Estimating Dynamic Games of Complete Information with an Application to the Generic Pharmaceutical Industry *

A. Ronald Gallant  
Fuqua School of Business  
Duke University

Han Hong  
Department of Economics  
Stanford University

Ahmed Khwaja  
Fuqua School of Business  
Duke University

Preliminary Draft  
Comments Welcome

Abstract

We estimate a dynamic oligopolistic entry model for the generic pharmaceutical industry that allows for dynamic spillovers from experience due to entry on future costs. Ex ante all firms are identical and heterogeneity arises endogenously based on past decisions of firms. Our paper contributes to both the estimation of dynamic games and the understanding of entry decisions in the pharmaceutical industry. Our dynamic model features serially correlated unobserved firm specific production costs. This introduces difficulty in the estimation of the dynamic game which we overcome using sequential importance sampling methods. We find that there are significant dynamic spillovers of entry on costs; each entry on average reduces costs by 7% at the next market opportunity, and the average annual cumulative reduction is 51%. The dynamic evolution of production cost plays an important role in the equilibrium path of the structure of the generic pharmaceutical industry. Our method is more generally applicable to estimating dynamic games when the choice set is discrete and there is serially correlated unobserved heterogeneity among agents.

Keywords: Dynamic Games, Dynamic Spillovers, Generic Pharmaceuticals, Sequential Importance Sampling.

JEL Classification: E00, G12, C51, C52

*Supported by the National Science Foundation.
1 Introduction

Entry behavior by pharmaceutical firms in the generic drug industry is an important topic of investigation in the empirical literature. For example, Scott-Morton (1999) estimates a static game of complete information to show that entry can be predicted by a firm’s organizational experience, size of the market and whether the entry opportunity is similar the firm’s existing portfolio of drugs. Ching (2008) uses a dynamic random utility model of demand for prescription drugs and finds that after patent expiration the rise in prices of brand name drugs can be explained by heterogeneity in consumer preferences, and the slow diffusion of generics in to the market by consumer learning about product quality. In a companion paper Ching (2004) develops a finite horizon dynamic oligopoly model to examine the effect of reducing the approval times for generics. He finds that although the policy change reduces the time to entry it also reduces the number of generic entrants in equilibrium as there tends to be greater entry in the early periods decreasing the average profitability of firms. Crawford and Shum (2005) employ a dynamic model to examine the effects of uncertainty and learning about the quality of match between a drug and the patient. They provide evidence of significant heterogeneity in drug efficacy across individuals and that there are benefits from learning through personal experience in dealing with uncertainty. However, much remains to be understood regarding the entry decisions of pharmaceutical firms especially whether this confers any strategic advantage when firms engage in a process of dynamic oligopolist competition.

Entry decisions of myopic pharmaceutical firms in a static competitive environment are drastically different from those in a forward looking dynamic competitive environment. In a dynamic setting, current entry can have a potential spillover effect on future entry, e.g. on the supply side through economies of scope and on the demand side through developing distributional channels and networks, and reputation about quality of the product. A firm might enter even if the current opportunity is loss generating as long as entry improves payoffs in the future.

In order to evaluate the effects of past experience on future cost and entry, and to evaluate
the effects of costs on entry, we formulate and estimate a structural model of a dynamic game
of oligopolistic competition that allows for spillovers from experience due to entry on future
costs leading to unobserved heterogeneity between firms. Ex ante all firms are identical and
heterogeneity arises endogenously based on decisions of firms. The model allows for serially
correlated unobserved cost. Although, there have been substantial recent developments in
the empirical literature on estimation of dynamic games, incorporating serially correlated
unobserved state variables remains prohibitively difficult. In our model production costs are
a serially correlated unobserved state variable which leads to severe computational obstacles
in estimating the model. We overcome this difficulty using sequential importance sampling
techniques. The sequential importance sampling method offers a drastic improvement in
the speed of computing the likelihood which makes the estimation of this dynamic model
feasible. In a seminal paper Keane (1994) used sequential importance sampling to develop
a computationally feasible simulation based estimator for panel data models when there are
potentially serially correlated errors. A special case of that estimator is the well-known GHK
(Geweke-Hajivassiliou-Keane) estimator that arises for a particular choice of the importance
sampling densities.

Our paper contributes to both the estimation of oligopolistic dynamic games and the
understanding of entry decisions in the pharmaceutical industry. We find that there are
significant dynamic spillovers of entry on costs; each entry on average reduces costs by 7% at
the next market opportunity, and the average annual cumulative reduction is 51%. Thus the
dynamic evolution of the production cost plays an important role in the equilibrium path
of the structure of the generic pharmaceutical industry. Our methods are more generally
applicable to estimating dynamic games in which (i) the choice set is discrete in nature, e.g.,
entry and exit from industry, expansion or reduction of product categories, introduction of
new or discontinuation of old brands, technology adoption or upgrades, relocation, start
up or shut down decisions of stores, firms, or factories etc, and (ii) when there is serially
correlated unobserved heterogeneity among agents.

The rest of the paper is organized as follows. We begin by discussing the related literature
in Section 2. Section 4 presents the model. The method used to solve the model is described
in Section 5 and the estimation procedure is described in Section 6. The results are discussed
in Section 7 and Section 8 concludes.

2 Related Literature

Static games under the incomplete information assumption have been studied by Bjorn and Vuong (1984), Bresnahan and Reiss (1991a), Bresnahan and Reiss (1991c), Haile, Hortacsu, and Kosenok (2003), Aradillas-Lopez (2005), Ho (2005), Ishii (2005), Pakes, Porter, Ho, and Ishii (2005), Augereau, Greenstein, and Rysman (2005), Seim (2005), Sweeting (2005), Tamer (2003), Manuszak and Cohen (2004), Rysman (2004), Gowrisankaran and Stavins (2004) and Bajari, Hong, Krainer, and Nekipelov (2006). Dynamic versions have been studied by Aguirregabiria and Mira (2002), Bajari, Benkard, and Levin (2004), Berry, Pakes, and Ostrovsky (2003), Pesendorfer and Schmidt-Dengler (2003) and Bajari, Chernozhukov, Hong, and Nekipelov (2007). These papers all make the strong assumption that there is no market or firm level unobserved heterogeneity other than a random shock that is independent and identically distributed across both time and players. This assumption is restrictive because it rules out unobserved dynamics in the latent state variables. This in turn rules out any private information that a player might have about competing firms that the researcher does not have.

On the other hand, Bresnahan and Reiss (1991b), Berry (1992), Tamer (2003), Ciliberto and Tamer (2003) and Bajari, Hong, and Ryan (2004) investigated static games of complete information. The complete information assumption allows substantial unobserved heterogeneity at the level of the firms. These games typically require the use of a combinatorial algorithm to search for an equilibrium instead of the continuous fixed point mapping used in incomplete information models to compute equilibria. To our knowledge, we are the first to undertake the challenge of applying the complete information model in a dynamic setting.

In the single agent dynamic framework, there is a considerable research that allows for unobserved heterogeneity that is time invariant, e.g., Keane and Wolpin (1997). However there is very little work that allows for serially correlated unobserved heterogeneity. Building on the work of Heckman (1981), a frequentist simulation based approach was developed by Khwaja (2001) to integrate out unobserved state variables in the context of a finite horizon dynamic discrete choice model. However this approach is only feasible when the state space
is of tractable size, e.g., discrete. Recent work by Imai, Jain, and Ching (2005) and Norets (2006) develops Bayesian methods for single agent dynamic discrete choice models with unobserved state variables that are serially correlated over time.

Our approach is likelihood based and we use MCMC methods in estimating the model. Thus, in principle, we could use either frequentist or Bayesian methods in the analysis because an MCMC chain can be used to compute the statistics that relate to either approach as shown by Chernozhukov and Hong (2003). However, our likelihood is nonlinear and is not differentiable making it extremely difficult to obtain an asymptotic theory that would justify frequentist inference. Conversely, Bayesian inference is theoretically justified under these conditions. Moreover, we use prior information to achieve identification which makes Bayesian methods doubly attractive. Therefore we apply Bayesian methods in the application in this paper.

Our implementation makes use of a sequential importance sampler. In recent work sequential importance sampling methods are have been used by Fernandez-Villaverde and Rubio-Ramirez (2005) for estimating macroeconomics dynamic stochastic general equilibrium models. The structure of dynamic stochastic general equilibrium models are very similar to dynamic discrete choice models. However, the game theoretical component of our implementation is new. More recently, Blevins (2008) has used the bootstrap filter to allow for serially correlated unobservable state variables in estimating dynamic single agent models, and dynamic games of incomplete information in a revealed preference framework.

In a continuous time setting, Nekipelov (2007) has developed a flexible indirect inference estimator for continuous time dynamic games in the context of eBay auctions without requiring the complete solution of the dynamic game. This is a novel approach that has potential applications in dynamic oligopolistic competition models. Relatedly, Benkard, Weintraub, and Roy (2007) have introduced the notion of oblivious equilibrium to facilitate the computation of dynamic game equilibria when the number of players is very large.

