Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 122 (2004) 213–246
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

An empirical model of learning and patient spillovers in new drug entry

Andrea Coscelli[a], Matthew Shum[b,*]

[a] *Lexecon, Ltd., Orion House, 5 Upper St. Martin's Lane, London WC2H9EA, UK*
[b] *Department of Economics, Johns Hopkins University, Mergenthler Hall 463, 3400N, Charles Street, Baltimore, MD 21218, USA*

Accepted 3 September 2003

Abstract

We specify and estimate a diffusion model for the new molecule *omeprazole* into the anti-ulcer drug market. Our model is based on a Bayesian learning process whereby doctors update their beliefs about *omeprazole*'s quality relative to existing drugs after observing its effects on the patients that have been prescribed this drug. The model also accommodates informational spillovers and heterogeneity in informativeness across patients with different diagnoses. We obtain estimates of the learning process parameters using a novel panel data set tracking doctors' complete prescription histories over a 3-year period.

JEL classification: I10; L10

Keywords: Entry of innovative drugs; Barriers to entry; Structural diffusion models

1. Introduction

First-mover advantage is a well-documented phenomenon in many differentiated product markets (see Urban et al. (1986) for a survey of the evidence). Economists have tended to attribute this phenomenon to lack of information among consumers about the quality or attributes of an entrant's product; for example, Shapiro (1982, p. 7) states that

> ...the fundamental source of the entry barrier is an information one: consumers have better information about established brands than about new ones [...] information is the basic barrier to be overcome by a new product...

* Corresponding author. Tel.: +1-410-516-8828; fax: +1-410-516-7600.
E-mail addresses: andrea.coscelli@lexecon.co.uk (A. Coscelli), mshum@jhu.edu (M. Shum).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2003.09.002

The doctor/patient relationship is fraught with uncertainty. Doctors have incomplete information on the medical condition of a patient, and which treatment is best for the patient. Doctors learn about the quality of alternative treatments both through direct experience (actual prescriptions of the new drug), and indirect experience (such as promotional activity by pharmaceutical companies, articles in medical journals and attendance at medical conferences). This paper focuses on direct information, which accumulates slowly, and is confounded by heterogeneity across diagnoses: what works for diagnosis X may not work as well for diagnosis Y.

Using a novel panel data set of complete prescription histories for a sample of doctors in the Rome (Italy) metropolitan area, we study the diffusion process of a new anti-ulcer drug (*omeprazole*) during a 3-year period (1990–1992). The evolution of *omeprazole*'s market share over time was marked by the gradual diffusion which characterizes new product entry into many product markets: *omeprazole*'s market share (as a proportion of total prescriptions) climbed from under 5% in the latter half of 1990 to about 15% in early 1992, and eventually up to 25% by the middle of 1995.

In this paper, we gauge how well this gradual diffusion pattern can be explained by a learning model in which doctors, initially uncertain about the quality differential between *omeprazole* and the incumbent drugs, update their beliefs about this differential after observing noisy signals from patients to whom they have prescribed *omeprazole*. To that end, we specify and estimate the parameters of such a learning model. Furthermore, in order to accommodate features specific to the pharmaceutical prescription process, we extend the basic learning model to allow for spillovers across all the patients of a given doctor, as well as heterogeneity in informativeness across patients. While there are alternative explanations for the individual-level diffusion process (such as the publication of the results of post-marketing clinical trials in medical journals), we focus on a learning explanation because our data includes especially rich detail on doctors' prescription histories.

Our results suggest that the learning model does very well in generating the observed slow diffusion path of *omeprazole* in the Italian market. The parameters of the learning model quantify, in informational terms, the disadvantage that *omeprazole* suffered relative to the existing drugs upon its entry into the Italian anti-ulcer market. This informational disadvantage can arise from either doctors' initial pessimism about *omeprazole*'s quality, or risk aversion. In addition, we find that the informational spillovers are *negative* across some diagnosis groups, which tends to *retard* the speed of learning. That is, we find that a positive outcome when prescribing *omeprazole* for certain diagnoses leads doctors to regard it as less attractive for other diagnoses.

The next section provides some background on the international and Italian anti-ulcer drug markets. Section 3 describes the doctor-level learning model, Section 4 describes our panel data set of complete prescription histories and Section 5 derives the estimating equations associated with the learning model. Results from several specifications of the learning model are presented and interpreted in Section 6, and we conclude in the last section.

2. Background

Several studies have documented the existence and, more importantly, the nature of barriers to entry into pharmaceutical markets. Bond and Lean (1977) found evidence of substantial pioneer advantage, but they also found that products containing some therapeutic novelty managed to gain large market shares when backed by heavy promotional campaigns. Berndt et al. (1997) document similar effects in the anti-ulcer drug market. Their findings clearly show that technological advances do not necessarily translate into large market shares without tremendous marketing muscle.[1] As striking as the results from the two studies are, however, they never explain the causes of pioneer advantage. The availability of doctor-level prescription histories allows us a unique opportunity to assess the role of information in explaining the diffusion patterns observed in many product markets.[2]

This paper joins a growing empirical literature examining behavioral explanations for diffusion patterns for new products in experience good markets. Among these studies, Ackerberg (2002) and Erdem and Keane (1996) estimated structural learning models to explain consumers' purchase patterns for, respectively, yogurt and laundry detergent. Ching (2000) has also estimated a demand model for pharmaceuticals based on a Bayesian learning procedure. Our work differs from these papers because we consider a more general learning model which allows for spillovers across all the patients of a given doctor, as well as heterogeneity in informativeness across patients. These extensions seem especially appropriate for pharmaceutical markets, since prescription drugs (and in particular anti-ulcer drugs) are usually prescribed for several different diagnoses.

Using aggregate market share data, Azoulay et al. (2003) estimate a diffusion model to study the importance of consumption externalities in explaining the diffusion patterns of H_2-antagonist drugs into the anti-ulcer drug market. Our analysis extends their work by using a novel micro-data set to quantify the extent of network-type spillovers across patients belonging to the same doctor.[3]

3. The learning model

In this section, we describe the behavioral model which forms the basis of our empirical analysis. In what follows, we index doctors by the subscript i, and assume that patients are heterogeneous in their diagnoses, which we subscript by j. We begin

[1] Using a similar data set, Azoulay (2002) investigates how promotional activity and scientific information arising from clinical trials affect the diffusion of competing molecules in the anti-ulcer drug market. King (2000) focuses on the role of marketing in increasing the perceived product differentiation (i.e., degree of substitutability) between competing anti-ulcer drugs.

[2] A related literature (Stern, 1996; Ellison et al., 1997) has investigated the extent of competition in pharmaceutical markets by estimating cross-price elasticities between the competing drugs in a market. Unlike these papers, we abstract away from competition between existing anti-ulcer drugs.

[3] Finally, there has been a long interest in diffusion models in the marketing literature. See Bass et al. (1990) for a review of this largely theoretical and macro-level empirical literature. Chandrashekaran and Sinha (1995) is one of the few papers in this literature which are formulated at the micro-level.

by describing a baseline version of the learning model in which doctors are assumed to be risk neutral. At the end of this section, we discuss an alternative model which allows for risk aversion.

Consider a given patient k, from diagnosis group j, who visits doctor i during period t. We assume that doctor i distinguishes between two treatment alternatives: the new molecule, *omeprazole* (alternative 1), and *any* of the other molecules (alternative 0). The utilities for a given patient k with diagnosis j during period t from each alternative are:

$$U^i_{1jkt} = \alpha^*_1 x^i + \beta p_{1t} + \xi^*_1(t) + \delta^*_{1j} + \varepsilon^i_{1jkt} \quad \text{if take } omeprazole, \tag{3.1}$$

$$U^i_{0jkt} = \alpha^*_0 x^i + \beta p_{0t} + \xi^*_0(t) + \delta^*_{0j} + \varepsilon^i_{0jkt} \quad \text{otherwise}, \tag{3.2}$$

where

- p_{1t} and p_{0t} are, respectively, the price of *omeprazole* and a weighted average of the prices of the incumbent drugs weighted by their market shares at time t. The vector x^i contains observed doctors' characteristics.
- δ^*_{1j} and δ^*_{0j} parameterize the "unobserved quality" of *omeprazole* and the incumbent drugs when treating diagnosis j. These are unobserved by the econometrician. Doctors, however, are presumed to know δ^*_{0j}, and have imperfect information about δ^*_{1j}. As described below, doctors learn about δ^*_{1j} by prescribing *omeprazole* to their patients.
- $\xi^*_1(t)$ and $\xi^*_0(t)$ are flexible functions of time, which parameterize period t factors which affect the attractiveness of, respectively, *omeprazole* and the incumbent drugs. These are the same over all doctors, patients, and diagnoses. In particular, the function $\xi^*_1(t)$ proxies for aspects of the learning process which we do not explicitly model, such as word of mouth, medical congresses, and articles in medical journals.
- ε^i_{1jkt} and ε^i_{0jkt} are i.i.d. (over doctors, patients, diagnoses, and time periods) shocks associated with, respectively, *omeprazole* and the incumbent drugs. They are observed by the doctors, but not by the econometrician.

Throughout, we abstract away from agency problems between the doctor and the patient, and assume the doctor maximizes the *patient's* utility from the prescription.[4] Doctor i chooses the option with the higher *per-period* utility.[5] The choice rule for

[4] The reputation effects resulting from the long-term nature of many patient–doctor relationships in Italy (the National Health Service requires each enrollee to list a general practitioner) tend to minimize the divergence between doctors' and patients' objective functions which potentially form the basis of agency problems.

[5] For computational tractability we have assumed that doctors are myopic in our model, so that in any given time period, a doctor chooses the molecule with the highest *per-period* utility based solely on her current information. If the doctor were forward-looking, she would choose the molecule with the highest *present discounted utility* and thereby take into account the information that she would gain about *omeprazole* by prescribing it this period. Ongoing work by Crawford and Shum (2000) examines issues of uncertainty and matching in pharmaceutical demand in a fully forward-looking framework. Ferreyra (1999) has recently estimated a forward-looking dynamic learning model, using the same data that we use in this paper, but without allowing for spillovers across patients.

the doctor is to prescribe *omeprazole* if $E_t(U^i_{1kjt}) > U^i_{0kjt}$. If we assume that ε^i_{1jkt} and ε^i_{0jkt} are i.i.d. with the type 1 extreme value distribution, the probability that doctor i prescribes *omeprazole* takes the familiar logit form: [6]

$$\text{Prob(prescribe } \textit{omeprazole}) = \frac{\exp(\alpha x^i + \beta \Delta p_t + \xi(t) + E_t\delta_j)}{1 + \exp(\alpha x^i + \beta \Delta p_t + \xi(t) + E_t\delta_j)} \tag{3.3}$$

where we have substituted $\alpha \equiv \alpha^*_1 - \alpha^*_0$, $\xi(t) \equiv \xi^*_1(t) - \xi^*_0(t)$, and $E_t\delta_j \equiv E_t\delta^*_{1j} - \delta_{0j}$. α, β, and the $\xi(t)$ function are to be estimated. [7]

By distinguishing between different diagnoses, we allow the entrant and incumbent anti-ulcer drugs to differ in their effectiveness and suitability across diagnoses. This accommodates "segmentation" or "horizontal differentiation" in the market on the basis of diagnosis, which we believe to be an important feature of the anti-ulcer drug market.

3.1. Bayesian updating

The main focus of the paper is to measure how well the diffusion pattern for *omeprazole* can be explained by doctors' learning about δ. We explain this learning process in this section. Throughout, we assume that the learning processes are independent across doctors. [8] Therefore, we describe the learning process for doctor i, omitting the superscript i in most of the equations below for expositional clarity. We assume that, at time $t = 0$ (i.e., at *omeprazole*'s entry), she (doctor i) has the following *initial* beliefs about $\vec{\delta}$, the J-dimensional vector of quality differentials between *omeprazole* and the incumbent drugs:

$$\vec{\delta} \sim N\left(\vec{\delta}_1 \equiv \begin{bmatrix} E_1\delta_1 \\ \vdots \\ E_1\delta_J \end{bmatrix}, \quad \Sigma_{\delta,1} \equiv \begin{bmatrix} \sigma^2_{\delta,1} & 0 & \dots & 0 \\ 0 & \sigma^2_{\delta,2} & \dots & 0 \\ 0 & \cdots & \cdots & \vdots \\ 0 & 0 & \dots & \sigma^2_{\delta,J} \end{bmatrix}\right). \tag{3.4}$$

Throughout, we adopt the indexing convention that the subscript t denotes the beginning of period t; therefore, $\vec{\delta}_1$ denotes the mean of doctors' beliefs at the beginning of period 1, corresponding to the mean of the doctors' initial beliefs (and $\Sigma_{\delta,1}$ is similarly the initial variance–covariance matrix). The assumption that the initial variance–covariance matrix $\Sigma_{\delta,1}$ is diagonal implies that the information that doctors had about

[6] By aggregating all the non-*omeprazole*-based drugs into one alternative, we are implicitly assuming that all these drugs are perfectly substitutable, and that an *omeprazole*-based drug substitutes equally well with all of them. We make this assumption because we want to focus on the diffusion of drugs based on *omeprazole* into the marketplace.

