# Dynamic Entry with Cross Product Spillovers: An Application to the Generic Drug Industry*

A. Ronald Gallant

Duke University and

New York University

Han Hong

Stanford University

Ahmed Khwaja

Yale University

### Abstract

Experience in one product market can potentially improve firm performance in a related product market in the future. Thus, entry into a market is determined not just by profits in that market but also by the future effect of entry on profitability in other markets. We formulate and estimate a dynamic game of entry decisions of firms in the presence of such spillovers using data on the generic drug industry. Spillovers imply that the unobserved component of a firm's profitability not only changes stochastically but is also is endogenous to past entry decisions. Therefore, the model needs to accommodate unobserved state variables that are endogenous to firm actions and serially correlated. We address the methodological challenge of estimating such a dynamic game using a sequential importance sampling based technique. Our estimates show significant spillover effects of entry on future profits. On average, each entry reduces costs by 7% at the next entry opportunity. In our sample there are on average eight entry opportunities annually. The average cumulative benefit of a firm that enters all eight markets in a year is 51%. We conclude that spillovers are critical in the equilibrium evolution of the structure of the generic drug industry.


Keywords: Dynamic Discrete Games, Sequential Importance Sampling, Dynamic Entry, Experience Spillovers, Unobserved Endogenous State Variables, Serial Correlation, Generic Pharmaceuticals.

JEL Classification: E00, G12, C51, C52

1

# 1 Introduction

One of the most significant decisions a firm makes is whether or not to enter a new market. Moreover, this decision can be drastically different in a dynamic setting when entry opportunities occur repeatedly over time from that in a static environment. In such a setting there may be dynamic spillovers of entry across products and over time, i.e., experience gained from entering a market could enhance profitability when entering other markets in the future. In the presence of dynamic spillovers a firm with a long term perspective may even place a high value on entering a market when the current opportunity is unprofitable provided the entry experience leads to greater cumulative profitability across all markets that the firm enters subsequently. Consequently, a forward looking firm's decision to enter a product market may be determined not just by expected profits in that market but also by the effect of the current entry decision on future profitability in other product markets. In this paper we analyze the entry decisions of forward looking firms to evaluate the effects of dynamic spillovers of current experience on future market performance. In order to do this we formulate and estimate a dynamic game theoretic model of entry decisions of firms in the presence of cross product spillovers.

There is a long history of research on spillovers of experience and more generally learning curves in various contexts and disciplines such as psychology (e.g., Ebbinghaus (1913)), engineering (e.g., Wright (1936)), and economics and marketing (e.g., Arrow (1962), Bass (1980), Dolan and Jeuland (1981)). A number of theoretical papers have examined the dynamic strategic effects of past experience on firm performance and industry structure (see e.g., Spence (1981), Fudenberg and Tirole (1983), Cabral and Riordan (1994)). Among these most relevant for us is a recent paper by Shen and Villas-Boas (2010) that includes analysis of the entry decisions of forward looking firms in dynamically evolving markets, in particular a case when firm capabilities differ.

Our paper contributes to the small but growing literature on estimating spillovers in a dynamic oligopolistic context. It is related to the work of Benkard (2004) who incorporates learning by doing in estimating an oligopolistic model of competition in the aircraft industry. However, an important difference is that whereas Benkard (2004) models the spillovers of

experience in producing a particular product, i.e., volume-cost relationship for a product, our analysis examines spillovers across products over time.[1]

There is an extensive literature on estimating strategic models of entry. Static games of entry have been analyzed among others by e.g. Bresnahan and Reiss (1990), Bresnahan and Reiss (1991a), Berry (1992), Scott-Morton (1999), Mazzeo (2002), Seim (2006), Zhu and Singh (2006), Orhun (2006), Ellickson and Misra (2008), Vitorino (2008), Datta and Sudhir (2009). There is also some empirical research on analyzing entry decisions in a dynamic oligopolistic context, e.g., Hitsch (2006), Ching (2009), Shen (2010). However, to our knowledge there are is no empirical work on estimating dynamic oligopoly models of entry when there are spillovers across markets.[2]