3 Institutional Background and Data

Our data come from Scott-Morton (1999) who in an pioneering study, based on the work of Berry (1992), analyzed the entry decisions of generic pharmaceutical manufacturers using a
static game of complete information.¹ Scott-Morton (1999) examined generic drug entries
in the period 1984 to 1994. This time period is particularly interesting because of the
1984 Waxman-Hatch Act which permitted Abbreviated New Drug Applications (ANDAs)
We summarize the relevant facts for our study here.

The preparation of an ANDA application takes months to years because it requires con-
struction of manufacturing facilities that need to be inspected and approved by the FDA.
The sunk cost of an ANDA market entry is high even though it is much less than a new
drug invention. For example, the revenues in one-firm markets average $10 million, while
the costs of submitting an ANDA can range from $250,000 to $20 million (Scott-Morton
(1999)). The size and heterogeneity of entry cost relative to the size of the market revenue
lead to a small number of entrants supported by each market. In addition, the FDA does
not reveal when and from whom it receives ANDA applications.

These features of the data are consistent with our modeling assumption of a dynamic
simultaneous entry game among a small number of competing pharmaceutical firms, in which
firms have to face substantial competition when they incur the sunk cost of entry.

As discussed in Scott-Morton (1999), announced entry is very rare, because firms do not
want to signal the common market value. They also fear that the delay in the approval will
invite competition. There are few late sequential movers who withdraw in response to rivals’
approvals. Simultaneous moves in a dynamic context are a more important feature of this
industry.

We describe the data briefly here but refer the reader to Scott-Morton (1999) for details.
The original data used by Scott-Morton (1999) consists of all ANDA approvals between 1984
and 1994. There is data on 1,233 ANDAs, and 363 markets entry opportunities for a total
of 123 firms. For each market opportunity there is information on:

1. Submission date, approval date, applicant name.


¹We are grateful to Fiona Scott Morton for providing us with her data, and to Derek Gurney for answering
our questions about the data.
3. Characteristics of drug markets: drug therapeutic class, patent expiration date, brand name drug, revenue of brand name drug the year before expiration, revenue from hospitals.

4. Characteristics of firms: stock of all drugs approved before 1984 for firms, parent or subsidiary firm, whether firm indicted in a bribery scandal (see below).

Based on our model specification and estimation strategy (described below) we only need information on total market revenues and entry decisions of potential entrants at each market entry opportunity to recover the model parameters. In our estimation we focus on the period after the FDA bribery scandal in 1989 because of the general upheaval and uncertainty in the generic drug industry surrounding this period. We take great care in processing the data between 1988 to 1993. We only look at ANDA applications for generic drugs that are orally ingested in the form of pills. In this category, for the sample period 1990-94, there are 40 market openings for which there is no missing revenue information and 51 firms who entered at least once. Each market category was defined as unique combination of primary ingredient, patent expiration date and total revenue for the branded drug for the last year before patent expiration.\(^2\) The top ten dominant firms in the sample after 1989 are (in descending order of dominance): Mylan, Novopharm, Lemmon, Geneva, Copley, Roxane, Purepac, Watson, Mutual and Lederle. The top firm, Mylan, entered 45% of the markets, the top two 48%, the top three 55%, the top four 60%, the top five 65%, and the top ten 73%. Individually, Novopharm entered 28%, and Lemmon and Geneva entered 25% of the markets. In our analysis we consider situations where the potential entrants are the top three or four firms. In each case the remaining firms are clubbed together in category referred to as “other.” The fraction of the market allocated to “other” is taken as a given that is anticipated by the top firms when considering entry. The procedure we use to implement this is described in Section 4. On average 3.3 firms enter a market (std. dev. is 2.6, min. is 1, and max. is 9). The mean revenue in thousands of dollars is 126,901 (std. dev. is 161,580, min. is 72, and max. is 614,593). In our estimation we use the log of revenue and in that case the mean is 10.47 (std. dev. is 2.1, min. is 4.3, and max. is 13.3).

\(^2\)Some amount of hand editing was required in constructing the sample, e.g., when the revenue number was different due to rounding error or there was a spelling error in the primary ingredient of the drug.
4 Model

In this section we formally describe the entry game. Although the empirical specification we adopt is motivated by our data the methods we develop to solve and estimate the game are more generally applicable to other dynamic oligopolistic competition models where the choice set is discrete. Firms maximize profits over an infinite horizon $t = 1, \ldots, \infty$, where each time the market is open counts as one time increment. A market opening is defined to be an entry opportunity that becomes available to generic manufacturers each time a branded product goes off patent. Since a time period uniquely identifies a market opening in what follows $t$ is used interchangeably to denote a market opening or the time period associated with it. The actions available to firm $i$ when market $t$ opens are to enter, $A_{it} = 1$, or not enter $A_{it} = 0$. Empirically this is determined by whether a firm submits an ANDA or not. There are $I$ firms in total so that the number who enter market $t$ is given by

$$N_t = \sum_{i=1}^{I} A_{it}$$  \hspace{1cm} (1)

The primary source of dynamics is through costs. The evolution of current costs, $C_{it}$, is determined by past entry decisions which accounts for spillovers of entry on current costs and random shocks. We follow the standard convention that a lower case quantity denotes the logarithm of an upper case quantity, e.g., $c_{it} = \log(C_{it})$. The log cost of a firm is assumed to follow a stationary autoregressive process of order one. The equation governing the log cost of firm $i$ at time $t$ is

$$c_{it} = \mu_c + \rho_c(c_{i,t-1} - \mu_c) - \kappa_c A_{i,t-1} + \sigma_c e_{it},$$  \hspace{1cm} (2)

where $e_{it}$ is normally distributed shock with mean zero and unit variance, $\sigma_c$ is a scale parameter, $\kappa_c$ is the entry spillover or immediate impact on cost at time $t$ if the market was entered at time $t - 1$, $\mu_c$ is a location parameter that represents the overall average of the log cost over a long period of time. The autoregressive parameter $\rho_c$ represents the degree of persistence between the current cost the its long run stationary level. We assume that all firms are ex ante identical, with all the heterogeneity arising endogenously based on the
effects of current decisions on future costs, hence none of these parameters are indexed by \( i \). Alternatively put, heterogeneity arises endogenously depending on the past actions of the firms.

We assume, as in Scott-Morton (1999), that all firms observe each other’s costs and hence this is a game of complete information. As far as the researcher is concerned the log cost can be decomposed into a sum of two components, known (or observable based on past actions), \( c_{k, i, t} \), and unobservable to the researcher, \( c_{u, i, t} \), as follows:

\[
\begin{align*}
\ c_{i,t} &= c_{u,i,t} + c_{k,i,t} \\
\ c_{u,i,t} &= \mu_c + \rho_c (c_{u,i,t-1} - \mu_c) + \sigma_c \epsilon_{it} \\
\ c_{k,i,t} &= \rho_c c_{k,i,t-1} - \kappa_c A_{i,t-1}
\end{align*}
\]  

From these equations it is seen that the location parameter \( \mu_c \) can be interpreted as the stationary long run mean of the unobservable portion of log cost and that the total impact or entry spillover at time \( t \) of a firm’s past entry decisions is \( c_{k, i, t} = -\sum_{j=0}^{\infty} \rho^j \kappa_c A_{i,t-j-1} \).

Two implications of the specification in equations (3)-(5) are that irrespective of the calendar time that has elapsed between any two adjacent market openings, (i) cost decreases are of the same magnitude, and (ii) the discount rate is held constant between market openings. These are plausible assumptions for our application as in our estimation sample (described earlier in Section 3) there are 40 openings in the period 1990-94 or on average a market opening every 1.5 months. This convention avoids insurmountable computational difficulties in solving for the equilibrium of the model that would arise if unequal spacing between market openings were assumed. Moreover, with this convention, we are only required to get the ordering of the data correct rather than to try to determine market entry dates precisely. We order markets according to the date when the first ANDA application was received by the FDA for a particular market opportunity.

Our timing convention underlying the dynamic cost process (equation 3), i.e., \( t \) represents sequence of market openings and not calendar time, implies that the cost advantage of entry dissipates with additional entry opportunities rather than the passage of calendar time. This may happen due to capacity or resource constraints. As the resources required for entry are stretched beyond their limits it may not be possible to expand the pool of resources that can
be devoted to additional projects easily. For example, a team that is working on formulating a particular drug or guiding it through the FDA approval process may only be able to work on a small number of projects at a given time and it may not be easy to hire additional members for the team. Furthermore, in view of the excellent fit to the data that we are able to achieve (Section 7 below) this convention appears reasonable a posteriori. Also of note is that although we are calling the latent variable cost for convenience, it is really any unobserved variable that could have dynamic spillover effects of entry on profits, e.g., on the supply side the mixture of experience could be learning by doing or economies of scope and on the demand side it could be reputation about quality or development of distribution networks.

The total revenue to be divided among firms who enter a market at time $t$ is $R_t = \exp(r_t)$, which is realized from the following independent and identical distribution,

$$r_t = \mu_r + \sigma_r \epsilon_{t+1,t},$$

(6)

where $\epsilon_{t+1,t}$ is normally distributed with mean zero and unit variance. In equation (6) $\mu_r$ is a location parameter that reflects the average total revenue over time for all the firms across all market opportunities, and $\sigma_r$ is a scale parameter. In our data the measure we have for total revenue is that for the last year the brand name drug was on patent. We interpret this value as being exogenously determined prior to the entry decisions of the generic firms, and being proportional to the total discounted value of the revenue flows to generic drugs after patent expiration.