[7] In most of the specifications reported below, we assume that the time function $\xi(t)$ is a quadratic time trend. As we point out below, since the price differential Δp_t only varies over time, it would be impossible to separately identify the price coefficient β apart from a full set of time dummies.

[8] Informational spillovers across doctors ("word of mouth") at the aggregate level are captured by the $\xi(t)$'s.

omeprazole before its entry is specific to particular diagnoses. This reflects the institutional feature that clinical trials—the results of which constitute most of doctors' prior information—are generally most informative as to a drug's effectiveness for particular diagnoses, and less informative regarding interactions of effects for different diagnoses, which would lead to non-zero off-diagonal terms in the initial variance–covariance matrix. [9]

The evolution of doctor i's beliefs over time can be derived period-by-period. Assume that doctor i begins time period t with beliefs that

$$\vec{\delta} \sim \mathrm{N}(\mathrm{E}_t\vec{\delta}, \Sigma_{\delta,t} \equiv \mathrm{E}_t\vec{\delta}\vec{\delta}' - (\mathrm{E}_t\vec{\delta})(\mathrm{E}_t\vec{\delta})') \tag{3.5}$$

(it will be clear later how these beliefs arise). During period t, the doctor prescribes *omeprazole* to k_j of her patients with diagnosis j, and observes k_j noisy signals of δ_j. We assume that these k_j signals ($\mu_{jtk},\ k = 1 \to k_j$) take the following form:

$$\mu_{jtk} = \delta_j + v_{jtk}, \tag{3.6}$$

where v_{jtk} is normally distributed, with zero mean. Doctors attempt to form estimates of δ_j from observations of the noisy signals μ_{jtk}'s.

Correlation structure: In order to accommodate informational spillovers across patients in different diagnosis groups (i.e., to capture the idea that "what is good for diagnosis X may not be good for diagnosis Y"), we assume that, within a given period t, the noise terms v are correlated across signals. We induce correlation across signals with the following variance components structure for each v:

$$v_{jtk} = \rho_j\theta_t + \eta_{jtk}, \quad j = 1,\ldots,J, \tag{3.7}$$

where (i) θ_t is distributed $\mathrm{N}(0,\sigma_\theta^2)$, i.i.d. over t; (ii) the η_{jtk}'s are independent over j, t, and k, and distributed $\mathrm{N}(0,\sigma_{\eta j}^2)$, $j = 1,\ldots,J$; and (iii) $\rho_1,\ldots,\rho_J$ are time-invariant parameters. Given these assumptions, then, the following correlation structure among all the signals (μ's) that doctor i observes in period t emerges:

1. $\mathrm{Var}(v_{jtk}) = \rho_j^2\sigma_\theta^2 + \sigma_{\eta j}^2$,
2. $\mathrm{Cov}(v_{jtk},\ v_{jtk'}) = \rho_j^2\sigma_\theta^2$, for $k \neq k'$,
3. $\mathrm{Cov}(v_{jtk},\ v_{j'tk'}) = \rho_j\rho_{j'}\sigma_\theta^2$, for $j \neq j'$ and $\forall\ k,\ k'$.

This one-factor variance components specification reduces the number of parameters, while placing mild restrictions on the correlation structure. [10] In Appendix A we calculate the variance–covariance matrix of a vector of signals, for a simple example.

Period-by-period updating: Given the normality assumptions on the signals μ as well as on the δ's, a doctor's posterior beliefs about $\vec{\delta}$ given $\vec{\mu}_t$ are described by a normal distribution with a mean and variance that can be derived using the multivariate normal conditional mean and variance formulas (Amemiya, 1985, p. 3). The computed posterior distribution in period t serves as the prior distribution for period $t+1$. In this way, we derive the sequence of a doctor's posterior distributions over all the periods

[9] Specific clinical evidence on *omeprazole*'s effectiveness for different diagnoses is presented further below.

[10] We have attempted to estimate an extended model with a 2-factor variance components structure, but we have experienced problems identifying some of the parameters in that case.

by repeatedly applying the conditional mean and variance formulas for jointly normally distributed random variables.

To this end, we characterize the joint distribution of $(\vec{\delta}, \vec{\mu}_t)$, during period t:

$$\begin{pmatrix} \vec{\delta} \\ \vec{\mu}_t \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mathrm{E}_t\vec{\delta} \\ \vec{\delta}_{\mu t} \end{bmatrix}, \begin{bmatrix} \Sigma_{\delta,t} & \Sigma_{\delta,\mu,t} \\ \Sigma'_{\delta,\mu,t} & \Sigma_{\mu,t} \end{bmatrix} \right), \tag{3.8}$$

where $\vec{\delta}_t$ and $\Sigma_{\delta,t}$ are, respectively, the mean and variance–covariance matrix of $\vec{\delta}$ conditional on all the signals received before period t. $\vec{\delta}_{\mu,t}$ and $\Sigma_{\mu,t}$ are the mean and variance–covariance matrix of the vector of signals $\vec{\mu}_t$ (Eqs. (A.2) and (A.3) in the appendix are examples of these formulas), and $\Sigma_{\delta,\mu,t}$ is the matrix of covariance terms between $\vec{\delta}$ and $\vec{\mu}_t$ (which is easy to derive given Eqs. (A.1), (3.6) and (3.7)).

Recall our indexing convention, whereby $\vec{\delta}_{t+1} \equiv \mathrm{E}(\vec{\delta}|\vec{\mu}_t)$ and $\Sigma_{\delta,t+1} \equiv \Sigma_t(\vec{\delta}|\vec{\mu}_t)$ are, respectively, the prior mean vector and variance–covariance matrix of the quality vector $\vec{\delta}$ at the beginning of period $t+1$ (i.e., conditional on all the information signals obtained up to, and including, period t). For the learning model described above, and given the initial beliefs (3.4), these quantities can be recursively defined as

$$\begin{aligned} \vec{\delta}_{t+1} &= \vec{\delta}_t + \Sigma'_{\delta,\mu,t}\Sigma^{-1}_{\mu,t}(\vec{\mu}_t - (\vec{\delta}_t)), \\ \Sigma_{\delta,t+1} &= \Sigma_{\delta,t} - \Sigma'_{\delta,\mu,t}\Sigma^{-1}_{\mu,t}\Sigma_{\delta,\mu,t} \end{aligned} \tag{3.9}$$

for period $t = 0, 1, 2, \ldots$

Eq. (3.9) yields the means of the posterior distribution of the δ's which are substituted into the logit prescription probabilities (cf. Eq. (3.3)). These probabilities form the basis for our likelihood function, which is described in the next section.

The parameters of the model which we estimate are: (i) the elements of the period zero initial mean vector $(\mathrm{E}_1\delta_1, \ldots, \mathrm{E}_1\delta_J)$; (ii) the diagonal elements of the initial variance–covariance matrix $(\sigma^2_{\delta,1}, \ldots, \sigma^2_{\delta,J})$; (iii) the true values $\delta_1, \ldots, \delta_J$; (iv) the parameters of the correlation structure $\rho_1, \ldots, \rho_J$, $\sigma^2_{\eta 1}, \ldots, \sigma^2_{\eta J}$, and σ^2_θ; and (v) the parameters which enter the utility specification α, β, and the time function $\xi(t)$.

3.2. Remarks

Rational expectations and risk aversion: In the preceding model, we have not allowed for risk aversion in the utility function. We can accommodate risk aversion directly in the utility specification of Eq. (3.1) above by including the posterior variance directly as an argument in the expected utility expression. Hence, the probability that doctor i prescribes *omeprazole* takes the form

$$\frac{\exp(\alpha x^i + \beta \Delta p_t + \xi(t) + \mathrm{E}_t\delta_j + \gamma\ \mathrm{Var}_t\,\delta_j)}{1 + \exp(\alpha x^i + \beta \Delta p_t + \xi(t) + \mathrm{E}_t\delta_j + \gamma\ \mathrm{Var}_t\,\delta_j)}, \tag{3.10}$$

where $\mathrm{Var}_t\ \delta_j$ denotes doctor i's posterior variance on δ_j based on the information she has obtained from prescriptions prior to period t, and γ measures the degree of risk

aversion.[11] Using the notation presented above in Eq. (3.8), we can write $\text{Var}_t\ \delta_j = \Sigma_{\delta,t}(j,j)$, the (j,j)th element of the period t variance–covariance matrix $\Sigma_{\delta,t}$.

Without additional assumptions, we cannot separately identify the prior mean $\text{E}_1\delta_j$, and the risk coefficient γ. To see this, consider the simplest case of only one diagnosis. In this case, the sequence of prior means and variances is given by the well-known formulas (cf. DeGroot, *Optimal Statistical Decisions*, p. 167), for periods $t = 2, 3, \ldots$

$$\text{E}_t\delta = \frac{\sigma_\mu^2\delta_1 + \sigma_1^2\left(\sum_{t'=1}^{r_{t-1}} \mu_{t'}\right)}{\sigma_\mu^2 + \sigma_1^2 * r_{t-1}},$$

$$\text{Var}_t\ \delta = \frac{\sigma_\mu^2\sigma_1^2}{\sigma_\mu^2 + \sigma_1^2 * r_{t-1}},$$

where r_{t-1} denotes the number of prescription of *omeprazole* up to (and including) period $t-1$, σ_μ^2 denotes the variance of the prescription signals, and δ_1 and σ_1^2 denote the initial mean and variance. By substituting these expressions into the expression for the choice probability (in Eq. (3.10) above), we see that the mean and variance above always enter the choice probability as the sum

$$\begin{aligned}
&\text{E}_t\delta_j + \gamma\ \text{Var}_t\ \delta_j \\
&= \frac{\sigma_\mu^2\delta_1 + \sigma_1^2\left(\sum_{t'=1}^{r_{t-1}} \mu_{t'}\right) + \gamma * \sigma_\mu^2\sigma_1^2}{\sigma_\mu^2 + \sigma_1^2 * r_{t-1}} \\
&= \frac{\sigma_\mu^2 * \left[\delta_1 + \gamma\sigma_1^2\right] + \sigma_1^2\left(\sum_{t'=1}^{r_{t-1}}(\delta + v_{t'} * \sigma_\mu)\right)}{\sigma_\mu^2 + \sigma_1^2 * r_{t-1}} \\
&= \frac{\sigma_\mu^2 * \left[\delta_1 + \gamma\sigma_1^2\right] + \sigma_1^2 * r_{t-1} * \delta + \sigma_1^2 * \sigma_\mu * \left(\sum_{t'=1}^{r_{t-1}} v_{t'}\right)}{\sigma_\mu^2 + \sigma_1^2 * r_{t-1}},
\end{aligned} \tag{3.11}$$

where we have re-written the signals as $\mu_{t'} = \delta + v_{t'} * \sigma_\mu$ with $v_{t'}$ as i.i.d. standard-normal random variables.

In our learning model, the parameters σ_μ^2 (the signal variance), σ_1^2 (the prior variance), δ (the true quality), δ_1 (the prior mean), and γ (the risk aversion parameter) affect the likelihood function only through expression (3.11) above. Clearly, if one only has cross-sectional data for the initial period $t = 1$, it is impossible to identify all these parameters separately (in this case, $r_0 = 0$ across all doctors, and the above expression reduces to the constant $\delta_1 + \gamma\sigma_1^2$).

However, inspection of the above expression yields that variation in r_t (the number of *omeprazole* prescriptions) across periods t and across doctors should be sufficient to identify σ_μ^2, σ_1^2, δ, and the sum $[\delta_1 + \gamma\sigma_1^2]$, just due to the non-linear updating formulas of the Gaussian learning model. Since the two remaining parameters δ_1 and γ only enter the above expression via the sum $[\delta_1 + \gamma\sigma_1^2]$, they cannot be separately identified (i.e., for any value of Z, the locus of pairs $(\delta_1, \gamma = (Z - \delta_1)/\sigma_1^2)$ yields the same likelihood function value).

[11] With CARA utility, $\gamma = \frac{1}{2}r$, where r is the coefficient of absolute risk aversion.

This discussion highlights the infeasibility of identifying the risk aversion parameter γ separately from the prior means $E_1\delta_j,\ j=1,2,3,4$. Therefore, in our estimation, we consider an additional restriction which rules out pessimism by setting the prior means equal to the true qualities:

$$E_1\delta_j = \delta_j, \quad j=1,2,3,4.$$

This is a *rational expectations* assumption which is standard in many learning models, and which is based on the assumption that doctors' beliefs should be right *on average* about the true quality of *omeprazole*. However, even with rational expectations, doctors still face uncertainty about its quality, as parameterized by the prior variances $\sigma^2_{\delta,j} \equiv \mathrm{Var}_1\ \delta_j,\ j=1,\ldots,4$.