The model we develop incorporates dynamic spillovers from experience through the effect of entry on future costs. It allows for serially correlated firm specific costs that evolve endogenously based on past entry decisions. Hence, endogeneity of costs to past entry decisions induces heterogeneity among firms even if they are identical ex ante, which they need not be. Alternatively put, spillovers imply that the unobserved component of a firm's profitability in a product market not only changes stochastically but is also is endogenous to past entry decisions. Therefore, the model needs to accommodate unobserved state variables that are endogenous to firm actions and serially correlated. We address the methodological challenge of estimating such a model using a sequential importance sampling based technique. Although, there have been substantial recent developments in the empirical literature on estimation of dynamic games, incorporating unobserved (to the researcher) state variables that are serially correlated and endogenous remains prohibitively difficult.[3] Our work is related to the seminal paper by Keane (1994) that used sequential importance sampling to develop a computationally feasible simulation based estimator for limited dependent variable panel data models in the presence of serially correlated errors.[4] A special case of that estimator

---

[1]In contrast to spillovers within a particular product, Xu (2008) examines the spillover effects on a firm's productivity of its competitors' R&D in a dynamic oligopolistic framework using data from the Korean electric motor industry.

[2]Exceptions are Jia (2008) and Ellickson, Houghton, and Timmins (2009) who examine such spillovers in a *static* context.

[3]See Aguirregabiria and Mira (2009) for a recent survey of the literature on dynamic models of strategic interactions. Other excellent surveys of discrete games of entry are Berry and Reiss (2007) and Berry and Tamer (2006).

[4]See Ackerberg (2009) for an alternative use of importance sampling to provide computational gains in

3

is the well-known Geweke-Hajivassiliou-Keane (GHK) estimator (see e.g., Hajivassiliou, Mc-Fadden, and Ruud (1996), Geweke and Keane (2001)) that arises for a particular choice of importance sampling densities.

Our work using sequential importance sampling or particle filters is related to the research on dynamic linear models that has employed Kalman-Filters (e.g., Naik, Mantrala, and Sawyer (1998), Naik, Raman, and Winer (2005), Sriram, Chintagunta, and Neelamegham (2006), Jap and Naik (2008), Bass, Bruce, Murthi, and Majumdar (2007)). The important difference is that Kalman-Filters require the dynamic relationships to be specified in a linear fashion and error terms to be Gaussian whereas our paper allows for the dynamic relationships to be non-linear with no requirement that the error terms be Gaussian. Our paper is also related to the work of Bruce (2008). The critical distinction is that the particle filter in our research embeds the solution to a dynamic game.[5] Moreover, the unobserved state variable (i.e., cost) in our model is endogenously determined in contrast to most particle filter models which treat such variables as exogenous.[6]

Our research also contributes to the growing literature on the estimation of games. There is a growing literature on estimating static[7] and dynamic[8] games of *incomplete* information. The literature on estimating games of incomplete information has mostly relied on a two step estimation strategy building on the Conditional Choice Probability (CCP) estimator of Hotz and Miller (1993). The two step estimation strategy requires the assumption that there is no market or firm level unobserved heterogeneity other than a random shock that is distributed *IID* across both time and players. This assumption is restrictive because it

estimating game theoretic and dynamic discrete choice models.

[5]This is through the measurement or observation equation which in our case is the decision to enter and requires the solution of the dynamic game to compute. Details are provided in Section 5.

[6]Another example of using a sequential importance sampler to estimate dynamic models is the work of Fernandez-Villaverde and Rubio-Ramirez (2005). They estimate a macroeconomic dynamic stochastic general equilibrium model. The structure of dynamic stochastic general equilibrium models is closely related to that of dynamic discrete choice models. However, the discrete outcome and game theoretic strategic interaction aspects of our model are novel.