A total of fifty one firms entered the market after the 1989 FDA bribery scandal. Computing a solution to a dynamic game of strategic interactions between fifty one players is not computationally feasible. So in our analysis we consider only the dominant firms. Therefore in the following $N_t$ is used to denote the number of entering dominant firms. We consider the case of three and four dominant firms. $N_t$ is less than or equal to $I$, which is the total number of dominant firms (i.e. 3 or 4), which is considered to be time-invariant. $N_t$ is to be differentiated from $N_t^a$, which is used to denote the total number of entrant firms at time $t$ including both dominant and nondominant firms.

We allow for nondominant firms as follows. Regressions indicate that $\log N^a = b \log R$,
with $b \approx 0.092$, is a reasonable approximation to the total number of firms that enter a market. The idea of this regression dates back to Bresnahan and Reiss (1991c) who showed that there is a close relationship between the number of entrants and the total market revenue. Therefore, when one of the dominant firms is considering entry, it can anticipate that the revenue available to be divided among all dominant firms should be larger than the average revenue available to each of the entering firms, which is \( \log R_{\text{anticipated}} = \log R - \log N^a = \log R - b \log R = \log \left( R^{1-b} \right) \). These considerations suggest that a reasonable functional form for dominant firm $i$’s per period profit at time $t$ is

\[
A_{it} \left( R_t^i / N_t - C_{it} \right),
\]

with $1 - b = 0.908$ being a reasonable lower bound for $\gamma$. The upper bound is one.

The firm’s total discounted profit at time $t$ is

\[
\sum_{j=0}^{\infty} \beta^j A_{i,t+j} \left( R_{t+j}^i / N_{t+j} - C_{i,t+j} \right),
\]

where $\beta$ is the discount factor, $0 < \beta < 1$. The firm’s objective is to maximize the present discounted value of its profit at each time period $t$ taking as given the equilibrium action profiles of other firms.

The Bellman equation for the choice specific value function, $V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t)$, for firm $i$’s dynamic problem at time $t$ is given by

\[
V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t)
= A_{it} \left( R_t^i / N_t - C_{it} \right)
+ \beta \mathcal{E} \left[ V_i(A_{i,t+1}^E, A_{-i,t+1}^E, C_{i,t+1}, C_{-i,t+1}, R_{t+1}) \mid A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t \right],
\]

where by convention $-i$ represents the other players. The choice specific value function represents the sum of current and future payoffs to firm $i$ from a choice $A_{i,t}$ at time $t$ explicitly conditioning on the choices that would be made by other firms $A_{-i,t}$ at time $t$ and with the expectation that firm $i$ and the other firms would be making equilibrium choices from period $t + 1$ onwards conditional on their current choices. The expectations operator here is over the distribution of the state variables in time period $t + 1$ conditional on the
realization of the time $t$ state variables. Therefore $V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t)$ is the payoff of firm $i$ at stage $t$ of the game.

A stationary pure strategy Markov perfect equilibrium of the dynamic game is defined by a best response strategy profile $(A_{i,t}^E, A_{-i,t}^E)$ that satisfies

$$V_i(A_{i,t}^E, A_{-i,t}^E, C_{i,t}, C_{-i,t}, R_t) \geq V_i(A_{i,t}, A_{-i,t}^E, C_{i,t}, C_{-i,t}, R_t) \quad \forall \ i, t,$$

(10)

where $A_{i,t}^E$ is the entry decision of firm $i$ for market $t$, $A_{-i,t}^E$ the vector of entry decisions of the other dominant firms, and $N_t^E$ is the number of firms that enter, which can be computed using equation (1), i.e., $N_t^E = \sum_{i=1}^{I} A_{i,t}^E$.

This is a game of complete information. Hence, if the state, which is the current cost of all firms $(C_{i,t}, C_{-i,t})$ and total revenue $(R_t)$, is known, then the equilibrium is known. Therefore, an ex ante value function can be computed from the choice specific value function

$$V_i(C_{i,t}, C_{-i,t}, R_t) = V_i(A_{i,t}^E, A_{-i,t}^E, C_{i,t}, C_{-i,t}, R_t).$$

(11)

The ex ante value function satisfies the Bellman equation

$$V_i(C_{i,t}, C_{-i,t}, R_t)$$

$$= A_{i,t}^E \left( R_t^i / N_t^E - C_{i,t} \right) + \beta \mathcal{E} \left[ V_i(C_{i,t+1}, C_{-i,t+1}, R_{t+1}) \mid A_{i,t}^E, A_{-i,t}^E, C_{i,t}, C_{-i,t}, R_t \right].$$

(12)

Equation (12) is different from the Bellman equation associated with the choice specific value function (equation 9) as it represents the sum of current and future payoffs to firm $i$ from an optimal choice $A_{i,t}^E$ at time $t$ explicitly conditioning on the equilibrium choices that would be made by other firms $A_{-i,t}^E$ at time $t$, and with the expectation that all firms would be making equilibrium choices from period $t + 1$ onwards conditional on their current choices. As in equation (9), the expectations operator here is over the conditional distribution of the state variables in time period $t + 1$.

Reny (1999) demonstrates the complexity of the conditions required to guarantee existence of pure strategy equilibria in games in which payoffs are discontinuous in strategies, as in our case. Dutta and Sundaram (1998) provide a comprehensive discussion of several results for existence of equilibria in Markovian games. The simplest possible case is when the state space can only take on a finite set of values, in which case, Theorem 3.1 of Dutta
and Sundaram (1998) implies that this game has a stationary Markov perfect equilibrium in mixed strategies. Parthasarathy (1973) showed that this can be relaxed to include a state space with countable values. The regularity conditions of Theorem 5.1 of Dutta and Sundaram (1998) come closer to the problem as we have posed it, notably the revenue and cost do not have to be discrete but they do need to be bounded. The equilibrium strategy profiles provided by Theorem 5.1 may depend on periods $t$ and $t - 1$ of the state vector.

We could modify our problem to meet the requirements of Theorem 3.1 that the state space be finite and countable. However we rely on Theorem 5.1 instead as we do not have trouble computing pure strategy equilibria for the problem as posed with an infinite state space. Theorem 3.1 is of interest to us because its proof relies on a dynamic programming approach that motivates our computational strategy, discussed below in Section 6 (see also Rust (2006) for a discussion of a similar computation strategy). We find that we can always compute pure strategy equilibria that depend only on period $t$ of the state vector, and hence trivially satisfy the regularity conditions of Theorem 5.1. Although, there are results that imply that a slightly modified version of the game proposed by us has equilibria, we rely mostly on the fact that we have no difficulty computing equilibria. In fact the key hurdle we face is not the lack of existence of equilibria but instead multiplicity of equilibria. In Section 6 we discuss how we resolve this problem.

5 Solving the Model

Our estimation strategy is based on a nested approach wherein the solution of the dynamic game is computed for each evaluation of a likelihood function that depends on both observable and latent variables.\(^3\) To compute a likelihood that depends only on observable variables, the latent state variables are integrated out using sequential importance sampling. Using the likelihood that depends only on observable variables, an MCMC algorithm generates draws from the posterior distribution of the parameters. The broad outline of the computational strategy is as follows: (1) Generate a parameter value by means of an MCMC algorithm. (2)

\(^3\)There are two likelihood functions: one that depends on both observable and latent variables and another that depends only on observable variables. In this paragraph we are specific. Usually we let meaning be determined by context.
For that parameter value, generate values for the latent variable over the sample period by means of the importance sampler. (3) Solve the dynamic game as function of the observed and unobserved state variables and the parameter value. (4) Use the equilibrium outcome generated from the model solution to compute a likelihood that depends on the observed and latent state variables (and the parameter value). (5) Integrate out the latent state variables by averaging the log likelihood over repetitions of the importance sampler to obtain a log likelihood that depends only observed variables (and the parameter value).\(^4\) (6) Use the likelihood that depends only on observed variables to make the accept/reject decision of the MCMC algorithm. Cycling through steps (1) through (6) generates an MCMC chain that is a sample from the posterior distribution of the parameters from which the posterior mean, mode, standard deviation, etc. can be computed.

In this section we describe the method used to solve for the equilibrium of the dynamic game given the observed and latent state variables and a set of parameter values. In Section 6 we describe how the likelihood is computed using the solution of the dynamic game and the MCMC algorithm. Since we use an infinite horizon model we looking for a stationary Markov perfect equilibrium which entails finding the fixed point of the Bellman equation (12).

Let the entry decisions of all \(i = 1, \ldots, I\) firms at time \(t\), i.e., the strategy profile of the dynamic game, be denoted by

\[
A_t = (A_{1t}, \ldots, A_{It}) .
\]  

(13)

As discussed in Section 4, the strategy profile \(A_t\) at time \(t\) of the dynamic game is a function of the current period state variables \((C_{1t}, \ldots, C_{It})\) and \(R_t\). The vector of the log of the state variables at time \(t\) is

\[
s_t = (c_{1t}, \ldots, c_{It}, r_t) .
\]  

(14)

We describe the solution algorithm for a given parameter vector \(\theta\) and a given state \(s_t\) at time \(t\).

