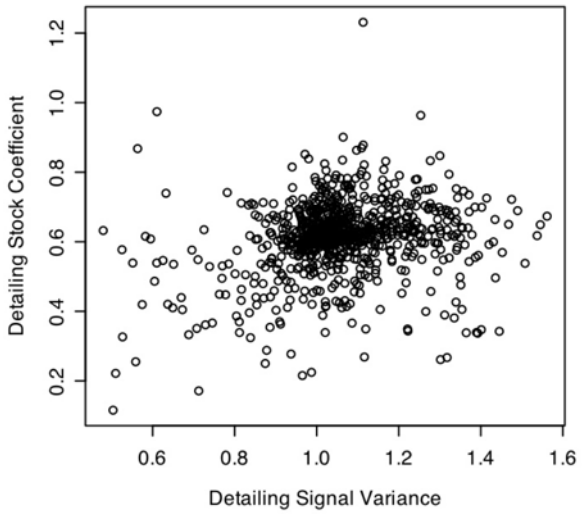
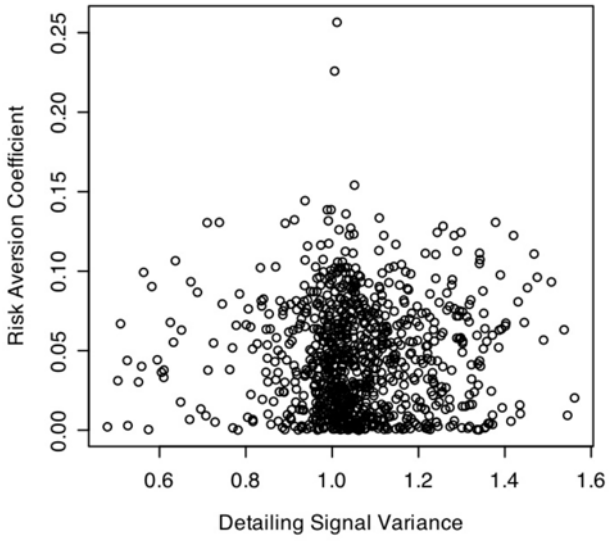


Figure 4: Plot of Informative Effect vs. Risk Aversion



Appendix: Full Conditional Distributions

Given the model described in the model section, the joint posterior distribution of all the parameters conditional on the data is given by is given by the following expression

$$L \propto \prod_{i=1}^{N} \left\{ \begin{array}{l} \prod_{t=1}^{T_i} \left[f\left(U_{it} \mid \{I_{ijt}\}, \{X_{ijt}\}, \{Q_{ijt}\}, r_i, \beta_i, \Sigma\right) \prod_{j=1}^{J} \left[f\left(Q_{ijt} \mid Q_{ij(t-1)}, nd_{ijt}, I_{ij(t-1)}, \sigma_{D_i}^2, \sigma_{F_i}^2, Q_{ij}, Q_{j0}\right)\right]\right] \\ \quad \cdot f\left(\gamma_i \mid \Lambda, Z_i, V_\gamma\right) \end{array} \right\}$$

$$\cdot f\left(\lambda \mid \bar{\lambda}, V_\lambda\right) f\left(V_\gamma \mid g, G\right) \prod_{j=2}^{J} \left[f\left(\sigma_j^2 \mid s_{1j}, s_{2j}\right)\right]$$

We shall now derive the full conditional distributions of all the parameters (and augmented parameters like U_{ijt} and Q_{ijt}), so that we can use a Gibbs sampling method. For those parameters for which the full conditional distributions are not from known distribution families, we shall use the Metropolis Hastings algorithm to draw from the respective full conditional distributions.

Suppose we define:

$$U_{it} = \begin{pmatrix} U_{i1t} \\ . \\ U_{iJt} \end{pmatrix}, \qquad \Omega_{it} = \begin{pmatrix} -\exp\left(-r_i Q_{i1t} + \frac{r_i^2 \sigma_{Q_{i1t}}^2}{2}\right) & & \\ & . & \\ & & -\exp\left(-r_i Q_{iJt} + \frac{r_i^2 \sigma_{Q_{iJt}}^2}{2}\right) \end{pmatrix}, \quad X_{it} = \begin{pmatrix} X_{i1t} \\ . \\ X_{iJt} \end{pmatrix}$$