[7]See e.g., Bjorn and Vuong (1984), Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991c), Haile, Hortaçsu, and Kosenok (2008), Aradillas-Lopez (2005), Ho (2009), Ishii (2005), Pakes, Porter, Ho, and Ishii (2005), Augereau, Greenstein, and Rysman (2006), Seim (2006), Sweeting (2006), Tamer (2003), Manuszak and Cohen (2004), Rysman (2004), Gowrisankaran and Stavins (2004), and Bajari, Hong, Krainer, and Nekipelov (2006).

[8]See e.g., Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2004), Pakes, Ostrovsky, and Berry (2007), Pesendorfer and Schmidt-Dengler (2003), Ryan (2005), Collard-Wexler (2006), Bajari, Chernozhukov, Hong, and Nekipelov (2007) and Gowrisankaran, Lucarelli, Schmidt-Dengler, and Town (2008).

rules out unobserved dynamics in the latent state variables. It also rules out any private information that a player might have about competing firms that the researcher does not have. Aguirregabiria and Mira (2007) and Arcidiacono and Miller (2008) have extended the literature on two step CCP estimation of dynamic discrete models using finite mixture methods to allow for discrete forms of unobserved heterogeneity. In contrast, our method is applicable even when the unobserved variable is continuous. Further, our method allows for the unobserved variable to be endogenous to past actions.[9]

Static games of *complete* information have been estimated by, e.g., Bresnahan and Reiss (1991a), Berry (1992), Tamer (2003), Ciliberto and Tamer (2009) and Bajari, Hong, and Ryan (2004). 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 estimate a dynamic game of complete information.

Our paper also contributes to the literature on incorporating unobserved heterogeneity in the estimation of discrete dynamic programming models. In the *single* agent framework, there is a considerable amount of research that allows for time invariant unobserved heterogeneity, e.g., Keane and Wolpin (1997). With very few exceptions (e.g., Erdem and Keane (1996), Khwaja (2001)), there is almost no work that allows for serially correlated unobserved *endogenous* state variables. Bayesian approaches for single agent discrete dynamic programming models with unobserved state variables that are serially correlated over time have been developed by Imai, Jain, and Ching (2005) and Norets (2009). These papers use MCMC for integrating out the unobserved state variables. In contrast, we use sequential importance sampling to integrate out the unobserved state variables and use MCMC to iterate through the parameter space in estimating the model. In addition we are the first to apply this method to estimate a discrete dynamic game with multiple agents whereas the previous literature has focused on single agent models. Incorporating such strategic interactions poses

---

[9]In addition, while two step methods can be computationally attractive, we think that a likelihood based method, as the one we are employing, has advantages when the model is potentially misspecified. In this case, the likelihood based approach still minimizes a well defined Kullback-Leibler distance between the model and the data. On the other hand, it is not clear whether two step methods minimize a well defined distance between the model and the data if the model is potentially misspecified.

a substantial computational burden in solving and estimating the model.

We use MCMC methods in estimating the model. 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 compute and to conduct frequentist inference. Conversely, Bayesian inference is both theoretically justified and computationally attractive under these conditions. Moreover, Bayesian methods facilitate use of prior information for identification. Therefore we apply Bayesian methods in the application in this paper.

We estimate the model using data from generic drug industry. The generic pharmaceutical industry provides a good context for applying our model and method because of the following characteristics: (i) there is a high frequency of new entry opportunities that arise in sequence, (ii) there are potential economies of scope, e.g., due to working with a particular ingredient, therapeutic class, or form of drug (e.g., oral liquid or liquid injectable), possibly leading to dynamic spillovers of experience (see e.g., Scott-Morton (1999)), (iii) it is an oligopolistic industry with firms that ostensibly take a strategic and long-term perspective in making decisions and (iv) pay-offs are relatively big as are the fixed costs of entry so entry decisions have important consequences.