For the more complex multivariate learning model with varying numbers of signals per diagnosis employed in this paper, the argument for non-identification is more difficult because the expressions for the posterior mean and variance cannot be written in the manner above, as a function of r_t, $\sum_t \mu_{t'}$, and the estimated parameters. Hence, we explore the separate identification of the γ and prior mean parameters by simulation, and our findings are presented in Appendix C.

Related work with the same data set: In ongoing work, one of the authors is using the same data set to estimate an explicitly dynamic (forward-looking) model of learning (cf. Crawford and Shum, 2000). Since the real world prescription process is much more complicated than either of these approaches, the models in the two papers accommodate contrasting sets of simplifying assumptions to shed light on different aspects of the learning problems which we expect to be important in pharmaceutical markets. We discuss several important differences between the two papers here.

First, the empirical questions considered in the two papers are quite different. The current paper addresses the question of new good *entry*, and focuses on considering micro-level explanation for *omeprazole*'s aggregate diffusion pattern. For this reason, we assume here that agents are uncertain about the quality of only *omeprazole*, but not the other drugs. The Crawford–Shum paper, on the other hand, addresses the issue of patient–drug *matches*, and focuses on estimating a model which explains the observed treatment lengths and "switches" of patients from one drug to another. Since matching problems arise only when agents face uncertainty about the returns from a *number* of competing choices, the Crawford–Shum model assumes that patients are ignorant of the relative qualities of *all* the competing drugs, not just *omeprazole*.[12]

Second, the model considered in Crawford and Shum (2000) is fully dynamic, and features patients who choose drugs via a dynamic discrete-choice optimization problem. Since the computational burden of such a model is quite severe, the information structure is kept quite simple, and no attempt is made to accommodate informational spillovers across patients at the doctor-level. In the current paper, however, we accommodate a more complicated information structure (including these spillovers) by abstracting away from the dynamic forward-looking aspect of the learning problem.

[12] The matching problem resembles the well-known "multi-armed bandit" problem in decision theory.

4. Data

The data used in this analysis was collected by the Italian National Institute of Health. It records, for a 10% sample [13] of the doctors in the metropolitan area of Rome, all prescriptions of anti-ulcer drugs (therapeutic class A02B [14]) to all their patients during a 3-year period (1990–1992). A prescription episode is the unit of observation in the data set. The data set contains more than 660,000 observations, each of which records the identity of the patient, the prescribing doctor, the drug prescribed, and the year and month of the prescription: 326 doctors, and 174,000 patients are represented in the data. The median number of prescriptions for the 326 doctors is around 2000 prescriptions during the 3-year sample period: 10% of the in-sample doctors have less than 1300 prescriptions, while only 10% have more than 2800 prescriptions. Appendix B provides more details on the data, in particular describing the covariates which we use when estimating the learning model.

The anti-ulcer drug market: The anti-ulcer drug market is the largest therapeutic drug market worldwide. It is naturally segmented, with preferred treatments differing across segments depending on the severity of the diagnosis, as summarized in the first three columns of Table 1.

The two most common diagnoses requiring treatment using anti-ulcer drugs are peptic ulcers and Gastroesophageal Reflux Disease (GERD). A peptic ulcer is an area of

Table 1
Segmentation in the anti-ulcer drug market: diagnoses and treatments

Diagnosis	Treatment[a]	Preferred drugs[a]	Empirical distinction[b]	Frequency	Percent
(1) Minor heartburn	Drugs or no Prescription	Anti-acids	Patient has $\leqslant 2$ in-sample prescriptions	135,466	20.49
(2) Pathological hypersecretory conditions	Anti-ulcer drugs	*omeprazole*	$Q \geqslant 133\%$ average monthly quantity for ulcer in *every* month	32,362	4.89
(3) Attack therapy for GERD or peptic ulcer	Anti-ulcer drugs	*omeprazole*	$Q \geqslant 133\%$ average monthly quantity for ulcer in the month	132,361	20.02
(4) Maintenance therapy for GERD or peptic ulcer	Anti-ulcer drugs	H_2-antagonists	$Q < 133\%$ average monthly quantity for ulcer in the month	361,054	54.60

[a] Medical Economics Co., 1997.

[b] Prescription assigned to diagnoses by the authors using the daily dosage for the average patient requiring an ulcer treatment as suggested by Medical Economics Co. (1997).

[13] The doctors in the sample were not chosen following any sampling technique, since the only information available was their "id" number.

[14] This four digit code, the ATC code, is an international classification system according to which drugs are divided into different categories by target organ, mechanism of action, and chemical and therapeutic characteristics.

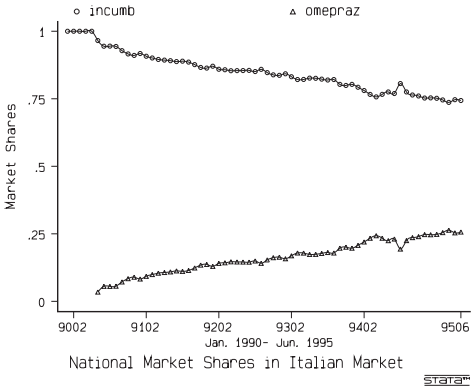

Fig. 1. Molecule market shares in the Italian anti-ulcer drug market. Monthly market shares of *omeprazole* and incumbent drugs. *Note*: *Omeprazole*razole-based drugs entered the market in June 1990. *Source*: IMS, Italy.

the stomach or duodenum that has been destroyed by stomach acid. Gastroesophageal Reflux Disease (GERD) refers to the backward flow of acid from the stomach up into the esophagus. During the 1990–92 period covered in our sample, traditional treatments for these two diagnoses consisted in alternating periods of intensive drug therapy to "attack" the ulcer, followed by periods of "maintenance" therapy where the drugs would be taken at lower dosages to prevent a recurrence. The third main diagnosis is a pathological hypersecretory condition. Different ailments are defined under this label, including Zellinger–Ellison syndrome. Finally, on occasion anti-ulcer drugs are also prescribed for minor heartburns.

Entry of omeprazole: Drugs based on the *omeprazole* molecule entered the Italian market in mid–1990. AstraZeneca holds the patent on *omeprazole*, and markets it worldwide under the brand name LOSEC (PRILOSEC in the US).[15] Its diffusion into the anti-ulcer market was gradual: as Fig. 1 shows, the market share of drugs based on *omeprazole* climbed from under 5% in the latter half of 1990 to about 15% in early 1992, and eventually up to 25% by the middle of 1995, despite little change in its price differential relative to other anti-ulcer drugs.[16] This gradual diffusion pattern echoes patterns found in many other product markets (cf. Urban et al., 1986).

[15] During the sample period, each molecule in the Italian market was usually marketed by more than one firm. Usually one of the firms was the patent-holder and the others were licensees. *Omeprazole* was introduced and jointly marketed by three firms (*Bracco*, *Malesci* and *Schering Plough*), none of which was the patent-holder. In this paper, we focus on competition *across* molecules and we do not distinguish between different brands of a particular molecule, i.e., between RANIDIL and ZANTAC, which are two different brands of the *ranitidine* molecule. Coscelli (2000) investigates competition among sellers of the same molecules.

[16] The price differential for *omeprazole* relative to a (market share-) weighted average of the prices for the other drugs decreased by only 7.4% over the sample period. Since prices were regulated, the change in relative prices reflects solely changes in the weights used in averaging the prices of the non-*omeprazole* drugs.

Patient heterogeneity: classifying patients in diagnosis groups: As discussed above, the anti-ulcer drug market is naturally segmented on the basis of diagnoses, so that a given molecule differs in its usefulness in treating different diagnoses (see Table 1). *Omeprazole* was found in clinical trials to be more effective than H_2-receptor antagonists in treating serious ulcers and severe esophagitis (a GERD-related condition). On the other hand, other clinical studies on animals found that *omeprazole* produced a dose-related increase in gastric carcinoid tumors, making it inappropriate for preventive maintenance therapies. [17]

These considerations motivated the multivariate learning model presented in Section 3, in which we permit *omeprazole*'s quality differential δ to vary across diagnoses. Since, we do not observe a patient's actual diagnosis in the data, we use the characteristics of patients' prescriptions to classify them into different diagnosis groups. After a review of the medical literature, we have classified observations into the four diagnosis groups based on the *length* and *intensity* of treatments. Table 1 describes the diagnoses, the possible treatments, the preferred drugs and the criteria we use to map prescriptions into diagnosis groups (Medical Economics Co., 1997).

The rightmost three columns in Table 1 show the proportion of the observations allocated to the 4 diagnostic classes. Cell 2 contains less than 5% of the prescriptions, while cell 4 contains more than 50% of the in-sample prescriptions. [18] The data also show that while the diffusion path of *omeprazole* is similar across the four cells, the market share of *omeprazole* increased more for diagnoses 2 and 3, which are those where the clinical trials indicated that *omeprazole* was more effective.

The large incidence of prescriptions for diagnosis 1 ("minor heartburns") evidenced in Table 1 is troubling. It suggests that "heartburn" may be a mischaracterization of these patients' gastro-intestinal woes—as discussed before, standard treatment for ulcers calls for a (short) period of intense "attack" therapy (using a powerful molecule such as *omeprazole*) followed either by a maintenance period or cessation of treatment. Diagnosis 1 may well encompass many patients who underwent attack therapy followed by cessation, for whom *omeprazole* would have been an appropriate molecule. [19] Therefore, as a specification check we have also estimated a model assuming patient homogeneity, without attempting to distinguish among diagnoses. [20]

[17] (Astra-Merck Inc., 1996) recommends that a standard treatment involving PRILOSEC last no longer than 8 weeks, making it inappropriate for preventive maintenance therapies. Continuous use of *omeprazole* was advised only for the pathological conditions (e.g., Zellinger–Ellison) where H_2-receptor antagonists were ineffective (Medical Economics Co., 1997). Furthermore, the molecule is too strong for heartburns.

[18] These proportions are broadly consistent with the incidence in the overall Italian population at the time (peptic ulcer 59%, GERD 5%, minor heart-burns 23%, and "others" 12%), according to a survey of Italian doctors in 1991 (Pertile, 1995).

[19] Furthermore, "survival bias" may be present here: perhaps those patients whose treatment lasted only two months or less (and therefore classified in diagnosis 1) are exactly those for whom *omeprazole* had a high success rate. More generally, if *omeprazole* is a very effective molecule, leading to less frequent prescriptions and shorter treatment lengths, then our classification scheme can lead to selection bias because, essentially, inclusion in a particular diagnostic cell is endogenous.

[20] Crawford and Shum (2000) explicitly model the effect of treatment option on treatment length ("success").

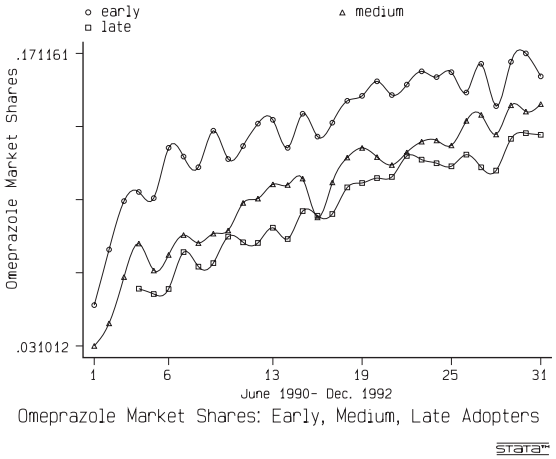

Fig. 2. Aggregate diffusion process: a composition artifact? Market shares of *omeprazole*: monthly averages for early, medium and late adopters *Late adopters*: those who did not adopt during the first three months; *Early adopters*: those who adopted both during the first and second month after entry; *Medium adopters*: all others. Market shares are the monthly averages across all doctors belonging to a certain group. Entry occurred in June 1990 ($t = 1$ in the graph). *Source*: authors' computations using the data described in Section 4.

Aggregate diffusion process: a composition artifact? Our learning model posits an individual doctor-level explanation for the aggregate diffusion of *omeprazole*. Our data indicate that almost 30% of the in-sample doctors "adopted" (i.e., prescribed *omeprazole* to at least one patient) in the first month after entry, 60% adopted before the end of the second month, and almost 75% prescribed *omeprazole* before the end of the third month. Finally, by the end of the tenth month *every* doctor had prescribed *omeprazole* to at least one patient. Given this heterogeneity in first adoption times, we wish to ensure that the market-level diffusion process is not simply a composition effect due to across-doctor heterogeneity in first adoption times.

Fig. 2 shows that even after controlling for first adoption times, a pronounced diffusion trend still persists in the data. In particular, we divided the doctors in our sample into three groups: early, medium and late adopters.[21] Early adopters are the doctors who prescribed *omeprazole* both in the first and second month after entry (approximately 25% of the sample), late adopters are all the doctors who did not adopt *omeprazole* in the first three months (approximately 25% of the sample), while the medium adopters category includes the other doctors. The graph clearly demonstrates that the diffusion process in market shares persists across the three groups, suggesting that the observed diffusion process is not simply an artifact due to aggregation across doctors with different first adoption dates.[22] Having established this, we next describe our econometric model.