We begin by defining a grid on the state space which determines a set of \((I + 1)\)-dimensional hyper-cubes. For each hyper-cube we use its centroid as its index or key \(K\).

\(^4\)As the name sequential importance sampler suggests, one does this averaging sequentially as one progress through the sample rather than storing all latent variable trajectories and then averaging. One can think of it as changing the order of a double sum.
A state $s_t$ within a rectangle can be mapped to its key $K$.\(^5\) Let the vector $V_K(s_t)$ have as its elements the ex ante value functions $V_{i,K}(s_t)$, i.e., $V_K(s_t) = (V_{1,K}(s_t), \ldots, V_{I,K}(s_t))$ (see equations 11 and 12). To each $K$ associate a vector $b_K$ of length $I$ and a matrix $B_K$ of dimension $I$ by $I + 1$. A given state point $s_t$ is mapped to its key $K$ and the value function at state $s_t$ is the affine function $V_K(s_t) = b_K + (B_K)s_t$.\(^6\) A value function $V_K(s_t)$ whose elements satisfy equation (12) is denoted $V_{K}^{*}(s_t)$.

The game as solved as follows:

1. Given a state point $s$, get the key $K$ that corresponds to it. (We suppress the subscript $t$ for notational convenience.)\(^7\)

2. Check whether the fixed point $V_{K}^{*}(s)$ of the Bellman equations (12) at this key has already been computed. If not, then use the following steps to compute it.

3. Start with an initial guess of the ex ante value function $V_{K}^{(0)}(s)$. An initial guess of the value function amounts to guessing the coefficients $(b_K^{(0)}, B_K^{(0)})$.

4. Obtain a set of points $s_j, j = 1, \ldots, J$, that are centered around $K$. The objective now is to obtain the ex ante value functions associated with these points to use in a regression to recompute (or update) the the coefficients $(b_K^{(0)}, B_K^{(0)})$.

5. Ex ante value functions are evaluated at best response strategies. In order to compute these we must, for each $s_j$, compute the choice specific value function (9) at as many strategy profiles $A$ as are required to determine whether or not (10) is satisfied. In this process we need to take expectations to compute the continuation value

$$\beta \mathbb{E}\big[ V_{K,i}^{(0)}(s_{t+1}) \big| A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_{t} \big]$$

that appears in (9), where we have used (11)

\(^5\)Grid increments are chosen to be (fractional) powers of two so points on the grid and the key have an exact machine representation. This facilitates compact storage of objects indexed by the key. The rounding rules of the machine resolve which key a state on a grid boundary gets mapped to, although lying on a boundary is a probability zero event in principle. The grid itself is never computed because all we require is the mapping $s \rightarrow K$, which is determined by the increments.

\(^6\)Keane and Wolpin (1997) adopt a similar approach for a single agent model. Our approach differs from Keane and Wolpin (1997) in that we let the coefficients of the regression depend on the state variables, specifically the key $K$, whereas Keane and Wolpin (1997) use an OLS regression whose coefficients are not state specific. Thus, our value function, unlike theirs, need not be continuous. Our value function can be thought of as an approximation by a step function.

\(^7\)In fact, because it is a stationary game, the subscript $t$ does not really matter
to express (9) in terms of \( V_K^{(0)}(s) \). To compute expectations over the conditional distribution of the random components of next period state variables, we use Gauss-Hermite quadrature. To do this, we obtain another set of points corresponding to each \( s_j \), i.e., \( s_{jl} \), \( l = 1, \ldots, L \). These points are the abscissae of the Gauss-Hermite quadrature rule which are located relative to \( s_j \) but shifted by the actions \( A \) under consideration to account for the dynamic effects of current actions on future costs (see equation 5). Expectations are computed using a weighted sum of the the value function evaluated at the abscissae (details below).

6. Having now the means to compute the continuation value at \( s_j \) for candidate strategies \( A_l \), we can compute the best response strategy profile \( A_j^E \) corresponding to \( s_j \) by checking the Nash condition (equation 10). Denote the choice specific value function evaluated at \((A_j^E, s_j)\) computed by means of \( V_K^{(0)}(s) \) using (9) as just described by 

\[
V_K^{(1)}(A_j^E, s_j) = (V_{i,K}^{(1)}(A_j^E, s_j), \ldots, V_{i,K}^{(1)}(A_j^E, s_j))
\]

7. Next we use the “data” \((V_K^{(1)}(A_j^E, s_j), s_j)\) to update the ex ante value function to \( V_K^{(1)}(s_j) \) by updating the coefficients of its affine representation to \((b_K^{(1)}, b_K^{(1)})\) via a multivariate regression on this “data” (as described in detail below).

8. We iterate (go back to step 5) over the ex ante value functions \( V_{i,K}^{(0)}(s), V_{i,K}^{(1)}(s), \ldots \) by finding a new equilibrium strategy profile \( A_j^E \) for each \( s_j \) until convergence is achieved for the coefficients \((b_K^{(0)}, b_K^{(0)}), (b_K^{(1)}, b_K^{(1)}), \ldots, (b_K^{(s)}, b_K^{(s)})\). This gives us \( V_K(s) = b_K + (b_K)s \) for every \( s \) that maps to key \( K \).

To summarize, the process of solving for the equilibrium begins with guessing a value for the value functions at a given state at iteration \( k = 0 \). These guesses are then used in computing the choice specific value functions at iteration \( k + 1 \) using equation (9). This computation involves taking expectations over the evolution of the future state variables, which is accomplished using Gaussian-Hermite quadrature. Once we have the choice specific value functions we compute the best response strategy profile at iteration \( k + 1 \) using equation (10). The best response strategy profile at iteration \( k + 1 \) is then used to compute the iteration

\[\footnote{\(V_K^{(1)}(A_j^E, s_j)\) will not equal \(V_K^{(1)}(s_j)\) because the former is “data” and the later is a regression prediction.}
The iteration \( k + 1 \) ex ante value functions are then used to compute the iteration \( k + 2 \) choice specific value functions using equation (9), and the entire procedure is repeated till a fixed point of equation (12) is obtained. In this iterative way the dynamic game is solved. We next provide additional details about the steps of the algorithm described above to solve the model.

To describe the Gauss-Hermite quadrature procedure used in Step 5, note that if one conditions upon \( s_t \) and \( A_t \), then the elements of \( s_{t+1} \) are independently normally distributed with means \( \mu_i = \mu_c + \rho_c(c_{it} - \mu_c) - \kappa_c A_{it} \) for the first \( I \) elements (see equation 2), mean \( \mu_{t+1} = \mu_R \) for the last element (see equation 6), standard deviations \( \sigma_s = \sigma_c \) for the first \( I \) elements, and standard deviation \( \sigma_{t+1} = \sigma_R \) for the last. Computing a conditional expectation of functions of the form \( f(s_{t+1}) \) given \( (A_t, s_t) \) such as appear in equations (9) and (12) is now a matter of integrating with respect to a normal distribution with these means and variances which can be computed by a Gauss-Hermite quadrature rule that has been subjected to location and scale transformation. The weights \( w_i \) and abscissae \( x_i \) for Gauss-Hermite quadrature may be obtained from tables such as Abramowitz and Stegun (1964) or by direct computation using algorithms such as Golub and Welsch (1969) as updated in Golub (1973). To integrate with respect to \( s_{i,t+1} \) conditional upon \( A_t \) and \( s_t \) the abscissae are transformed to \( \tilde{s}_{t+1,i} = \mu_i + \sqrt{2}\sigma_i x_i \), and the weights are transformed to \( \tilde{w}_i = w_i / \sqrt{\pi} \). Then, using a \( 2K + 1 \) rule,

\[
E[f(s_{t+1})|A_t, s_t] \approx \sum_{i_1=\cdots=-K}^{K} \cdots \sum_{i_t=\cdots=-K}^{K} \sum_{i_{t+1}=\cdots=-K}^{K} f(\tilde{s}_{t+1,i_1}, \tilde{s}_{t+1,i_2}, \tilde{s}_{t+1,i_3}) \tilde{w}_{i_1} \cdots \tilde{w}_{i_2} \tilde{w}_{i_3}. \tag{15}
\]

Note from above that the abscissae \( \tilde{s}_{t+1} \) depend on \( A_t \) and \( s_t \).\(^9\)

If, for example, there are three firms and a three point quadrature rule is used, then

\[
E[f(s_{t+1})|A_t, s_t] \approx \sum_{i=-1}^{1} \sum_{j=-1}^{1} \sum_{k=-1}^{1} f(\tilde{s}_i, \tilde{s}_j, \tilde{s}_k) \tilde{w}_i \tilde{w}_j \tilde{w}_k.
\]

We use three point rules throughout. A three point rule will integrate a polynomial in \( s_{t+1} \)

\(^9\)If the \( \tilde{s}_{t+1} \) cross a grid boundary when computing (9) in Step 5, we do not recompute \( K \) because this would create an impossible circularity due to the fact that the value function at the new \( K \) may not yet be available. Our grid increments are large relative to the scatter of the quadrature rule so that crossing a boundary will be a rare event, if it happens at all.
up to degree five exactly. In particular, equations (9) and (12) can be expressed in terms of $s_t$ using $C_{it} = \exp(s_{it})$ for $i = 1, \ldots, I$ and $R_t = \exp(s_{I+1,t})$.