Our paper provides evidence on the dynamic spillover effects of experience in one product market on subsequent performance in the market for another product. Our estimates show significant spillover effects of entry on future profits. On average, each entry reduces costs by 7% at the next entry opportunity. In our data there are on average eight entry opportunities annually. The average cumulative benefit of a firm that enters all eight markets in a year is 51%. We conclude that spillovers are critical in the equilibrium evolution of the structure of the generic drug 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, (ii) when there are serially correlated unobserved endogenous state variables (either discrete or continuous), and (iii) an algorithm to solve the game is

available.

The rest of the paper is organized as follows. Section 2 describes the background and data and Section 3 the model. The estimation procedure is in explained in Sections 4 and 5. The results are presented in Section 6 and Section 7 concludes.

## 2   Background and Data

Generic pharmaceutical sales in the U.S. were valued at $58.5 billion in 2007. In the same year generics made up 65% of all prescriptions in the U.S. Moreover, in 2007 generic equivalents existed for 8,730 of the 11,487 branded drugs approved by the FDA and listed in its "Orange Book." Consequently analysis of consumption and production decisions in the generic pharmaceutical firm industry has been an topic of considerable empirical research. In an important paper Scott-Morton (1999) estimated a static entry model to show that entry can be predicted by a firm's organizational experience, size of the market and whether the entry opportunity is similar to the firm's existing portfolio of drugs. In subsequent work, Reiffen and Ward (2005) provided evidence on various features of the industry such as: (i) prices of generic drugs fall as the number of competitors increase but remain above long run marginal costs for less than eight entrants, (ii) greater number of firms enter larger markets, and (iii) profits first increase for early entrants and then decline with more entry.

In examining consumption decisions Crawford and Shum (2005) estimated a dynamic matching model of anti-ulcer drug choice for forward looking consumers. They provided evidence of (i) significant heterogeneity in drug efficacy across individuals and (ii) the benefits of learning from personal experience with drugs in dealing with uncertainty about efficacy. Ching (2010) in an analysis based on a dynamic random utility model of demand for prescription drugs found that after patent expiration (i) the rise in prices of brand name drugs can be explained by heterogeneity in consumer preferences, and (ii) the slow diffusion of generics into the market by consumer learning about product quality. In a companion paper analyzing firm decisions, Ching (2009) developed a finite horizon dynamic[10] oligopoly model

---

[10]Our work differs from Ching (2009) in that like Benkard (2004) he focuses on the dynamics of demand and supply over time within a product market. In contrast we focus on the dynamics that come from spillovers across product markets.

to examine the effect of reducing the approval times for generics. Estimating the model using data on the anti-hypertension drug Clonidine he found 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 which decreases the average profitability of firms. However, much remains to be understood about the generic pharmaceutical industry, especially with regard to the entry decisions of firms and whether past industry experience confers any strategic advantage when firms engage in dynamic oligopolist competition.

In this paper we use the unique data assembled by Scott-Morton (1999) on the entry decisions of generic drug manufacturers from 1984 to 1994.[11] This time period is particularly interesting because of the 1984 Waxman-Hatch Act which lowered barriers to entry for generic firms by permitting Abbreviated New Drug Applications (ANDAs). This vastly increased entry in to the generic market as it relaxed the requirements for generic firms to gain FDA approval. Generic firms did not have to repeat all the tests that the manufacturer of the pioneer branded product had undertaken but instead were only required to show that the generic product was bioequivalent and had identical characteristics in strength, dosage form and route. Also in 1989 there was a "generic scandal" in which some FDA reviewers were caught accepting bribes to expedite ANDAs, and some firms were found to have used the original branded drug in place of their own in the FDA review process.