[21] Finer categorizations of the adoption times yielded the same results.

[22] This evidence also rules out theories of "immediate learning" whereby doctors learn *omeprazole*'s true quality immediately upon making their first prescription. This is because otherwise we would observe the group market share for the *early adopters* jump to the asymptotic one immediately after the first prescription.

5. Empirical model

Our data set consists of observations of doctors' prescription sequences of different anti-ulcer drugs over time. Given our assumption that, conditional on diagnosis, all the patients are homogeneous,[23] we can aggregate up to the (doctor, month, diagnosis) level. Therefore, a single observation in this model is the number of doctor i's patients in the diagnostic cell j which are prescribed *omeprazole* in month $t(r_{jt}^i)$ out of the total number of patients with diagnosis j (N_{jt}^i) who are prescribed *any* anti-ulcer drug in month t by doctor i.

Assuming that the ε's are distributed i.i.d. (over patients and time) extreme value, the likelihood function for a given observation takes the *grouped logit* form. Throughout, we use the shorthand notation $V_t^{i*} \equiv \alpha x^i + \beta \Delta p_t + \xi(t) + \mathrm{E}_t\delta_j + \gamma \ \mathrm{Var}_t \ \delta_j$ (where $\gamma = 0$ for the specifications where doctors are risk-neutral). Then, omitting the i superscript,

$$\mathrm{Prob}(r_{jt} \mid N_{jt}) = \binom{N_{jt}}{r_{jt}} \left(\frac{\exp(V_t^{i*})}{1+\exp(V_t^{i*})}\right)^{r_{jt}} \left(\frac{1}{1+\exp(V_t^{i*})}\right)^{N_{jt}-r_{jt}}. \tag{5.12}$$

A significant problem we encounter in the estimation process is that while the doctor observes the realizations of the signals, μ_{jt}, which she uses to derive the posterior means $\mathrm{E}_t\delta_j$ and variances $\mathrm{Var}_t\delta_j$ of *omeprazole*'s quality, we as econometricians do not. This implies that the likelihood function for a given sequence of prescription frequencies for a given (doctor–diagnosis type) combination involves a multivariate integral over the distribution of the unobserved signals:

$$\begin{aligned}&\mathrm{Prob}(r_{11},\ldots,r_{1T},,\ldots,r_{J1},\ldots,r_{JT}|N_{j1},\ldots,N_{jT},,\ldots,N_{J1},\ldots,N_{JT})\\ &\quad=\int_{-\infty}^{\infty}\left[\prod_{t=1}^{T}\prod_{j}\binom{N_{jt}}{r_{jt}}\left(\frac{\exp(V_t^{i*})}{1+\exp(V_t^{i*})}\right)^{r_{jt}}\times\left(\frac{1}{1+\exp(V_t^{i*})}\right)^{N_{jt}-r_{jt}}|\vec{\mu}\right]\\ &\qquad\times \mathrm{d}F(\vec{\mu}),\end{aligned} \tag{5.13}$$

where $\vec{\mu}$ is the vector of signals observed by doctor i for patients with diagnosis j.

Since a doctor observes one signal for each patient to whom she prescribes *omeprazole*, the dimensionality of $\vec{\mu}$ is $R_T \equiv \sum_{j=1}^{J}\sum_{t=1}^{T} r_{jt}$, the number of prescriptions of *omeprazole* which the doctor wrote during the entire 31-month sample period to *all the* patients. In our data, R_T can often exceed 20 or 30, making the integral intractable using traditional quadrature methods. We use simulation techniques to evaluate these integrals, and estimate our model using simulated maximum likelihood.

Simulated maximum likelihood estimation: We employ a simple frequency simulator for the grouped logit choice probabilities (5.13).[24] This simulation procedure consists of three steps. First of all, for each doctor i and period t, we draw M vectors of

[23] This assumption implies that, among all the patients receiving diagnosis j by doctor i at time t, the unobserved i.i.d. component drives whether or not a particular patient receives *omeprazole*. This is a standard assumption in group data models. More general treatments would require multiple integration over patient heterogeneity distribution within each cell.

[24] Alternative simulators, such as the GHK simulator, are not applicable to our setting. GHK is useful for evaluating rectangular, multivariate truncated probabilities, but Eq. (5.13) does not satisfy those criteria.

signals: $\vec{\mu}_{it}^{m}$, $m = 1,\ldots,M$. Each vector is of dimension $\sum_j r_{ijt}$, the total number of times that doctor i prescribed *omeprazole* in period t, and drawn from a multivariate normal distribution with mean $\vec{\delta}_{it}$ and Σ_{it}, where these two quantities depend on the model parameters and are induced by the one-factor assumptions (cf. Section 3.2.1).[25]

Secondly, for each sequence of signals, we calculate the 31 (one for each month) means $\mathrm{E}_t\delta_j$, $t = 1,\ldots,31$ using the standard normal conditional mean formula (reported in Eq. (3.9)). Furthermore, the posterior mean (variance) of δ_j in month t will serve as the prior mean (variance) for month $t+1$. Having calculated the $\mathrm{E}_t\delta_j$'s for $t = 1,\cdots,31$ for a given drawn sequence of signals, we can calculate the grouped logit probabilities (in Eq. (5.12)) by substituting the calculated $\mathrm{E}_t\delta_j$'s and $\mathrm{Var}_t\delta_j$'s. Finally, we average the calculated grouped logit likelihood function (one for each drawn sequence) over all the drawn sequences. In the results reported in this paper, we used $M = 10$.

Hence, the log-likelihood function for all (doctor–diagnosis type) observations is

$$\mathrm{Log}\ L = \sum_{i=1}^{N} \log \frac{1}{M} \sum_{m=1}^{M} \left[\prod_{t=1}^{31} \prod_{j=1}^{J} \binom{N_{jt}}{r_{jt}} \left(\frac{\exp(V_t^{i*})}{1+\exp(V_t^{i*})} \right)^{r_{jt}} \left(\frac{1}{1+\exp(V_t^{i*})} \right)^{N_{jt}-r_{jt}} |\vec{\mu}^{m} \right], \tag{5.14}$$

where $N = 326$ is the total number of doctors in the sample and J, the number of diagnosis groups, is 4 (the signals $\vec{\mu}$ are now indexed by m, which labels the drawn sequences).

Pakes and Pollard (1989) derive a general asymptotic theory for estimators obtained by maximizing simulated objective functions. We utilize an asymptotic approximation for the variance–covariance matrices of the estimators which are based on the assumption that $M/\sqrt{N} \to \infty$ as $N \to \infty$. Under these assumptions, the familiar "outer product of the gradient" form for the asymptotic variance–covariance matrix is valid. Later, we discuss the validity of using this approximation for statistical inference.

5.1. Controlling for unobserved heterogeneity across doctors

Our learning model provides a structural (or behavioral) explanation for serial correlation (conditional on covariates like *omeprazole*'s price and doctor characteristics) at the individual doctor level in prescriptions of *omeprazole*. In order to ensure that learning is explaining the serial correlation, we must also control for unobserved heterogeneity, which can generate (spurious) serial correlation.[26] We assume that unobserved

[25] More precisely, we draw M $\sum_j r_{ijt}$-dimension vectors of standard normal variates. Then, for each vector v_{it}^{m}, for $m = 1,\ldots,M$, we generate signals via $\vec{\mu}_{it}^{m} = \vec{\delta}_{it} + \Sigma_{it}^{-1/2} v_{it}^{m}$ where $\Sigma_{it}^{-1/2}$ denotes the Cholesky factorization of Σ_{it}. We do not construct the signals based on the one-factor representation (3.7), but rather generate them directly from the multivariate normal distribution induced by the one-factor representation.

[26] To see this, assume that doctors know the true quality of *omeprazole*, but vary in their (time-invariant) propensity to prescribe it. Doctors with high propensities will prescribe *omeprazole* often, while those with low propensities prescribe a small amount—leading to a high degree of serial correlation in *omeprazole* prescription at the individual doctor level.

heterogeneity is summarized by a variable κ_i, which shifts the intercept of doctor i's net per-period utility from prescribing *omeprazole* at time t. The time-invariant κ_i captures doctor i's unobserved propensity to prescribe the *omeprazole*-based drugs over the incumbent drugs, arising from, for example, promotional activity to doctor i from manufacturers of *omeprazole*-based drugs, which are unobserved by the econometrician.

κ_i is modeled as a random effect: we assume that κ_i is drawn (identically across all doctors i) from a parametric distribution $G(\kappa;\Gamma)$, where Γ are parameters to be estimated. Therefore, the log-likelihood function for the prescriptions written by doctor i becomes:

$$\log \int_{\infty}^{\infty}\int_{\infty}^{\infty}\left[\prod_{t=1}^{31}\prod_{j=1}^{J}\binom{N_{jt}}{r_{jt}}\left(\frac{\exp(V_t^{i*}+\kappa)}{1+\exp(V_t^{i*}+\kappa)}\right)^{r_{jt}}\left(\frac{1}{1+\exp(V_t^{i*}+\kappa)}\right)^{N_{jt}-r_{jt}} |\vec{\mu}\right] \mathrm{d}F(\vec{\mu})\ \mathrm{d}G(\kappa;\Gamma). \tag{5.15}$$

In our work, we assume that $G(\kappa;\Gamma)$ is the normal distribution with mean zero and variance σ_κ^2, which is a parameter to be estimated. The multidimensional integral in this expression can be simulated, in a manner exactly analogous to that presented in the previous section.

6. Estimation results

Table 2 contains estimates from two specifications of the learning model. Model A is the baseline specification where we have categorized the various patient–visit observations into four diagnosis cells. In this specification, we assume that doctors are risk-neutral with preferences given by Eqs. (3.1) and (3.2), and allow for the prior means $\mathrm{E}_1\delta_j,\ j=1\ldots4$ to differ from the actual qualities $\delta_j,\ j=1\ldots4$.

In Model B, we consider the alternative specification of doctors' preferences which allows for risk aversion (so that the probability of prescribing *omeprazole* is given by Eq. (3.10). As discussed above, in order to identify the risk aversion parameter, we assume that doctors have rational expectations in the sense that their initial prior means for the quality of *omeprazole* coincide with the "true" quality, across all diagnoses (that is, $\mathrm{E}_1\delta_j=\delta_j,\ j=1,\ldots,4$).

6.1. Baseline results (Models A and B)

Since the likelihood function value is substantially higher for Model A (-52648.64) than for Model B (-53401.51), we focus on the Model A results first. However, as we discuss later, the plausibility of the Model B results will lead us to consider the simulated market shares using the results from both Models A and B.

The actual magnitudes of the estimated initial means and "true" quality differentials are difficult to interpret, since they depend highly on our choice of the other covariates to include in the utility specification, as well as our aggregation of all non-*omeprazole*

Table 2
Results for learning model. Number of observations: 10,106 (326 doctors/31 months). *Dependent variable*: $r^i_{jt} = k$ if doctor i prescribes omeprazole k times to patients with diagnosis j during month t, when her total number of anti-ulcer prescriptions to patients with diagnosis j is N^i_{jt}[a]. Model A: Baseline four-diagnosis specification; Model B: Specification with risk aversion

	Model A		Model B	
	Estimates	S.E.	Estimates	S.E.
Choice characteristics				
Δp : (PRICE$_{omep}$ − PRICE$_{og}$)/1000	−0.0098	0.0909	−0.0381	0.0723
γ:RISK AVERSION PARAMETER	—	—	−2.8304	0.1178
Doctors' characteristics				
HERFINDAHL PRODUCT LEVEL	0.0088	0.0018	−0.0082	0.0016
HERFINDAHL MOLECULE LEVEL	−1.0725	0.0217	−0.9762	0.0183
QUANTITY/100	−0.0095	0.0004	−0.0155	0.0004
Learning process parameters				
Initial distribution				
INITIAL MEAN–DIAG. 1 $E_1\delta_1$	−2.9482	0.0350		
INITIAL MEAN–DIAG. 2 $E_2\delta_2$	−2.1517	0.0203		
INITIAL MEAN–DIAG. 3 $E_3\delta_3$	−2.4244	0.0191		
INITIAL MEAN–DIAG. 4 $E_1\delta_4$	−3.2540	0.0264		
INITIAL VAR-COV MATRIX	see Appendix E.1		—	—
Other parameters				
TRUE QUALITY DIFF (δ_1)	−1.3665	0.0167	−3.2101	0.0196
TRUE QUALITY DIFF (δ_2)	−0.0920	0.0152	−1.8495	0.0141
TRUE QUALITY DIFF (δ_3)	−0.7838	0.0197	−2.1291	0.0135
TRUE QUALITY DIFF (δ_4)	−1.6024	0.0143	−2.5386	0.0131
VAR–COV MATRIX OF SIGNALS	see Appendix E.2		—	—
Unobserved heterogeneity				
σ_κ	0.6884	0.0052	0.8717	0.0031
TIME TREND (QUADRATIC)	yes		yes	
Log-likelihood fxn	−52,648.64		−53,401.51	
M (# sim. draws)	10		10	

[a]We have set σ_0^2 (the variance of the common component θ's) equal to 1, as our attempts to estimate it have not been successful.

drugs into a composite "outside" good. Therefore, in this section, we focus on differences between the initial and true δ's, and also between the δ's across different diagnoses.