Step 7 involves updating the ex ante value function using a regression. We next describe how we do this. We have a grid over the state space whose boundaries are (fractional) powers of two over the state space. We approximate the value function $V(s_t)$ by an indexed affine representation as described above. We set the grid increments that determine the index at 16 times the standard deviation of the state variables rounded to a nearby (fractional) power of two to scale the grid appropriately.\(^{10}\) We compute the coefficients $b_K$ and $B_K$ as follows. We initialize to zero. We generate a set of abscissae $\{s_j\}$ clustered about $K$ and solve the game with payoffs (9) to get corresponding equilibria $\{A^F_j\}$. We substitute the $(A^F_j, s_j)$ pairs into (9) to get $\{V(A^F_j, s_j)\}_{j=1}^J$. Using the pairs $\{(V(A^F_j, s_j), s_j)\}$ as data, we compute $b_K$ and $B_K$ by multivariate least squares. We repeat until the $b_K$ and $B_K$ stabilize. We have found that twenty iterations suffices for three firms and thirty for four firms.\(^{11}\) The easiest way to get a cluster of points $\{s_j\}$ about a key is to use abscissae from the quadrature rule described above with $s$ set to $K$ and $A$ set to zero. However, one must jitter the points so that no two firms have exactly the same cost (see next paragraph). Of more importance in reducing computational effort is never to recompute the payoff (9) when checking (10). Our strategy is to store payoff vectors in a binary tree indexed by $A$ and check for previously computed payoffs before computing new ones.

There will, at times, be multiple equilibria in solving the game. We therefore adopt an equilibrium selection rule as follows. Multiple equilibria usually take the form of a situation like the prisoner’s dilemma where one or another firm can profitably enter but if both enter they both will incur losses whereas if neither enters they would get the continuation value of the game. In the three firm game the frequency of multiple equilibria is 4%. We resolve this situation by assuming a coordination game. The firms that have the lowest costs $C_{it}$ are those that are allowed to enter. This idea is similar to that used by Berry (1992) and

\(^{10}\)The set of keys that actually get visited in any MCMC repetition is about the same for grid increments as small as 4 for our data. For a three firm game the number of rectangles that actually are visited in any one repetition is about six.

\(^{11}\)An alternative is to apply a modified Howard acceleration strategy as described in Kuhn (2006); see also Rust (2006) and Howard (1960). The idea is simple: The solution $\{A^F_j\}$ of the game with payoffs (9) will not change much, if at all, for small changes in the value function $V(s)$. Therefore, rather than recompute the solution at every step of the $(b_K, B_K)$ iterations, one can reuse a solution for a few steps.
Scott-Morton (1999). That is, the strategy profiles $A_t$ are ordered by increasing aggregate cost, $C = \sum_{i=1}^I A_{it} C_{it}$ and the first $A_t$ that satisfies (10) is accepted as the solution. Note that our distributional assumptions on $s_t$ guarantee that no two $C$ can be equal so that this ordering of the $A_t$ is unique. Moreover, none of the $C_{it}$ can equal one another and when that is true we have never failed to be able to compute a pure strategy equilibrium.

6 Computing the Likelihood and Estimation

In this section describe our estimation strategy. The parameters of the model are

$$\theta = (\mu_c, \rho_c, \sigma_c, k_c, \mu_r, \sigma_r, \gamma, \beta, p_a).$$

(16)

The meaning of the first eight parameters has been discussed in Section 4. The meaning of $p_a$ will be discussed immediately below. From the computational standpoint, the setting is as follows. There are $I$ firms, $i = 1, \ldots, I$, who can enter the market or not at each time period $t$. If firm $i$ enters at time $t$, then $A_{it} = 1$; if not, $A_{it} = 0$. The number of firms that enter at time $t$, is $N_t = \sum_{i=1}^I A_{it}$. The total anticipated revenue available to the firms in each market is $R_t^N$, which is divided equally among those firms that enter. We can observe both $R_t$ and $A_t = (A_{1t}, \ldots, A_{It})$. Log cost, $c_{i,t} = \log C_{i,t}$, is the sum of two components. The first is $\log C_{u,i,t}$, which is known by all firms but not by us. The second is $\log C_{k,i,t}$, which is known by all firms and by us as it depends only on the past actions which are observable (see equation (21) below). Both evolve as a Markov process. Cost together with revenue $R_t$ make up the state vector.

Denote the part of the state vector that is unobservable to us by

$$X_t = (C_{u,1,t}, \ldots, C_{u,l,t}).$$

(17)

Denote the variables that we can observe by

$$Y_t = (A_{1t}, \ldots, A_{It}, C_{k,1,t}, \ldots, C_{k,l,t}, R_t).$$

(18)

As previously, a lower case variable denotes the logarithm of an upper case variable with the exception that $a_t = A_t$; i.e. $x_{it} = \log X_{it}$, $c_{u,i,t} = \log C_{u,i,t}$, $c_{k,i,t} = \log C_{k,i,t}$, $r_{it} = \log R_{it}$. With
these conventions, \( x_t = (c_{u,1,t}, \ldots, c_{u,J,t}) \), and \( y_t = (a_{1t}, \ldots, a_{Jt}, c_{k,1,t}, \ldots, c_{k,J,t}, r_t) \). Recall that cost evolves as

\[
c_{i,t} = c_{u,i,t} + c_{k,i,t} \tag{19}
\]

\[
c_{u,i,t} = \mu_c + \rho_c (c_{u,i,t-1} - \mu_c) + \sigma_c e_{it} \tag{20}
\]

\[
c_{k,i,t} = \rho_c c_{k,i,t-1} - \kappa_c A_{i,t-1} \tag{21}
\]

and revenue evolves as

\[
r_t = \mu_r + \sigma_r e_{I+1,t} \tag{22}
\]

We have data for both the pre- and post-scarelder periods. The pre-scarelder period is indexed by \( t = -n_0, \ldots, 0 \) and the values of \( Y_t \) over the pre-scarelder period are denoted by \( Y_{pre} \). The post-scarelder period is indexed by \( t = 1, \ldots, n \) with values over it denoted by \( Y_{post} \).

While the scandal changed the market structure thus rendering the pre-scarelder data unsuitable for general estimation, it can still be used for two purposes: The entry decisions \( \{A_{it}\}_{t=-n_0}^0 \) can be used to compute the last two pre-scarelder values \( c_{k,i,-1} \) and \( c_{k,i,0} \) of the observable part of log cost for each firm; and the pre-scarelder log revenue \( \{r_t\}_{t=-n_0}^0 \) can be used to help identify the parameters \( \mu_r \) and \( \sigma_r \).

We compute the initial values \( c_{k,i,-1} \) and \( c_{k,i,0} \) for each firm by running the recursion on equation (21) started at zero over the observed choices \( \{A_{it}\}_{t=-n_0}^0 \). This gives us the vectors \( y_{-1} \) and \( y_0 \) because \( (R_{-1}, A_{-1}) \) and \( (R_0, A_0) \) are in \( Y_{pre} \).

While the scandal may have affected which firms participated in the market post-scarelder, there is no reason to believe that market opportunities were different pre- and post-scarelder. Therefore the pre-scarelder data can be used to help identify the revenue distribution. From \( Y_{pre} \) we can compute a normal likelihood for log revenue over the period \(-n_0, \ldots, 0\). Although this likelihood actually only depends on two elements \( (\mu_r, \sigma_r) \) of \( \theta \), we denote it as

\[
p(Y_{pre} | \theta) \tag{23}
\]

for convenience.

Since \( A_t \) is a deterministic function of \((x_t, r_t, y_{t-1}, \theta)\), a density for \( A_t \) computed according to the considerations discussed thus far would puts mass one on a single value of \( A_t \). The implication is that a likelihood over the post-scarelder data would be one if we predict every
entry decision perfectly and zero otherwise. This situation arises frequently in likelihood based inference; prominent examples are option pricing and yield curve estimation. The standard approach to this problem is to assume measurement error. Therefore, we adopt the following density for $A_t$

$$
p(A_t | r_t, x_t, y_{t-1}, \theta) = \prod_{i=1}^{I} (p_a)^{I(A_{it}=A^c_{it})} (1 - p_a)^{I(A_{it} \neq A^c_{it})} \tag{24}
$$

where $0 < p_a < 1$ and $A^c_{it}$ is computed from $(x_t, r_t, y_{t-1}, \theta)$ using the methods described in Section 5.

Doucet, de Freitas, and Gordon (2001) present a concise description of the sequential importance sampler that is adequate to follow our analysis. The densities relevant to a sequential importance sampler are the transition density of the hidden state vector

$$
p(x_t | x_{t-1}, \theta), \tag{25}
$$

which is defined by recursion equation (20), the initial density

$$
p(x_0 | \theta), \tag{26}
$$

which, from equation (20), is normal with mean $\mu_c$ and standard deviation $\sigma_c/\sqrt{1 - p_c^2}$, and the observation density

$$
p(y_t | y_{t-1}, x_t, \theta) = p(A_t | r_t, y_{t-1}, x_t, \theta) p(r_t | y_{t-1}, x_t, \theta), \tag{27}
$$

where, from equation (22), $p(r_t | y_{t-1}, x_t, \theta)$ is normal with mean $\mu_r$ and standard deviation $\sigma_r$.