The full expression for the joint posterior distribution can then be written as

$$L \propto \prod_{i=1}^{N}\left\{\prod_{t=1}^{T_i}\left[\left|\Sigma\right|^{-1/2}\exp\left(-\frac{1}{2}\left(U_{it}-\Omega_{it}-X_{it}\beta_i\right)'\Sigma^{-1}\left(U_{it}-\Omega_{it}-X_{it}\beta_i\right)\right)\prod_{j=1}^{J}\left[1\left(U_{ijt}>\max\left(U_{ikt}\right),k\neq j\right)\right]^{I_{ijt}}\right]\right.$$
$$\left.\prod_{j=1}^{J}\left[\frac{1}{\sigma^2_{Q_{ijt}}\sqrt{\frac{nf_{ijt}}{\sigma^2_{F_i}}+\frac{nd_{ijt}}{\sigma^2_{D_i}}}}\exp\left(-\frac{1}{2}\frac{\left[Q_{ijt}-\frac{\sigma^2_{Q_{ijt}}}{\sigma^2_{Q_{ij(t-1)}}}Q_{ij(t-1)}-\sigma^2_{Q_{ijt}}Q_{ij}\left(\frac{nf_{ijt}}{\sigma^2_{F_i}}+\frac{nd_{ijt}}{\sigma^2_{D_i}}\right)\right]^2}{\left(\frac{nf_{ijt}}{\sigma^2_{F_i}}+\frac{nd_{ijt}}{\sigma^2_{D_i}}\right)\sigma^4_{Q_{ijt}}}\right)\right]\right]$$
$$\left.\left|V_\gamma\right|^{-1/2}\exp\left(-\frac{1}{2}\left(\gamma_i-\Lambda Z_i\right)'V_\gamma^{-1}\left(\gamma_i-\Lambda Z_i\right)\right)\right\}$$
$$\left|V_\lambda\right|^{-1/2}\exp\left(-\frac{1}{2}\left(\lambda-\overline{\lambda}\right)'V_\lambda^{-1}\left(\lambda-\overline{\lambda}\right)\right)\qquad\cdot\frac{V_\gamma^{(g-k-1)/2}}{\left|G\right|^{g/2}}\exp\left(-\frac{1}{2}tr\left(G^{-1}V_\gamma\right)\right)\cdot\prod_{j=1}^{J-1}\left[\frac{\exp\left(-\frac{1}{s_{2j}\sigma_j^2}\right)}{\sigma_j^{2(s_{1j}+1)}}\right]$$

From this joint posterior, we can derive the full conditional distributions as follows:

1. $\sigma_j^2 \mid \{U_{ijt}\},\{X_{ijt}\},\{Q_{ijt}\},\{\beta_i\},\{r_i\} \sim IG\left(\frac{\sum_{i=1}^{N}T_i}{2}+s_{1j},\ \frac{2s_{2j}}{2+s_{2j}\sum_{i=1}^{I}\sum_{t=1}^{T_i}\left(U_{ijt}+\exp\left(-r_iQ_{ijt}+\frac{r_i^2\sigma^2_{Q_{ijt}}}{2}\right)-X_{ijt}\beta_i\right)^2}\right)$

2. $V_\gamma \mid \gamma_i,\Lambda,Z_i,g,G \sim Inverse\ \ Wishart\left(g+N,\left[\sum_{i=1}^{N}\left[\left(\gamma_i-\Lambda Z_i\right)\left(\gamma_i-\Lambda Z_i\right)'\right]+G^{-1}\right]^{-1}\right)$

3. $\lambda \mid \{\gamma_i\},V_\gamma,Z,V_\lambda,\overline{\lambda} \sim N\left(\left[V_\gamma^{-1}\otimes Z'Z+V_\lambda^{-1}\right]^{-1}\left[V_\gamma^{-1}\otimes Z'Z\hat{\lambda}+V_\lambda^{-1}\overline{\lambda}\right],\left[V_\gamma^{-1}\otimes Z'Z+V_\lambda^{-1}\right]^{-1}\right)$

where

$$\hat{\lambda}=vec\left[\left(Z'Z\right)^{-1}Z'\Gamma\right],\quad Z=\begin{pmatrix}Z_1'\\ .\\ Z_N'\end{pmatrix},\quad \Gamma=\begin{pmatrix}\gamma_1'\\ .\\ \gamma_N'\end{pmatrix}$$

4. Each physician is independent conditional on the *X* and *Z* matrices. And each observation for the physician is independent conditional on the vector of Q_{ijt}. Thus, we can draw the latent utilities for a particular physician and a particular observation separately from the other observations. This involves sequentially drawing from a truncated multivariate normal distribution for each time period.

$$\begin{bmatrix} U_{i1t} \\ . \\ U_{iJt} \end{bmatrix} \sim \textit{Truncated MVN} \left(\begin{pmatrix} -\exp\left(-r_i Q_{i1t} + \frac{r_i^2 \sigma^2_{Q_{i1t}}}{2} \right) + X_{i1T}\beta_i \\ . \\ -\exp\left(-r_i Q_{iJt} + \frac{r_i^2 \sigma^2_{Q_{iJt}}}{2} \right) + X_{iJT}\beta_i \end{pmatrix}, \Sigma \right)$$

with the truncation such that $U_{ijt} > U_{ikt},\ \forall k \neq j,\ I_{ijt} = 1$

5. The full conditional distributions for the individual level parameters, γ_i and the quality means Q_{i1t} are not from known families of distributions. Hence, draws from the distribution of these parameters for each individual physician are obtained using the Metropolis Hastings algorithm. We use a Random Walk Metropolis Hastings algorithm (Chib and Greenberg, 1995) with a normal candidate density to make these draws. The variances of these densities were obtained from the hessian of the pooled maximum likelihood estimates for these parameters, which was scaled up to obtain the best numerical efficiency.