Our estimates of the δ's, the true quality differentials, show that they exceed the initial means across all diagnoses (that is, $E_1\delta_j < \delta_j,\ j = 1,\ldots,4$). This is readily interpreted to mean that doctors are initially overly pessimistic about *omeprazole*'s quality at the time of its entry. This initial pessimism biases doctors against prescribing *omeprazole*, and is an important aspect of incumbent advantage in this market. As we will see later, risk aversion provides a plausible alternative explanation for doctors' initial reluctance to prescribe *omeprazole*, even without initial pessimism.

For Model A, the ordering of the initial means $E_1\delta_j,\ j=1,\ldots,4$ is 2, 3, 1, and 4. Since the true qualities $\delta_j,\ j=1,\ldots,4$ follow the same ordering, our results imply that, while doctors systematically underestimate *omeprazole*'s quality upon its entry into the market, they are correct in their assessment of *omeprazole*'s relative efficacy across different diagnoses. These estimates indicate that *omeprazole* is most beneficial (relative to the basket of incumbent drugs) in treating pathological hypersecretory conditions (diagnosis 2), which is consistent with the medical evidence cited earlier.[27]

Appendices E.1 and E.2 contain the estimated initial (i.e., before *omeprazole*'s entry) variance–covariance matrix, as well as the estimated variance–covariance matrix for a hypothetical vector of disturbances in the signals (the v_{jtk}'s) for a time period t, for all the models which we estimated. For the Model A results, the estimated magnitudes for the initial variance–covariance matrix are extremely small (roughly three orders of magnitude smaller than the estimates of the initial means), implying a great deal of confidence in these initial (and, as we saw above, pessimistic) estimates.

On the other hand, the finding that doctors are very confident in their overly pessimistic initial estimates is troubling. Fortunately, the magnitudes of the diagonal elements of the estimated variance–covariance matrix of the signals are also small, suggesting that the signals are very precise, so that doctors update "up" very quickly after a few prescriptions of *omeprazole* despite their initial pessimism. The results indicate that the signals from diagnosis 3 patients—those undergoing "attack" therapy—are the most precise. In Section 6.2 below, we present some simulation results which illustrate the extent to which doctors update their beliefs concerning *omeprazole*'s quality differential, as implied by these estimates.

Information spillovers: The spillover parameters (which are the off-diagonal elements in the matrices reported in Appendix E.2) are precisely estimated. The estimates of these parameters indicate mostly positive spillovers, except for the signals associated with diagnosis 4 ("maintenance" therapy). The negative values for the covariances between the signals for diagnosis 4 and all other diagnoses suggest that a positive outcome in prescribing *omeprazole* to a patient for diagnoses 1–3 leads doctors to regard *omeprazole* as less attractive for maintenance purposes (and vice versa). This is not surprising given information in the medical press, since (as described before) maintenance therapy typically requires longer than the 8 weeks which is deemed the appropriate length of a treatment involving *omeprazole*. This confirms the last column of Table 1 (derived independently from medical sources) that the anti-ulcer market is *horizontally differentiated*, in the sense that doctors will rank the two alternatives (*omeprazole*-based drug versus incumbent drugs) differently depending on the diagnosis of the patient to whom she is prescribing. However, as we will see below, these negative spillovers have important implications for the rate of learning.

[27] On the other hand, we would be cautious in interpreting the negative magnitudes for the δ's to mean that *omeprazole*'s true quality is lower than that of the competing drugs, since this result partly reflects our aggregation scheme, whereby all non-*omeprazole* drugs are aggregated into a composite drug. Since *omeprazole*'s market share does not exceed 30%, even by the end of our sample period, it is not surprising that we estimate the δ's to be negative.

The coefficient on Δp, the price differential between *omeprazole* and all the other anti-ulcer molecules, is negative but not significant across all models. This finding is not surprising, since the Italian health care institutions (almost all anti-ulcer drugs are paid for by the National Health Service) do not force doctors to internalize the costs of the drugs that they prescribe. The two doctor-level covariates which measure dispersion in prescription patterns prior to *omeprazole*'s entry are statistically significant but have different signs: we find that greater dispersion at the *molecular* level (lower value for the molecular-level Herfindahl index) increases the probability of prescribing *omeprazole*, but not greater dispersion at the product level. This is consistent with previous evidence (cf. Stern and Trajtenberg, 1998) that these variables proxy for doctors' willingness-to-experiment with new drugs.

Model B: Risk aversion: For the Model B specification, which allows doctors' utilities to be characterized by risk aversion, the ordering of the true qualities δ_j, $j=1,\ldots,4$ is 2, 3, 4, 1, which differs from the ordering in the Model A results. The Model B ordering is perhaps more plausible, since the medical literature seems to indicate that *omeprazole* is least appropriate for minor heartburns (Diagnosis 1). Furthermore, the estimates for the initial variance–covariance matrix (reported in Appendix E.1) are three orders of magnitude larger than the estimates for Model A, thereby implying much more initial uncertainty about *omeprazole*'s quality. In addition, the risk aversion parameter is estimated to be large in magnitude, and significant $\gamma(-2.8304)$. These findings suggest that initial pessimism (implied by the Model A results) and risk aversion (implied by the Model B results) constitute two observationally equivalent explanations for the diffusion patterns for *omeprazole* that we observe in the data. Given the plausibility of the Model B results, therefore, we examine counterfactual simulations for both Models A and B, despite the lower maximized likelihood function value for Model B.

In the appendix, we present and discuss results from three alternative specifications (models C–E) meant to gauge the robustness of the results discussed above. We will not discuss them further here.

6.2. Counterfactual simulations

The role of information: In Figs. 3 and 4, we decompose the across-time changes in the predicted market shares of *omeprazole* into a part due to outside sources of learning (parameterized by the quadratic time trend in our specifications), and a part due to learning by actual prescriptions using the estimated coefficients from Models A and B, to assess the ability of the learning model to fit the market-level diffusion trends of *omeprazole*. The curve marked avg represents the average doctor in our sample (whose evaluation of the quality of *omeprazole* changes in every period), and the curve marked nopres represents a doctor who never prescribes *omeprazole*, who therefore learns about it only on the basis of "word of mouth" and the dissemination of information in medical journals (as proxied by the quadratic time trend) alone. For purposes of comparison, the actual (in-sample) market shares of *omeprazole* for each diagnosis are also plotted, by the curve marked Actual.

It is clear that the results for both Models A and B predict the actual observed market share path remarkably well for "average" doctors (the avg lines), but severely

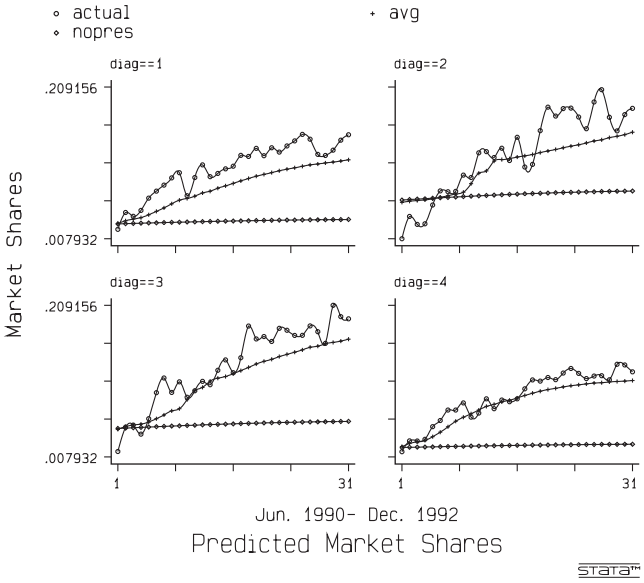

Fig. 3. Predicted market shares: using model a results. On the *y*-axis, market share of *omeprazole* in treating different diagnoses over time for three scenarios: (1) *actual*: actual (in-sample) market share of *omeprazole*; (2) *avg*: predicted market shares for doctor who prescribes the "average" amount of *omeprazole*; (3) nopres: predicted market share for doctor who never prescribes *omeprazole* (i.e., who learns about *omeprazole* on the basis of promotional activity and the time trend alone). Only time trend and posterior means/variances vary over time; the other covariates (price and doctor characteristics) are set at their sample means.

underpredict for doctors who do not prescribe at all (the nopres lines). These graphs suggest that the likelihood that a doctor prescribes *omeprazole* does not change much over time unless the doctor actually starts prescribing it. This provides some micro-level support for the results presented in Azoulay et al. (2003), obtained using aggregate data.

However, Figs. 3 and 4 also illustrate that the "initial pessimism" story from the Model A results, and the risk aversion story from the Model B results do equally well in matching the observed diffusion pattern. While both initial pessimism and risk aversion fall under the rubric of "informational" barriers to entry, these results also suggest that it is difficult in practice to separately identify initial pessimism from risk aversion, since they have observationally equivalent implications for the resulting diffusion paths of new products. Our analysis does not resolve this issue, but points out, in the framework of our learning model, *how large* the initial pessimism or risk aversion must be in order to generate the observed diffusion patterns.

The effects of across-diagnosis spillovers on diffusion paths: In a second set of simulations, we examined the importance of across-diagnosis spillovers in explaining the observed diffusion patterns. Using the results from Model A, we simulated prescription patterns for the doctors in the sample, under the alternative assumption that the signals they received from prescribing *omeprazole* to their patients are only correlated within, but not across diagnosis groups.

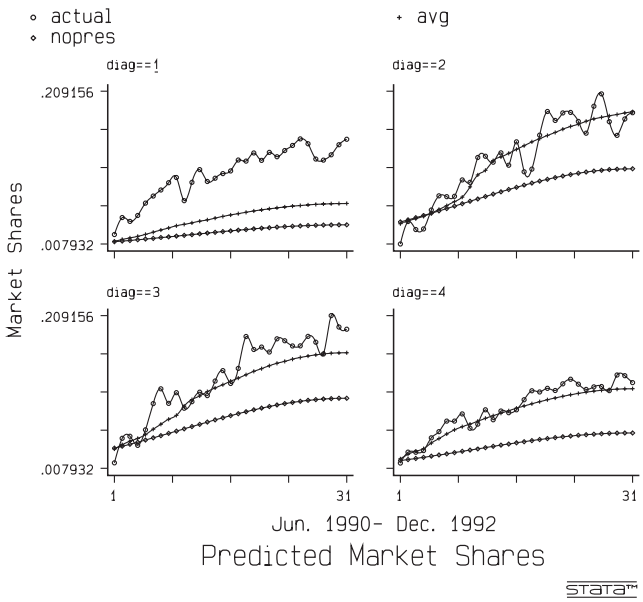

Fig. 4. Predicted market shares: using model B results. On the *y*-axis, market share of *omeprazole* in treating different diagnoses over time for three scenarios: (1) *actual*: actual (in-sample) market share of *omeprazole*; (2) *avg*: predicted market shares for doctor who prescribes the "average" amount of *omeprazole*; (3) *nopres*: predicted market share for doctor who never prescribes *omeprazole* (i.e., who learns about *omeprazole* on the basis of promotional activity and the time trend alone). Only time trend and posterior means/variances vary over time; the other covariates (price and doctor characteristics) are set at their sample means.

Differences in the speed of learning with and without spillovers are quantified in the graphs in Fig. 5. Comparing the top and bottom panels in that figure, it is evident that the convergence of posterior means to the true δ's is much quicker *in the absence of spillovers*. While this finding may appear paradoxical, we note that this faster learning does not imply that, in the absence of spillovers, signals are "more informative" in any way; rather, doctors sample *omeprazole* more frequently in the absence of spillovers, and it is this increased sampling which leads to faster learning.

7. Conclusions

In this paper, we specified and obtained parameter estimates from a structural diffusion model of new product entry into an experience good market. This model is based on a Bayesian learning process, specified at the individual level. Using this model, we employ a unique panel data set on doctor-level prescription histories to explicitly quantify the magnitude of informational barriers to entry into the anti-ulcer drug market. In order to accommodate features specific to the pharmaceutical prescription process, we extend the basic learning model to allow for spillovers across all the patients of a given doctor, as well as heterogeneity in informativeness across patients.