The sequential importance sampler is as follows:

1. For $t = 0$

   (a) Start $N$ particles by drawing $x^{(j)}_0$ for $j = 1, \ldots, N$ from the initial density equation (26).

   (b) Compute

   $$
p(y_0 | \theta) = \int p(y_0 | y_{-1}, x_0, \theta) p(y_{-1}, x_0 | \theta) \, dx_0 
   \quad \triangleq \quad \frac{1}{N} \sum_{j=1}^{N} p(y_0 | y_{-1}, x^{(j)}_0, \theta). \tag{21}
$$
2. For \( t = 1, \ldots, n \)

(a) For each particle, draw \( \tilde{x}^{(j)}_{0:t} \) from the transition density equation (25) and set
\[
\tilde{x}^{(j)}_{0:t} = (x_{0:t-1}^{(j)}, \tilde{x}^{(j)}_t).
\]

(b) For each particle compute the particle weights \( \tilde{w}^{(j)}_t \) using the observation density equation (27); i.e.
\[
\tilde{w}^{(j)}_t = p(y_t | y_{t-1}, \tilde{x}^{(j)}_t, \theta).
\]

Here is where we run into trouble with a deterministic function because the weights could all be zero.

(c) Normalize the weights so that they sum to one
\[
\tilde{w}^{(j)}_t = \frac{\tilde{w}^{(j)}_t}{\sum_{j=1}^{N} \tilde{w}^{(j)}_t}.
\]

(d) For \( j = 1, \ldots, N \) sample with replacement the particles \( x^{(j)}_{0:t} \) from the set \( \{\tilde{x}^{(j)}_{0:t}\} \) according to the weights \( \{\tilde{w}^{(j)}_t\} \). (Note the convention: Particles with unequal weights are denoted by \( \{x^{(j)}_{0:t}\} \). After resampling the particles are denoted by \( \{x^{(j)}_{0:t}\} \).

(e) Compute
\[
p(y_t | y_{1:t-1}, \theta, p_a) = \int p(y_t | y_{t-1}, x_t, \theta) p(y_{t-1}, x_t | y_{1:t-1}, \theta) \, dx_t.
\]
\[
= \frac{1}{N} \sum_{j=1}^{N} p(y_t | y_{t-1}, x_t^{(j)}, \theta).
\]

Note that \( p(y_t | y_{t-1}, x_t^{(j)}, \theta) \) does not have to be recomputed here if the weights \( \tilde{w}^{(j)}_t \) are associated to \( x_t^{(j)} \) in the resampling step and saved. If each firm’s entry decisions are similarly associated, then classification error rates can be computed at this step.

3. Done

(a) The likelihood is
\[
\mathcal{L}(\theta) = p(y_{0:t} | \theta) = p(Y_{pre} | \theta, p_a)p(y_0 | \theta) \prod_{t=1}^{n} p(y_t | y_{0:t-1}, \theta).
\]
The log likelihood surface is plotted on a fine grid in Figure 1 and on a coarse grid in Figure 2 for the three firm model. Figures 3 and 4 are the same for the four firm model. The endpoints of the abcissae are tenth of a standard deviation to the left and right of the maximum in Figure 1 and 24 standard deviations to the right and left in Figure 2. For Figures 3 and 4 they are a tenth and 48. These are profile likelihoods; i.e., in each panel the indicated parameter is moved and all others are fixed at the values that maximize the likelihood.

As seen from Figures 3 and 3, the surface is, basically, a step function so that curvature at the maximum will not provide a reliable basis for inference. The reason, of course, is that small changes in the parameters cause the decisions of the firms to change. In this situation, accurate frequentist inference would be difficult and would be prohibitively computationally intensive if bootstrapping were involved. On the other hand, Bayesian inference in this situation is conceptually straightforward and computationally feasible.

However, Figures 2 and 4 do suggest that implementing a Bayesian strategy that explores the surface well will be a challenge. They also suggest that the standard deviations of the posterior will be extremely tight. The horizontal line is at three orders of magnitude below the maximum. If an MCMC chain is near the maximum, the chance that it will move to a point below the horizontal line line is less than 0.001.

An MCMC chain that uses a move-one-at-a-time random walk proposal density will usually do a good job of exploring surfaces such as seen in Figures 2 and 4; see Gamerman and Lopes (2006). One pays a price for safety because an MCMC chain that uses a move-one-at-a-time random walk proposal strategy is usually inefficient relative to those that use
other proposal strategies. Briefly, the method is as follows: The proposal density \( q(\theta^o, \theta^*) \)
defines a distribution of potential new values \( \theta^* \) given an old value \( \theta^o \). Denote the likelihood
by \( \mathcal{L}(\theta) \) and the prior by \( \pi(\theta) \). Given the value \( \theta^o \) at the end of the MCMC chain, one moves
the chain forward one step to \( \theta' \) as follows:

1. Draw \( \theta^* \) according to \( q(\theta^o, \theta^*) \).
2. Let \( \alpha = \min\left(1, \frac{\mathcal{L}(\theta^*) \pi(\theta^*) q(\theta^*, \theta^o)}{\mathcal{L}(\theta^o) \pi(\theta^o) q(\theta^o, \theta^*)}\right) \).
3. With probability \( \alpha \), set \( \theta' = \theta^* \), otherwise set \( \theta' = \theta^o \).

For our particular \( q \), one randomly chooses an element \( j \) of \( \theta^o \) to move and then proposes
a new value by replacing \( \theta^o_j \) with a draw from the normal with mean \( \theta^o_j \) and scale \( \sigma_j \) where
\( \sigma_j \) is chosen such that acceptance at Step 3 occurs with a frequency of about 30\% (see e.g.,
Gelman, Roberts, and Gilks (1996), Roberts and Rosenthal (2001)). The vertical lines in
Figures 2 and 4 indicate the range of the MCMC chain’s excursions after transients have
died out.

The likelihood is hierarchical in that given model parameters and conditional upon the
latent cost variable, it can be evaluated by solving the game. Given this structure, estimation
can be viewed as a double nesting of the conditional likelihood within an outer MCMC loop
and an inner importance sampling loop. The MCMC proposal density fixes \( \theta \) in the outer
loop. The sequential importance sampler generates a cost trajectory within the inner loop.
Solving the game both evaluates the conditional likelihood along this trajectory and provides
the importance sampler with the information needed to adjust costs sequentially along the
trajectory to take into account the effect of entry decisions on the trajectory. When one
falls through the inner loop, the likelihood has been averaged over costs thereby removing
the conditioning on costs. At this point the MCMC accept/reject decision is made and the
MCMC chain incremented. When one falls through the outer loop one has the complete
MCMC chain.

We implement our computational algorithm using code that is in the public domain
and available at http://econ.duke.edu/webfiles/arg/emm. This code is based on Cherno-
zhukov and Hong (2003). Full details regarding the proposal density and other conventions
are in the User’s Guide distributed with the code. One needs enough draws to accurately compute averages such as standard deviations, histograms, and other characteristics of the posterior distribution. Our chains are highly correlated so that very long chains with a stride (sampling rate) of 375 are required to break the dependence. This is the price we pay for safety. As explained in the User’s Guide, computations can be accelerated if the values of \( \theta \) visited by the chain are restricted to (fractional) powers of two. We impose this restriction on the chain.

The parameter \( p_a \) can either be estimated or be fixed at various values. We tried values from 0.75 to 0.95. We find that estimates of the other elements of \( \theta \) are hardly affected. What we do find is that varying \( p_a \) affects the rate at which particles die out at Step 2d in the sequential importance sampler. Since we are not using the sequential importance sampler as a smoother, the rate at which particles die out is of no concern. We always have a large number of points available at Step 2e of the sequential importance sampler. When \( p_a \) is treated as a parameter to be estimated, the performance of the MCMC algorithm is degraded somewhat. We think that fixing \( p_a \) is preferred because doing so improves performance and permits a cleaner comparison of results across the cases \( I = 3, 4 \) that we consider in Section 7.

The firm’s discount rate \( \beta \) is extremely difficult to estimate in studies of this sort (see e.g., Magnac and Thesmar (2002) and Rust (1994)) and we find this to be the case here. A common rule of thumb in business is not to undertake a project whose internal rate of return is less than 20%. Grabowski, Vernon, and DiMasi (2002) state that estimates of internal rates specific to the drug industry range “from 13.5% to over 20%.” As to theory, a firm should not undertake a project whose rate of return is less than its cost of capital. The historical risk premium in the drug industry is 12.55%, (e.g., Gebhardt, Lee, and Swaminathan (2001)). If one adds to this a nominal borrowing rate of 5% one arrives at the value 17.55%. Grabowski, Vernon, and DiMasi (2002) arrive at a nominal cost of capital of 14% using a CAPM method that they regard as biased downward. On the basis of these considerations we set the firm’s discount rate at 20%. There are 40 market entry opportunities in our five years of data. That implies an expected time increment of 0.125 years between prospective projects for the firms in our data. Therefore, using a internal rate of 20%, allowing for compounding, and
rounding to a nearby fractional power of two, we set $\beta = 0.96875$.