We refer the reader to Scott-Morton (1999) for details of the data set and the institutional environment. We summarize the facts relevant for our study here. In particular we emphasize the features of the data that make it attractive for applying our model. (i) The preparation of an ANDA takes months to years because it requires the existence of manufacturing facilities that need to be inspected and approved by the FDA prior to launch of full scale operations. (ii) The sunk cost of submitting an ANDA is high even though it is much less than a new drug invention. For example, the average revenues for generic firms in one-firm markets are $10 million, while the costs of submitting an ANDA can range from $250,000 to $20 million. (iii) Furthermore, the size and heterogeneity of entry cost relative to the size of market revenue lead to a small number of entrants supported by each market. (iv) In addition, the FDA does

---

[11]We are grateful to Fiona Scott Morton for providing us with her data, and to Derek Gurney for answering our questions about the data.

not reveal when and from whom it receives ANDA applications. (v) 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. (vi) There are few late sequential movers who withdraw in response to rivals' approvals. (vii) Simultaneous moves in a dynamic context are an important feature of this industry. These characteristics 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 after they incur the sunk cost of entry.

In constructing our estimation sample we use the following information for each market opportunity: ANDA submission date, ANDA approval date, characteristics of drug (i.e., ingredient, concentration, route, form), characteristics of drug markets (i.e., drug therapeutic class, patent expiration date, revenue of brand name drug the year before expiration), characteristics of firms (i.e., parent or subsidiary firm, whether firm was indicted in a bribery scandal). In estimating the model 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 the scandal period. We take great care in processing the data between 1988 to 1993. We only study ANDAs for generic drugs that are orally ingested in the form of pills. Thus, we focus on spillovers from experience in producing drugs in the form of oral solids. 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 is defined as a unique combination of primary ingredient, patent expiration date and total revenue for the branded drug for the last year before patent expiration.[12] 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

---

[12]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.

four firms. In each case the remaining firms are combined into a category referred to as "other." The fraction of the market allocated to "other" is taken as given and is anticipated by the top firms when considering entry. The procedure we use to implement this is described in Section 3. 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. On average 3.3 firms enter a market, with the minimum number of entrants being one and the maximum being nine firms. The mean revenue in thousands of dollars is 126,901, the std. dev. is 161,580, the minimum is 72, and maximum 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, minimum is 4.3, and maximum is 13.3.

Table 1 about here

# 3 Model

In this section we formally describe the dynamic oligopolistic game of entry. 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. One could also think of the dynamics arising from evolution of demand, revenues and costs for a particular generic product as it diffuses through the market over time. This would lead to two time indices, one for the sequence of product markets opening over time and the other for profits over time within a product market. For computational feasibility, we abstract from the latter and assume that once a firm enters a market it realizes all the payoffs associated with that product market as a lump sum at the date of entry.

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

10

are $I$ firms in total so that the number of entrants in market $t$ is given by

$$N_t = \sum_{i=1}^{I} A_{it} \tag{1}$$

The primary source of dynamics is through costs. The evolution of current costs, $C_{it}$, is determined by past entry decisions and random shocks. The past entry decisions account for spillovers of past industry experience on production costs in the current entry opportunity. We will 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 market $t$ is

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

where $e_{it}$ is a 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 market $t$ if there was entry in market $t-1$,[13] $\mu_c$ is a location parameter that represents the overall average of the log cost over a long period of time.[14] The autoregressive parameter $\rho_c$ represents the degree of persistence between the current cost and its long run stationary level. We assume that all firms are ex ante identical, with the effects of current decisions on future costs creating heterogeneity between firms. Hence, none of these parameters are firm specific, i.e., indexed by $i$. Alternatively put, heterogeneity arises endogenously in the model depending on the past actions of the firms.

We assume, as in Scott-Morton (1999), that all firms observe each others' 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, a known component (or observable to the researcher based on past actions), $c_{k,i,t}$, and a component unobservable to the researcher,

---

[13]It should be noted that a priori the model allows for both positive ($\kappa_c > 0$) or negative spillovers ($\kappa_c < 0$).