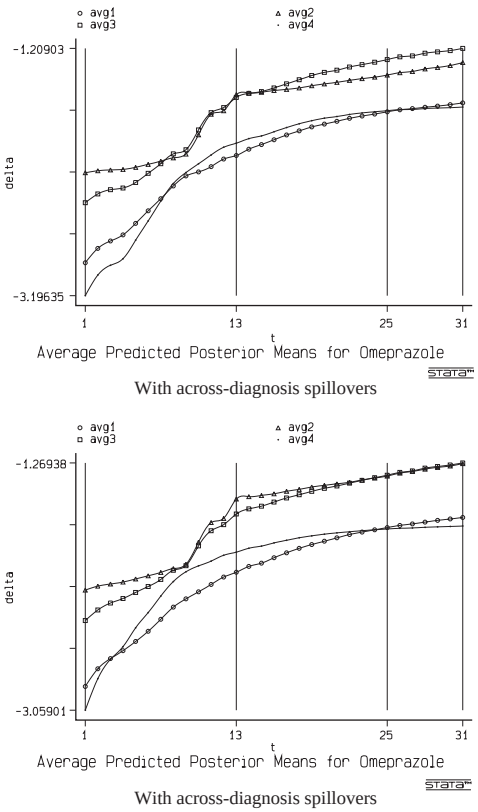

Fig. 5. Simulated average posterior means: using model A results. True deltas: Diagnosis 1 (−1.3665); Diagnosis 2 (−0.0920); Diagnosis 3 (−0.7838); Diagnosis 4 (−1.6024).

We have focused on the entry of the new molecule *omeprazole* into the Italian market in 1990. Although *omeprazole* represented a very significant innovation upon previous treatments, its market share grew only gradually during our sample period (1990–1992). According to our model, *omeprazole* suffered a competitive disadvantage upon its entry into the Italian anti-ulcer market which we attribute to two aspects of doctors' preferences: either (i) initial pessimism about, or (ii) aversion to the risk due to uncertainty about *omeprazole*'s quality. Even after controlling for time effects and doctors' heterogeneity, we find that most of the observed gradual growth in *omeprazole*'s market share is predominantly explained by doctors' accumulation of first-hand experience of the drug via actual prescriptions. This finding implies that one potentially effective strategy for manufacturers of new drugs to overcome the informational advantages enjoyed by incumbent producers is to offer free samples, which represent a direct attempt by pharmaceutical companies to convince doctors to try the drug and accumulate the requisite first-hand experience. Free samples were, in fact, widespread in the Italian pharmaceutical market in the early 1990s. As a postscript, we add that

by 1996, LOSEC (the flagship drug based on *omeprazole*) became the world's best selling drug, grossing more than US$2 billion, and gaining about half of the worldwide anti-ulcer market.

Our work also raises the interesting issue that, in practice, it is difficult to separately identify initial pessimism from risk aversion in empirical learning models, since they have observationally equivalent implications for the resulting diffusion paths of new products. Even with our unique doctor-level prescription data, it is difficult to disentangle these two competing explanations. In future work, we plan to explore what types of data would enable researchers to distinguish between these explanations.

Acknowledgements

We thank Dr. Giuseppe Traversa and Roberto Da Cas at the *Istituto Superiore di Sanita*' and IMS Italy for providing the data, and Tim Bresnahan, Costas Meghir, Fiona Scott Morton, Peter Reiss, and Frank Wolak for their suggestions. We acknowledge financial support from a European Commission TMR fellowship #ERBFMBICT972232 (Coscelli) and a Sloan Foundation dissertation fellowship (Shum). We are grateful to an associate editor and two anonymous referees for detailed comments on earlier drafts.

Appendix A. Illustrative example of the learning model

To illustrate the model described in Section 3.1, let us assume that $J = 4$, and that in period t, two signals are observed for the first two diagnoses, and one each for diagnoses 3 and 4 (in other words, we observe $\vec{\mu}_t \equiv (\mu_{1t1}, \mu_{1t2}, \mu_{2t1}, \mu_{2t2}, \mu_{3t1}, \mu_{4t1})$). Also assume that, conditioning on all the signals received up to (but not including) period t, the doctor believes that the vector $\vec{\delta}$ has the following distribution:

$$\begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{pmatrix} \sim \left(\begin{bmatrix} \mathrm{E}_t\delta_1 \\ \mathrm{E}_t\delta_2 \\ \mathrm{E}_t\delta_3 \\ \mathrm{E}_t\delta_4 \end{bmatrix} \equiv \mathrm{E}_t\vec{\delta}, \begin{bmatrix} \gamma_{1t}^2 & & & \\ \gamma_{12t} & \gamma_{2t}^2 & & \\ \gamma_{13t} & \gamma_{23t} & \gamma_{3t}^2 & \\ \gamma_{14t} & \gamma_{24t} & \gamma_{34t} & \gamma_{4t}^2 \end{bmatrix} \equiv \Sigma_{\delta t} \right). \tag{A.1}$$

Given these assumptions, then, in period t, doctor i believes that the vector of signals $\vec{\mu}_t$ has mean

$$\vec{\delta}_{\mu,t} \equiv \begin{bmatrix} \mathrm{E}_t\delta_1 \\ \mathrm{E}_t\delta_1 \\ \mathrm{E}_t\delta_2 \\ \mathrm{E}_t\delta_2 \\ \mathrm{E}_t\delta_3 \mathrm{E}_t\delta_4 \end{bmatrix} \tag{A.2}$$

and variance–covariance matrix

$$\Sigma_{\mu,t} \equiv \begin{bmatrix} \gamma_{1t}^2+\rho_1^2\sigma_\theta^2+\sigma_{\eta 1}^2 & \gamma_{1t}^2+\rho_1^2\sigma_\theta^2 & \gamma_{12t}+\rho_1\rho_2\sigma_\theta^2 & \gamma_{12t}+\rho_1\rho_2\sigma_\theta^2 & \gamma_{13t}+\rho_1\rho_3\sigma_\theta^2 & \gamma_{14t}+\rho_1\rho_4\sigma_\theta^2 \\ & \gamma_{1t}^2+\rho_1^2\sigma_\theta^2+\sigma_{\eta 1}^2 & \gamma_{12t}+\rho_1\rho_2\sigma_\theta^2 & \gamma_{12t}+\rho_1\rho_2\sigma_\theta^2 & \gamma_{13t}+\rho_1\rho_3\sigma_\theta^2 & \gamma_{14t}+\rho_1\rho_4\sigma_\theta^2 \\ & & \gamma_{2t}^2+\rho_2^2\sigma_\theta^2+\sigma_{\eta 2}^2 & \gamma_{2t}^2+\rho_2^2\sigma_\theta^2 & \gamma_{23t}+\rho_2\rho_3\sigma_\theta^2 & \gamma_{24t}+\rho_2\rho_4\sigma_\theta^2 \\ & & & \gamma_{2t}^2+\rho_2^2\sigma_\theta^2+\sigma_{\eta 2}^2 & \gamma_{23t}+\rho_2\rho_3\sigma_\theta^2 & \gamma_{24t}+\rho_2\rho_4\sigma_\theta^2 \\ & & & & \gamma_{3t}^2+\rho_3^2\sigma_\theta^2+\sigma_{\eta 3}^2 & \gamma_{34t}+\rho_3\rho_4\sigma_\theta^2 \\ & & & & & \gamma_{4t}^2+\rho_4^2\sigma_\theta^2+\sigma_{\eta 4}^2 \end{bmatrix}. \tag{A.3}$$

Appendix B. Additional details regarding the data

Table 3 presents the summary statistics for the variables used in the estimates. The first block summarizes the number of observations we have for each (doctor, month) cell. There is an observation in a (doctor, month, diagnosis) cell if at least one patient receives the particular diagnosis by the doctor in that month. Therefore, the upper bound on the number of observations we have for any diagnosis is 10,106 (326 doctors $\times$ 31 months). The least frequent is diagnosis 2: in only 5567 (doctor–months) combinations at least one patient was diagnosed with a pathological hypersecretory condition. The "Mean" column indicates the average number of patients who respectively received a prescription of *any* anti-ulcer drug (the *total* row), or a prescription of *omeprazole* across the (doctor–month) cells *conditional on at least one* prescription of anti-ulcer drugs taking place. For example, for diagnosis 3 (attack therapy): on average there are 1.62 prescriptions of *omeprazole*, and 13.55 prescriptions of anti-ulcer drugs, conditional on at least one prescription of anti-ulcer drugs by doctor i in month s.

The second block summarizes the values for the covariates proxying for observed heterogeneity across doctors. They are: (i) the monthly average of Herfindahl index[28] at *brand* level; (ii) the monthly average of the Herfindahl index at the *molecule* level; and (iii) the average monthly quantity of anti-ulcer drugs prescribed. All of these are calculated based on doctors' observed prescription behavior *before* omeprazole's *entry*.

[28] The Herfindahl index is defined to be the sum of the squares of the market shares. A high value of the Herfindahl index indicates that the doctors are concentrated in their prescriptions across brands.

Table 3
Summary statistics

Variable	4 diagnoses cells ($N = 10,106$: DOCTORS=326*31=T)				
	# Obs.	Mean	St.Dev.	MIN	MAX
Dependent variables–number of prescriptions in each (doctor, month) cell					
DIAG 1: MINOR HEART-BURNS					
omeprazole	10,015	1.157	1.747	0	17
Total	10,015	11.466	6.86	1	54
DIAG 2: HYPERSECRETORY CONDITIONS					
omeprazole	5567	0.706	1.613	0	26
Total	5567	5.342	6.373	1	70
DIAG 3: ATTACK THERAPY					
omeprazole	8989	1.62	2.934	0	42
Total	8989	13.557	17.8	1	148
DIAG 4: MAINTENANCE THERAPY					
omeprazole	10,092	2.748	3.092	0	33
Total	10,092	30.651	12.94	1	101
Doctors' characteristics					
Herfindahl brand level	326	0.252	0.068	0.13	0.564
Herfindahl molecule level	326	0.503	0.113	0.232	0.805
Monthly quantity	326	798.37[a]	234.15	241	1682.08
Choice characteristics					
Price diff.	31	1,188.724[b]	30.661	1131.17	1237.15

[a]The unit of measurement is *defined daily doses* (days of therapy).
[b]Italian Lire; 1US$=1600 ItL in the early 1990s; therefore, the average price difference between a day of treatment with *omeprazole*, and a day of treatment with the outside good was about 75 cents (this is approximately 25% of the price of a day of treatment with *omeprazole*).

The third block summarizes drug-specific covariates. *Omeprazole* had on average a 25% price mark-up over its competitors.[29]

Appendix C. Identification of the risk aversion specification: multivariate case

In this section, we present and discuss simulation results pertaining to the identification of the parameters of the multivariate learning model employed in this paper, assuming that doctors are risk-averse. Specifically, we simulated two small data set of prescription sequences for 50 doctors (roughly one-sixth the size of our actual data set). In one data set, we allowed for both risk aversion (i.e., $\gamma \neq 0$) as well as non-rational

[29] Over the sample period, *omeprazole*'s price differential decreases, but since prices are tightly regulated in Italy, this decrease is due less to price decreases in the non-*omeprazole* molecules, than to shifts in the composition of the non-*omeprazole* "basket" towards cheaper therapeutic alternatives.

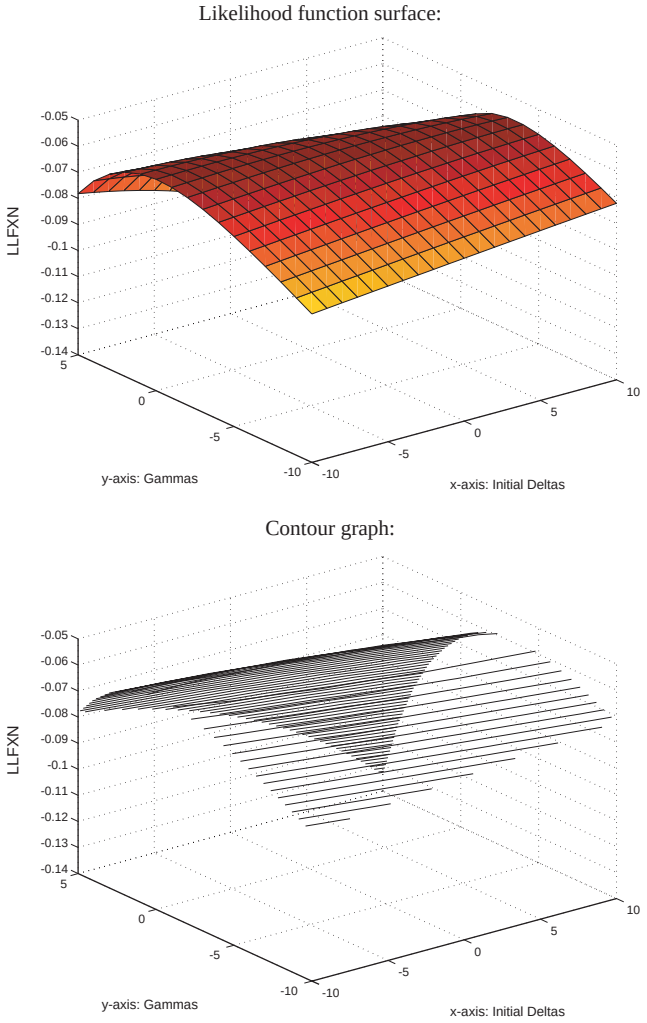

Fig. 6. Likelihood function for simulated data (without rational expectations) likelihood function surface. *Notes*: On the x-axis, we consider proportional deviations of $\vec{\delta}_1$ away from the true value. In particular, letting $\vec{\delta}_1^*$ and $\vec{\delta}^*$ denote, respectively, the true values of the prior mean and actual quality vectors used to generate the simulated data (these are the Model A estimates), we consider deviations of the prior mean vector $\vec{\delta}_1$ from the true value $\vec{\delta}_1^*$ given by $\vec{\delta}_1 = \lambda * (\vec{\delta}_1^*) + (1-\lambda) * (\vec{\delta}^*)$, $\lambda = -1.0, -0.9, -0.8 \ldots 0.8, 0.9, 1.0$ so that $\lambda = 0$ denotes the true data-generating value for $\vec{\delta}_1$, and $\lambda = 1.0$ denote the rational expectations case. Values of λ (multiplied by a factor of 10) are given on the x-axis. Values of γ are plotted on the y-axis. In this graph, the true data-generating value is given by $\gamma = -2$ and $\lambda = 0$.

expectations ($E_1 \delta_j \neq \delta_j, \ \forall j$). In the second data set, we allowed for risk aversion, but imposed rational expectations ($E_1 \delta_j = \delta_j, \ \forall j$).