Examination of equation (9) indicates that were $\gamma$ to enter as a linear factor then $\gamma$ would not be identified. That in fact it enters to the first order as $(1 + \gamma \log R)$ does not help matters much. Attempts to estimate $\gamma$ anyway yields estimates that meander about 0.93. Therefore, based on the plausible lower bound of 0.908 derived in Section 4 and our experience from trying to estimate $\gamma$, we take 0.93 to be a reasonable value. Rounding to a nearby fractional power of two, we set $\gamma = 0.9375$.

For the remaining parameters we use flat, noninformative priors that impose these support conditions: $-1 \leq \rho_c \leq 1$, $0 \leq \kappa_c$, $0 < \sigma_c$, and $0 < \sigma_r$.\(^{12}\)

7 Results

We estimate the model for two cases: (1) the top three dominant firms are the only potential entrants that are strategic competitors (the actions of the remaining 48 firms are accounted for by the parameter $\gamma$), and (2) the top four dominant firms are only potential entrants (as before the actions of the other 47 entrants are accounted for by $\gamma$). The mode and standard deviations of the posterior distribution are reported in Table 2. We focus on the mode of the multivariate posterior distribution because it actually corresponds to a value at which the model has been evaluated. Other measures of central tendency of the posterior distribution can be misleading when studying the behavior of a structural model because they may have never appeared in the MCMC chain and could give a distorted view of the model were it to be evaluated at such a point. Histograms of the marginal posterior distributions are displayed in Figures 5 and 6 for the three and four firm cases, respectively.\(^{13}\)

Table 2 about here

Figure 5 about here

Figure 6 about here

\(^{12}\)Open ended ranges actually should have large upper and lower bounds to assure stationarity of the MCMC chain. As seen from our histograms, Figures 5 and 6, bounds do not interfere with the chain.

\(^{13}\)Compare Figures 2 and 4.
The parameters are tightly estimated\textsuperscript{14} and, as seen from the extremely low classification error rates, model predictions are quite accurate. The large value of $\rho_c$ implies costs are persistent. A value of $\kappa_c$ of 0.07 implies that an immediate cost reduction of 7\% going into the next market opening, and the average annual cumulative reduction computed using the AR(1) process is 51\%. However, $\sigma_c$ is large, so that this reduction is attenuated by noise.

Figure 7 plots the log cost of the three dominant firms in the three firm model in the upper three panels. The circles indicate that the firm entered that market. The logarithm of cost is computed by averaging at Step 2e of the importance sampler. The bottom panel shows log total revenue; the numbers at the bottom of this panel are the number of dominant firms who entered the market at that time point. The top firm, Mylan, has a clear cost advantage over its competitors. Broad trends in cost are about the same for all firms.

Figure 8 overplots the log cost of the three dominant firms in both models. The circles at the bottom of the upper panel indicate which markets Mylan entered, the crosses in the middle panel are the same for Novopharm, and the asterisks in the lower panel are the same for Lemmon. The construction of the plots is the same as for Figure 7. The salient feature of this plot is that costs for the three dominant firms are estimated as being about the same in the three and four firm models.

Figure 9 displays the entry decisions of the dominant firms, period by period, as circles and the model’s average prediction of their entry, period by period, as crosses. The average prediction is computed by averaging game solutions at Step 2e of the importance sampler at the mode of the posterior density. The classification error rates shown in Table 2 can be viewed as the errors that would obtain if decisions were predicted by thresholding the average predictions shown in Figure 9.

\textsuperscript{14}Despite the small standard deviations shown in Table 2, the profile likelihoods in Figures 2 and 4 suggest that the MCMC chain adequately explored the posterior density. The likelihood is proportional to the posterior because priors are flat.
Another way is to assess results is to directly explore the possibility that the firms play a different game than the game we propose rather than inferring the importance of the dynamics from the estimate of $\kappa_c$. Consider two other games that might be played instead of the game with payoffs (9). They could play a game with payoffs

$$V_i(A_{i,t}, A_{i^{-1}, i}, C_{i,t}, C_{i^{-1}, t}, R_t) = A_{it} (R_t^0 / N_t - C_{u,i,t}),$$

(28)

where no attention at all is paid to the cost reductions arising from past market entries ($\kappa_c = 0$) or to dynamic spillovers of entry ($\beta = 0$). We call this the myopic game ($\beta = 0, \kappa_c = 0$).

Or they could play a game with payoffs

$$V_i(A_{i,t}, A_{i^{-1}, i}, C_{i,t}, C_{i^{-1}, t}, R_t) = A_{it} (R_t^0 / N_t - C_{it})$$

(29)

where they take cognizance of the effect of entry on costs but ignore the continuation value of the game, i.e., $\beta = 0$. We call this the static game ($\beta = 0, \kappa_c > 0$).

For the three firm game the myopic game ($\beta = 0, \kappa_c = 0$) has an equilibrium that agrees with the solution of the game we propose (i.e., the game with payoffs (9)) in 49% of the cases. The game that ignores the continuation value ($\beta = 0, \kappa_c > 0$) has an equilibrium that agrees in 81% of the cases. For the four firm game, these values are 31% and 68%, respectively.

These values were computed by using the posterior modes shown for the game in Table 2 and finding all equilibria for the three games for all costs that obtained at Step 2b of the sequential importance sampler. Incidentally, we can also compute the incidence of multiple equilibria for these three games. For the three firm game they are 5% ($\beta = 0, \kappa_c = 0$), 5% ($\beta = 0, \kappa_c > 0$), and 4% ($\beta > 0, \kappa_c > 0$), respectively. For the four firm game they are 5%, 7%, and 4%, respectively. As discussed earlier, these cases of multiple equilibria are like prisoners dilemma games and we assume a coordination game whereby the firms with the lowest costs are those that are allowed to enter.

These computations suggest that the myopic and static games would do a poor job of rationalizing the data. To check, we use our parameter estimates, impose $\beta = \kappa_c = 0$, and find that the overall classification error rate for the myopic game exceeds the overall value in Table 2 by a factor of 3.8 for the three player game and 3.6 for the four player game. Similarly, imposing $\beta = 0$, we find that the classification error rate for the static game exceeds the values in Table 2 by a factor of 2.0 for both the three and four player games.
It is worth asking the question whether what is recovered is the dynamic spillover effect of entry, i.e., entry reduces costs or whether the causality is reversed and it is low cost firms that enter. In the latter case one could think of a situation where there is persistent heterogeneity in costs across firms and the low cost firms always enter and the high cost firms stay out. Recall that in our model all firms are the same ex ante. Heterogeneity in costs arises endogenously based in part on past actions. Therefore Figures 7, 8, and 9 are to be viewed as ex-post reconstructions of the history of the game. However, one might surmise from the volatility of the plots that too much emphasis is being placed on the trajectory of equation (4) and not enough on (5). Stated differently, one might surmise that the effects of the random shocks (operating through $\sigma_c$) is too large and of the dynamic spillovers (through $\kappa_c$) is too small or that more generally $\sigma_c$ and $\kappa_c$ are correlated. One way to check this is to set $\sigma_c$ to smaller values and re-run the MCMC chain. Setting $\sigma_c$ to 0.25, 0.125, and 0.0625 has very little effect on $\kappa_c$ although it does dramatically reduce the likelihood evaluated at the mode. Thus we conclude that we are estimating the effect of entry on costs and not vice versa.

8 Conclusions

We estimate a dynamic oligopolistic entry model for the generic pharmaceutical industry. Our stylized model fits the data well, i.e., the classification error rates are small. We find that costs affect entry decisions and that past entry decisions (e.g., experience) affect future costs. Hence we find a dynamic spillover effect of entry in reducing future costs.

Our paper contributes to both the estimation of the oligopolistic dynamic games and the understanding of entry decisions in the pharmaceutical industry. The dynamic model features unobserved firm-level production costs that are serially correlated over time. This introduces difficulty in the estimation of the dynamic game theoretic model which we overcome using sequential importance sampling methods. The empirical findings show that the dynamic evolution of the production cost plays an important role in the equilibrium path of the generic pharmaceutical industry structure. Our method is more generally applicable to estimating dynamic games in which the choice set is discrete and when there is serially correlated unobserved heterogeneity among agents. In future work we hope to extend our
methods to allow for estimation of dynamic games where the strategy set is mixed discrete-continuous, e.g., introduction of a new brand and associated decisions about advertising expenditure. We also plan to examine whether our method can be employed in estimating dynamic games of incomplete information.