Using these data sets, we graphed the likelihood function in Figs. 6 and 7. In each figure, the "true" data-generating values of the parameters is given by the point $(0, -2)$

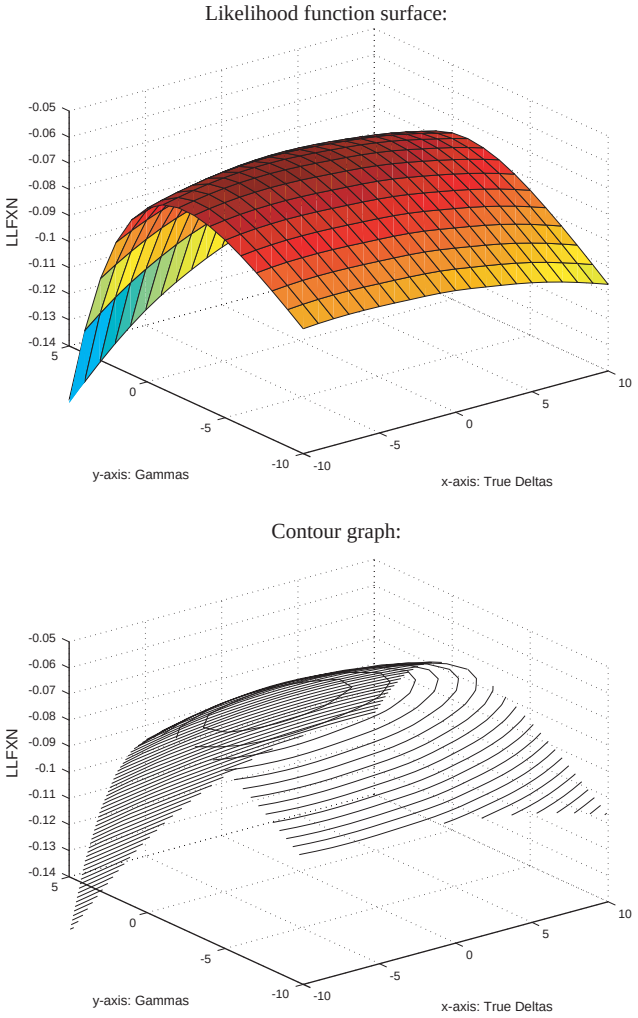

Fig. 7. Likelihood function for simulated data (with rational expectations) likelihood function surface. *Notes*: In the x-dimension, we consider proportional deviations of $\vec{\delta}$ away from the true value. In particular, we consider deviations of the prior mean vector $\vec{\delta}$ from the true value $\vec{\delta}^*$ given by $\vec{\delta}\vec{\delta} = \lambda * (\vec{\delta}^*), \lambda = -1.0, -0.9, -0.8 \ldots 0.8, 0.9, 1.0$ so that $\lambda = 1.0$ denotes the true data-generating value for $\vec{\delta}$, which is given by the Model B estimates. Values of λ (multiplied by a factor of 10) are given on the x-axis. Values of γ are plotted on the y-axis. In this graph, the true data-generating value is given by $\gamma = -2$ and $\lambda = 0$.

in $(x,\ y)$ space. The labeling of the axes in each figure is described in notes at the bottom of each figure. Next, we describe each figure in some detail.

In Fig. 6 the small range in color for the top graph indicates that the likelihood function without the rational expectations assumption is very flat. More significantly, the contour graph plotted at the bottom of the figure (each line gives combinations of

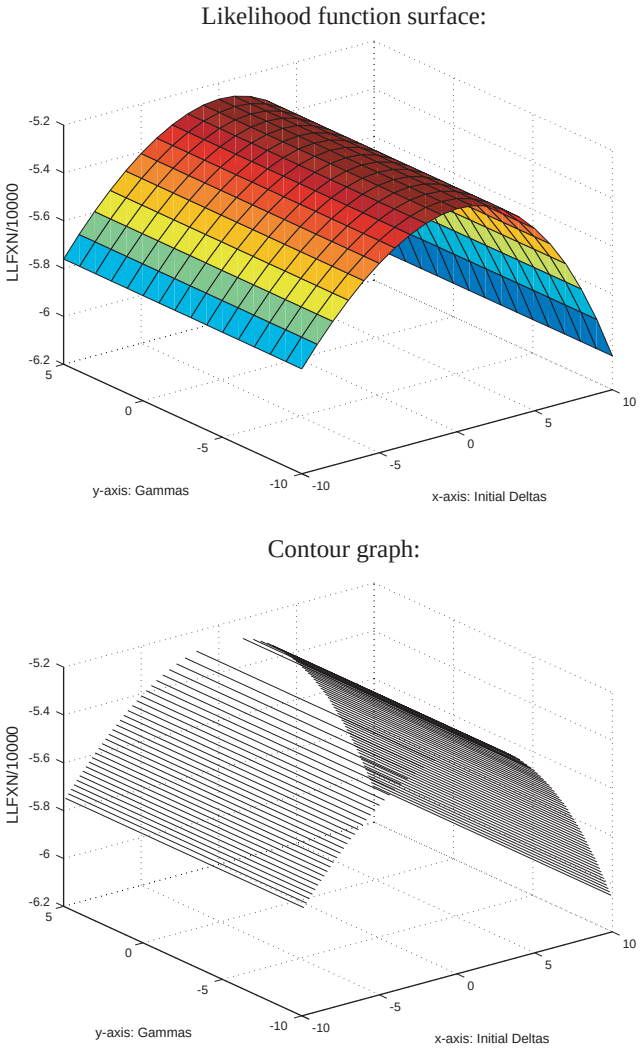

Fig. 8. Likelihood function for actual data (without rational expectations) likelihood function surface. *Notes*: On the x-axis, we consider proportional deviations of $\vec{\delta}_1$ away from the value given in the Model A estimates. In particular, letting $\vec{\delta}_1^*$ and $\vec{\delta}^*$ denote, respectively, the true values of the prior mean and actual quality vectors in the Model A results, we consider deviations of the prior mean vector $\vec{\delta}_1$ given by $\vec{\delta}_1 = \lambda * (\vec{\delta}_1^*) + (1-\lambda) * (\vec{\delta}^*)$, $\lambda = -1.0, -0.9, -0.8 \ldots 0.8, 0.9, 1.0$ so that $\lambda = 0$ denotes the Model A estimates for $\vec{\delta}_1$, and $\lambda = 1.0$ denote the rational expectations case. Values of λ (multiplied by a factor of 10) are given on the x-axis. Values of γ are plotted on the y-axis.

γ and the prior means for which the likelihood function has the same value) indicates that the contour lines are very linear and close to parallel,[30] which imply that γ and

[30] For the one-signal model above, the contour curves in $(\gamma,\ \delta_1)$ space are exactly linear, with slope $1/\sigma_1^2$.

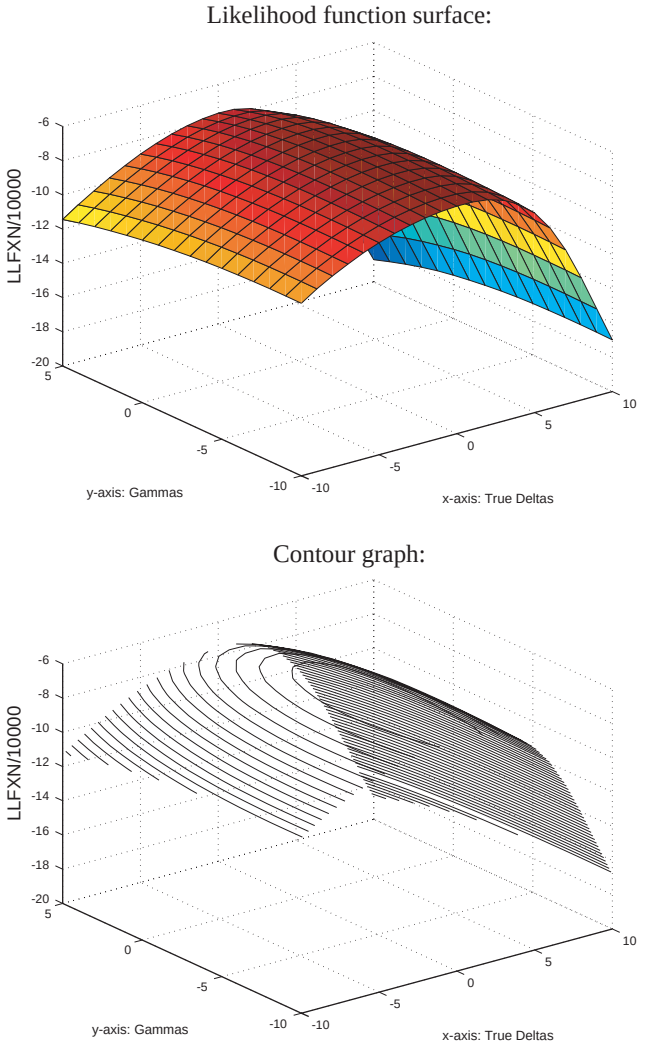

Fig. 9. Likelihood function for actual data (with rational expectations) likelihood function surface. *Notes*: In the x-dimension, we consider proportional deviations of $\vec{\delta}$ away from the Model B estimates (denoted by $\vec{\delta}$). In particular, we consider deviations of the prior mean vector $\vec{\delta}$ given by: $\vec{\delta} = \lambda * (\vec{\delta}^*)$, $\lambda = -1.0, -0.9, -0.8 \ldots 0.8, 0.9, 1.0$ so that $\lambda = 1.0$ denotes the Model B estimates. Values of λ (multiplied by a factor of 10) are given on the x-axis. Values of γ are plotted on the y-axis.

the prior means are not separately identified (at least locally to the true data-generating values).

For comparison's sake, we also computed the likelihood function for the data set simulated under the assumption of rational expectations (Fig. 7). The larger range of colors in the top graph indicates that this likelihood function is not so flat; furthermore,

the contour lines are circular implying that γ is now separately identified apart from the *true* means (at least locally to the true values).

Finally, Figs. 8 and 9 are the equivalent of Figs. 6 and 7, but graphed for the actual data used in this paper. The parallel contour lines are clearly evident in Fig. 8, which presents the likelihood function of the model with risk aversion, but without rational expectations. This explains why we were unable to jointly estimate γ as well as the prior mean parameters. In Fig. 9, which presents the likelihood function for the risk aversion model imposing rational expectations, we see that the contour lines (now in γ and δ space) are circular, indicating that these parameters are (at least locally) identified.

Appendix D. Robustness checks

Table 4 contains estimates from three alternative specifications (Models C–E) meant to gauge the robustness of the results discussed above.[31] First, in order to gauge the statistical superiority of the 4-diagnosis model, which is prone to the criticism that the criteria used to group patients in diagnosis cells may be endogenous, we also estimated a 1-diagnosis model (Model C) in which all the signals which a doctor observes about *omeprazole*'s quality are i.i.d. draws from a distribution with mean δ and variance σ_η^2. These estimates vary little qualitatively and quantitatively from those obtained from the 4-diagnosis model, thus providing some indirect evidence that this type of endogeneity is not a big problem here. On the basis of the likelihood function values alone, Model A provides, by far, the better fit of the data.

Second, we address potential worries about the validity of statistical inferences based on the reported standard errors given that we use only ten simulation draws.[32] To assess these issues, we re-estimated Model A for $M = 20$. The results, given as Model D in Table 4, show little difference in the estimates, or the standard errors. Indeed, the (average) percentage changes between the Models A and D results in (1) the parameter estimates; and (2) the standard errors, computed from the asymptotically valid covariance matrix are, respectively: (1) 1%, and (2) 19%. We take these rather small changes as an indication not only that variation in the parameter estimates due to simulation may be relatively small, but also that the parameter estimates and standard errors are relatively stable for an increased M, so that worries about the validity of statistical inferences based on the reported standard errors may be minimized.