References


Table 1. Data

<table>
<thead>
<tr>
<th>Drug</th>
<th>ANDA Date</th>
<th>Mylan</th>
<th>Novopharm</th>
<th>Leommon</th>
<th>Geneva</th>
<th>Total Entrants</th>
<th>Revenue ($'000s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SULINDAC</td>
<td>03 Apr. 90</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>189010</td>
</tr>
<tr>
<td>ERYTHROMYCIN STE</td>
<td>15 May 90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>13937</td>
</tr>
<tr>
<td>ATENOLOL</td>
<td>31 May 90</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>69802</td>
</tr>
<tr>
<td>NIFEDIPOINE</td>
<td>04 Jul. 90</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>302983</td>
</tr>
<tr>
<td>MINOCYLCE HYDR</td>
<td>14 Aug. 90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>55491</td>
</tr>
<tr>
<td>METHOTREXATE SOD</td>
<td>15 Oct. 90</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>24848</td>
</tr>
<tr>
<td>PYRIDOSTIGMINE B</td>
<td>27 Nov. 90</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2113</td>
</tr>
<tr>
<td>ESTROPIRATE</td>
<td>27 Feb. 91</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6820</td>
</tr>
<tr>
<td>LOPERAMIDE HCL</td>
<td>30 Aug. 91</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>31713</td>
</tr>
<tr>
<td>PHENDIMETRAZINE</td>
<td>30 Oct. 91</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1269</td>
</tr>
<tr>
<td>TOLMETIN SODIUM</td>
<td>27 Nov. 91</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>59108</td>
</tr>
<tr>
<td>CLEMASTINE FUMAR</td>
<td>31 Jan. 92</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9077</td>
</tr>
<tr>
<td>CINOXACIN</td>
<td>28 Feb. 92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6281</td>
</tr>
<tr>
<td>DILTIAZEM HCL</td>
<td>30 Mar. 92</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>439125</td>
</tr>
<tr>
<td>NORTRIPZYLINE HC</td>
<td>30 Mar. 92</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>187683</td>
</tr>
<tr>
<td>TRIAMETEREN</td>
<td>30 Apr. 92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2292</td>
</tr>
<tr>
<td>PIROXICAM</td>
<td>29 May 92</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>309756</td>
</tr>
<tr>
<td>GRIZEOFULVIN ULT</td>
<td>30 Jun. 92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11727</td>
</tr>
<tr>
<td>PYRAZINAMIDE</td>
<td>30 Jun. 92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>306</td>
</tr>
<tr>
<td>DIFLUNISAL</td>
<td>31 Jul. 92</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>96488</td>
</tr>
<tr>
<td>CARBIDOPA</td>
<td>28 Aug. 92</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>117233</td>
</tr>
<tr>
<td>PINDOLOL</td>
<td>03 Sep. 92</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>37648</td>
</tr>
<tr>
<td>KETOPROFEN</td>
<td>22 Dec. 92</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>107047</td>
</tr>
<tr>
<td>GEMFIBROZIL</td>
<td>23 Jan. 93</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>330539</td>
</tr>
<tr>
<td>BENZONATATE</td>
<td>29 Jan. 93</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2597</td>
</tr>
<tr>
<td>METHADONE HCL</td>
<td>15 Apr. 93</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1858</td>
</tr>
<tr>
<td>METHAZOLAMIDE</td>
<td>30 Jun. 93</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4792</td>
</tr>
<tr>
<td>ALPRAZOLAM</td>
<td>19 Oct. 93</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>614593</td>
</tr>
<tr>
<td>NADOLOL</td>
<td>31 Oct. 93</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>125379</td>
</tr>
<tr>
<td>LEVONORGESTREL</td>
<td>13 Dec. 93</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>47836</td>
</tr>
<tr>
<td>METOPROLOL TARTR</td>
<td>21 Dec. 93</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>235625</td>
</tr>
<tr>
<td>NAPROXEN</td>
<td>21 Dec. 93</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>456191</td>
</tr>
<tr>
<td>NAPROXEN SODIUM</td>
<td>21 Dec. 93</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>164771</td>
</tr>
<tr>
<td>GUANABENZ ACETAT</td>
<td>28 Feb. 94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>18120</td>
</tr>
<tr>
<td>TRIAZOLAM</td>
<td>25 Mar. 94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>71282</td>
</tr>
<tr>
<td>GLIPIZIDE</td>
<td>10 May 94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>189717</td>
</tr>
<tr>
<td>CIMETIDINE</td>
<td>17 May 94</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>547218</td>
</tr>
<tr>
<td>FLURBIPROFEN</td>
<td>20 Jun. 94</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>155329</td>
</tr>
<tr>
<td>SULFADIAZINE</td>
<td>29 Jul. 94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>HYDROXYCHLOROQUI</td>
<td>30 Sep. 94</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8492</td>
</tr>
</tbody>
</table>

The table shows the post-scandal data used in the study. The entry decisions of the four dominant firms are indicated by 1 for entry and 0 for no entry. Total Entrants are how many of the fifty-one potential entrants entered, including the dominant firms. Revenue is in thousands of dollars, and is the revenue of the branded product in the year before patent expiration.
Table 2. Posterior Distribution

<table>
<thead>
<tr>
<th>Parameter</th>
<th>3 firms</th>
<th>4 firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_c$</td>
<td>10.05</td>
<td>10.07</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>$\rho_c$</td>
<td>0.9866</td>
<td>0.9873</td>
</tr>
<tr>
<td></td>
<td>(0.00086)</td>
<td>(5.6e-05)</td>
</tr>
<tr>
<td>$\sigma_c$</td>
<td>0.3721</td>
<td>0.3675</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(3.0e-04)</td>
</tr>
<tr>
<td>$\kappa_c$</td>
<td>0.06655</td>
<td>0.07067</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(1.1e-04)</td>
</tr>
<tr>
<td>$\mu_r$</td>
<td>9.906</td>
<td>10.008</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td>$\sigma_r$</td>
<td>1.591</td>
<td>1.682</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.0023)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.9375</td>
<td>0.9375</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.96875</td>
<td>0.96875</td>
</tr>
<tr>
<td>$p_a$</td>
<td>0.9375</td>
<td>0.9375</td>
</tr>
</tbody>
</table>

CER firm 1 0.09 0.12
CER firm 2 0.08 0.09
CER firm 3 0.10 0.11
CER firm 4 0.14
CER all firms 0.09 0.11
MCMC Reps 3000000 3000000
stride 375 375

Shown is the mode of the multivariate posterior distribution not the modes of the marginal posterior distributions. The multivariate posterior mode does correspond to a set of parameter settings that actually occur in the MCMC chain whereas other measures of central tendancy such as the mean or marginal medians might not. Standard deviations are shown in parentheses. CER is the classification error rate when the parameters are set to the posterior mode. They are computed at Step 2e of the importance sampler. At that point in the algorithm the predicted actions $A_{i,t,j}^c$ are known for each firm $i$ at each time $t$ for each particle $j$ and can be compared to the observed actions $A_{it}$. The CER is the the proportion of the cases where $A_{it} \neq A_{i,t,j}^c$ computed both by firm and overall.
Figure 1. Profile Log Likelihood, Three Firm Model. Shown is the logarithm of the profile likelihood plotted for a tenth of the posterior standard deviation to the left and right of the maximum of the likelihood.
Figure 2. Profile Log Likelihood, Three Firm Model. Shown is the logarithm of the profile likelihood plotted for 24 posterior standard deviations to the left and right of the maximum of the likelihood. Points that violate support conditions and points below $10^{-6}$ of the maximum of the likelihood are not plotted. The horizontal line is at $10^{-3}$ of the maximum. The vertical lines indicate the range of the MCMC chain’s excursions after transients have died out.
Figure 3. Profile Log Likelihood, Four Firm Model. Shown is the logarithm of the profile likelihood plotted for a tenth of a posterior standard deviations to the left and right of the maximum of the likelihood.
Figure 4. Profile Log Likelihood, Four Firm Model. Shown is the logarithm of the profile likelihood plotted for 48 posterior standard deviations to the left and right of the maximum of the likelihood. Points that violate support conditions and points below $10^{-6}$ of the maximum of the likelihood are not plotted. The horizontal line is at $10^{-3}$ of the maximum. The vertical lines indicate the range of the MCMC chain’s excursions after transients have died out.
Figure 5. Marginal Posterior Distributions, Three Firm Model. Shown are histograms constructed from an MCMC chain for the three firm model with 3,000,000 repetitions at a stride of 375 for 8000 net. The salient feature of this graphic is the contrast of the histogram for the parameter $\kappa_c$ compared to that shown in Figure 6.
Figure 6. Marginal Posterior Distributions, Four Firm Model. Shown are histograms constructed from an MCMC chain for the four firm model with 1,400,000 repetitions at a stride of 375 for 3733 net. The salient feature of this graphic is the contrast of the histogram for the parameter $\kappa_c$ compared to that shown in Figure 5.
Figure 7. Cost, Revenue, and Entry Decisions. Plotted as a solid line in the first three panels is the logarithm of cost for the three dominant firms in the three firm model. The logarithm of cost is computed by averaging at Step 2e of the importance sampler at the maximum likelihood estimate. The circles in these plots indicate that the firm entered the market at that time point. The bottom panel shows the logarithm of total revenue. The numbers at the bottom are the count of the number of dominant firms who entered the market at that time point.
Figure 8. Cost and Entry Decisions of the Dominant Firms. Plotted is the logarithm of cost for the three dominant firms. The dashed line is under the three firm model, and the solid under the four firm model. The circles indicate the markets that Mylan entered, crosses the same for Novopharm, and the asterisks for Lemmon. The logarithm of cost as described in the legend of Figure 7.
Figure 9. Actual and Predicted Entry Decisions. Plotted as circles are the entry decisions of the three dominant firms in the three-firm model. The crosses are the average predictions of the three-firm model computed by averaging game solutions at Step 2e of the importance sampler at the maximum likelihood estimate.