Third, we estimated the Model E specification, where we included a full set of (month-specific) time dummies. Since our price data has variation only at the monthly level, we were not able to identify the price coefficient in this specification, and so

[31] The variance–covariance matrices corresponding to doctors' priors and the signals corresponding to Models D and E are given in Appendix E.

[32] Gourieroux and Monfort (1996, Chapter 3) point out that, as long as we assume $M/\sqrt{N} \to \infty$, the "outer product of the gradient" estimator for the asymptotic variance is valid.

Table 4
Results for learning model: robustness checks. Number of observations: 10,106 (326 doctors/31 months). *Dependent variable*: $r^i_{jt} = k$ if doctor i prescribes omeprazole k times to patients with diagnosis j during month t, when her total number of anti-ulcer prescriptions to patients with diagnosis j is N^i_{jt}.[a] Model C: One-diagnosis specification; Model D: 20 simulation draws; Model E: Include time period dummies rather than time trend

	Model C		Model D		Model E	
	Estimates	S.E.	Estimates	S.E.	Estimates	S.E.
Choice characteristics						
Δp: $(\text{PRICE}_{omep} - \text{PRICE}_{og})/1000$	−0.0217	0.0982	−0.0098	0.0245	—[b]	—
Doctors' characteristics						
HERFINDAHL PRODUCT LEVEL	−0.1977	0.0110	0.0088	0.0025	0.0089	0.0039
HERFINDAHL MOLECULE LEVEL	−0.6341	0.0173	−0.9719	0.0334	−0.3135	0.0116
QUANTITY/100	−0.0113	0.0004	−0.0095	0.0006	−0.0319	0.0005
Learning process parameters						
Initial distribution						
INITIAL MEAN–DIAG. 1 $E_1\delta_1$	−3.2808	0.0188	−2.9773	0.0396	−3.6940	0.0313
INITIAL MEAN–DIAG. 1 $E_2\delta_2$	—	—	−2.2071	0.0230	−3.1163	0.0289
INITIAL MEAN–DIAG. 1 $E_3\delta_3$	—	—	−2.4836	0.0240	−3.4170	0.0276
INITIAL MEAN–DIAG. 1 $E_4\delta_4$	—	—	−3.3396	0.0287	−4.0790	0.0259
INITIAL VAR–COV MATRIX	—	—	see Appendix E.1		see Appendix E.1	
INITIAL ST.DEV.	0.0214	0.0080	—	—	—	—
Other parameters						
TRUE QUALITY DIFF (δ_1)	−1.2802	0.0097	−1.4452	0.0204	−0.6797	0.0287
TRUE QUALITY DIFF (δ_2)	—	—	−0.0916	0.0199	0.0604	0.0246
TRUE QUALITY DIFF (δ_3)	—	—	−0.9050	0.0236	−0.3805	0.0308
TRUE QUALITY DIFF (δ_4)	—	—	−1.6911	0.0200	−1.8298	0.0377
VAR–COV MATRIX OF SIGNALS	—	—	see Appendix E.2		see Appendix E.2	
ST.DEV. OF SIGNALS	0.0581	0.0133	—	—	—	—
Unobserved heterogeneity						
σ_κ	0.7392	0.0046	0.5468	0.0079	0.0036	0.0404
TIME TREND (QUADRATIC)	Yes		Yes		No	
TIME PERIOD DUMMIES	No		No		Yes	
Log-likelihood fxn	−53656.61		−52380.56		−54347.21	
M (# sim. draws)	10		20		10	

[a]We have set σ_0^2 (the variance of the common component θ's) equal to 1, as our attempts to estimate it have not been successful.
[b]Price coefficient not identified given period-specific time dummies.

omit price from the model. Generally, the results are qualitatively and quantitatively similar to those obtained in Model A. One important difference is that the estimate for σ_κ, the standard error of the unobserved heterogeneity distribution, is much smaller in this specification (0.036) than in the Model A specification (0.6884). As we would expect, the period-specific dummies in the Model E specification may be soaking up much of what the previous specifications attributed to doctor heterogeneity.

Appendix E. Variance–covariance matrices: Models A, B, D, E

E.1. Initial variance–covariance matrices

For Model A Results :

$$10^{-4} * \begin{bmatrix} 2.117 & 0 & 0 & 0 \\ (0.179) & & & \\ 0 & 2.370 & 0 & 0 \\ & (0.204) & & \\ 0 & 0 & 1.808 & 0 \\ & & (0.153) & \\ 0 & 0 & 0 & 2.349 \\ & & & (0.197) \end{bmatrix}$$

For Model B results :

$$10^{-1} * \begin{bmatrix} 2.543 & 0 & 0 & 0 \\ (0.117) & 0 & 0 & 0 \\ 0 & 2.936 & 0 & 0 \\ 0 & (0.131) & 0 & 0 \\ 0 & 0 & 2.200 & 0 \\ 0 & 0 & (0.100) & 0 \\ 0 & 0 & 0 & 3.079 \\ 0 & 0 & 0 & (0.135) \end{bmatrix}$$

For Model D Results :

$$10^{-4} * \begin{bmatrix} 2.227 & 0 & 0 & 0 \\ (0.216) & 0 & 0 & 0 \\ 0 & 2.487 & 0 & 0 \\ 0 & (0.250) & 0 & 0 \\ 0 & 0 & 1.900 & 0 \\ 0 & 0 & (0.185) & 0 \\ 0 & 0 & 0 & 2.483 \\ 0 & 0 & 0 & (0.235) \end{bmatrix}$$

For Model E Results :

$$10^{-5} * \begin{bmatrix} 1.64 & 0 & 0 & 0 \\ (0.24) & 0 & 0 & 0 \\ 0 & 1.95 & 0 & 0 \\ 0 & (0.28) & 0 & 0 \\ 0 & 0 & 1.32 & 0 \\ 0 & 0 & (0.19) & 0 \\ 0 & 0 & 0 & 1.97 \\ 0 & 0 & 0 & (0.28) \end{bmatrix}$$

E.2. Variance–covariance matrices for the signals

Hypothetical example: Two signals each for diagnoses 1 and 2; one signal each for diagnoses 3 and 4

For Model A Results

$$10^{-4} * \begin{bmatrix} 6.763 & & & & & \\ (0.614) & & & & & \\ 0.340 & 6.763 & & & & \\ (0.046) & (0.614) & & & & \\ 0.336 & 0.336 & 17.775 & & & \\ (0.034) & (0.034) & (1.650) & & & \\ 0.336 & 0.336 & 0.333 & 17.775 & & \\ (0.034) & (0.034) & (0.062) & (1.650) & & \\ 1.523 & 1.523 & 1.508 & 1.508 & 8.702 & \\ (0.124) & (0.124) & (0.145) & (0.145) & (0.652) & \\ -1.020 & -1.020 & -0.1009 & -1.009 & -4.570 & 9.416 \\ (0.095) & (0.095) & (0.102) & (0.102) & (0.346) & (0.799) \end{bmatrix}$$

For Model B results

$$10^{-4} * \begin{bmatrix} 6.248 & & & & & \\ (2.099) & & & & & \\ 0.356 & 6.248 & & & & \\ (0.163) & (2.099) & & & & \\ 0.340 & 0.340 & 62.109 & & & \\ (0.206) & (0.206) & (19.879) & & & \\ 0.340 & 0.340 & 0.324 & 62.109 & & \\ (0.206) & (0.206) & (0.370) & (19.879) & & \\ -7.531 & -7.531 & -7.177 & -7.177 & 160.380 & \\ (1.882) & (1.882) & (4.105) & (4.105) & (24.277) & \\ -4.489 & -4.489 & -4.279 & -4.279 & 94.857 & 62.381 \\ (1.172) & (1.172) & (2.472) & (2.472) & (11.071) & (13.888) \end{bmatrix}$$

For Model D Results

$$10^{-4} * \begin{bmatrix} 6.307 & & & & & \\ (0.636) & & & & & \\ 0.329 & 6.306 & & & & \\ (0.051) & (0.636) & & & & \\ 0.367 & 0.367 & 20.763 & & & \\ (0.039) & (0.039) & (2.125) & & & \\ 0.367 & 0.367 & 0.408 & 20.763 & & \\ (0.039) & (0.039) & (0.076) & (2.125) & & \\ 1.479 & 1.479 & 1.646 & 1.646 & 8.090 & \\ (0.141) & (0.141) & (0.175) & (0.175) & (0.729) & \\ -0.956 & -0.956 & -1.065 & -1.065 & -4.292 & 8.692 \\ (0.097) & (0.097) & (0.109) & (0.109) & (0.381) & (0.821) \end{bmatrix}$$

For Model E Results

$$10^{-4} * \begin{bmatrix} 2.582 & & & & & \\ (0.369) & & & & & \\ 0.489 & 2.582 & & & & \\ (0.072) & (0.369) & & & & \\ 0.163 & 0.163 & 1.842 & & & \\ (0.023) & (0.023) & (2.63) & & & \\ 0.163 & 0.163 & 0.054 & 62.109 & & \\ (0.023) & (0.023) & (0.010) & (19.879) & & \\ 0.540 & 0.540 & 0.054 & -7.177 & 160.380 & \\ (0.075) & (0.075) & (0.010) & (4.105) & (24.277) & \\ -0.907 & -0.906 & -0.302 & -4279 & 94.857 & 62.381 \\ (0.130) & (0.129) & (0.043) & (2.472) & (11.071) & (13.888) \end{bmatrix}$$

References

Ackerberg, D., 2002. Advertising, learning, and consumer choice in experience good markets: a structural examination, International Economic Review (forthcoming).

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

Astra-Merck Inc., 1996. 1996 Annual Report.

Azoulay, P., 2002. Do pharmaceutical sales respond to scientific evidence? Journal of Economics and Management Strategy 11, 551–594.

Azoulay, P., Berndt, E., Pindyck, R., 2003. Network externalities and diffusion in pharmaceutical markets: anti-ulcer drugs. Journal of Industrial Economics 51, 243–270.

Bass, F., Mahajan, V., Muller, E., 1990. New product diffusion models in marketing. Journal of Marketing 54, 1–26.

Berndt, E., Bui, E., Reiley, D., Urban, G., 1997. The roles of marketing, product quality and price competition in the growth and composition of the US anti-ulcer drug industry. In: Bresnahan, T., Gordon, R. (Eds.), The Economics of New Goods. University of Chicago Press, Chicago.

Bond, R.S., Lean, D., 1977. Sales promotion and product differentiation in two prescription drug markets. Discussion Paper, Federal Trade Commission.

Chandrashekaran, M., Sinha, R., 1995. Isolating the determinants of innovativeness: a split-population tobit (SPOT) duration model of timing and volume of first and repeat purchase. Journal of Marketing Research 32, 443–456.

Ching, A., 2000. Dynamic equilibrium in the US prescription drug market after patent expiration. Ohio State University, Mimeo.

Coscelli, A., 2000. The importance of doctors' and patients' preferences in the prescription decision. Journal of Industrial Economics 48, 349–369.

Crawford, G., Shum, M., 2000. Uncertainty and Learning in Pharmaceutical Demand. Duke University, Mimeo.

Ellison, S., Cockburn, I., Griliches, Z., Hausman, J., 1997. Characteristics of demand for pharmaceutical products: an examination of four cephalosporins. RAND Journal of Economics 28, 426–446.

Erdem, T., Keane, M., 1996. Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets. Marketing Science 15, 1–20.

Ferreyra, M., 1999. Physicians' learning about new drugs: an empirical approach. Carnegie-Mellon University, Mimeo.

Gourieroux, C., Monfort, A., 1996. Simulation-Based Econometric Methods. Oxford University Press, Oxford.

King, C., 2000. Marketing, product differentiation, and competition in the market for anti-ulcer drugs. Harvard University, Mimeo.

Medical Economics Co., 1997. (Ed.), Physicians' Desk Reference, Print Edition. Oradell, NJ.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027–1057.

Pertile, M., 1995. Zantac: Il Successo é un Qualcosa di Piu (Zantac: its phenomenal success), M.Sc. Thesis, CUOA—Consorzio Universitario per gli studi di organizzazione aziendale (in Italian).

Shapiro, C., 1982. Advertising as a barrier to entry? Federal Trade Commission Bureau of Economics, Working Paper #74.

Stern, S., 1996. The Demand for Pharmaceuticals. Kellogg School, Mimeo.

Stern, S., Trajtenberg, M., 1998. Empirical implications of physician authority in pharmaceutical decisionmaking, NBER Working Paper, #6851.

Urban, G.L., Carter, T., Gaskin, S., Mucha, Z., 1986. Market share rewards to pioneering brands: an empirical analysis and strategic implications. Management Science 32, 645–659.