[14]We do not model the $\mu_c$ as firm specific. However, in estimating the model we use a pre-estimation sample to allow the initial costs of firms to be different based on their past actions. Our approach emphasizes path-dependent unobserved heterogeneity in favor of firm fixed effects. Moreover, given the scarcity of data it will be difficult to distinguish between the two. We find the heterogeneity in the initial cost captures the key features of the data well. See Section 4 for details.

$c_{u,i,t}$, as follows:

$$c_{i,t} = c_{u,i,t} + c_{k,i,t} \tag{3}$$

$$c_{u,i,t} = \mu_c + \rho_c \left(c_{u,i,t-1} - \mu_c\right) + \sigma_c e_{it} \tag{4}$$

$$c_{k,i,t} = \rho_c \, c_{k,i,t-1} - \kappa_c A_{i,t-1} \tag{5}$$

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 of entry spillover at market $t$ 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 2) there are 40 openings in the period 1990-94, i.e. on average a market opens 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 obtain a correct chronological order of the data rather than to determine market entry dates precisely. We order markets according to the date when the first ANDA was received by the FDA for a particular market opportunity.

Our timing convention underlying the dynamic cost process (equation (2)), i.e., each time period $t$ represents the sequence of market openings and not calendar time, implies that the cost advantage of entry dissipates with additional entry 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 (see Section 6 below) this timing convention appears reasonable a posteriori. Also of note is that although we are calling the latent variable "cost"

for expositional convenience, it represents any unobserved variable that could have dynamic spillover effects of entry on profits. For example, the underlying sources of the spillover could be supply side factors like learning by doing or economies of scope or for that matter demand side factors like reputation about quality of a firm's products or its development of distribution networks. As stated earlier, the spillover effect estimated in this paper pertains to experience gained in producing drugs in the form of oral solids. Quantifying other sources of spillovers is beyond the scope of this paper and is left for future research as it would require expanding the state space leading to computational intractability.

The total (lump sum) 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,

$$r_t = \mu_r + \sigma_r e_{I+1,t} , \tag{6}$$

where $e_{I+1,t}$ is normally distributed with mean zero and unit variance. In equation (6), $\mu_r$ is a location parameter that reflects the average total revenue 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 from the last year the brand name drug was on patent. We interpret this value as being exogenously determined solely by the firm manufacturing the branded product 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.[15] Therefore, we consider only the dominant firms. 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 $\log N_t^a = b \log R_t$, with $b \approx 0.092$, is a reasonable approximation to the total number of firms that enter

---

[15]See Benkard, Weintraub, and Roy (2007) for a discussion of the concept of oblivious equilibrium and the associated method to compute the solution to dynamic games when the number of players is very large.

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 $\log R_{anticipated} \geq \log R - \log N^a = \log R - b \log R = \log \left( R^{1-b} \right)$. These considerations suggest that a reasonable functional form for dominant firm $i$'s per period profit at time $t$ is

$$A_{it} \left( R_t^\gamma / N_t - C_{it} \right), \tag{7}$$

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

$$\sum_{j=0}^{\infty} \beta^j A_{i,t+j} \left( R_{t+j}^\gamma / N_{t+j} - C_{i,t+j} \right), \tag{8}$$

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

$$V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t) \tag{9}$$

$$\begin{aligned}
= \quad & A_{it} \left( R_t^\gamma / N_t - C_{it} \right) \\
+ \quad & \beta \, \mathcal{E} \left[ V_i(A_{i,t+1}^E, A_{-i,t+1}^E, C_{i,t+1}, C_{-i,t+1}, R_{t+1}) \, | \, A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t, \right],
\end{aligned}$$

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 and the action profile at time $t$. Therefore $V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t)$ represents the expected 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

$$V_i(A_{i,t}^E, A_{-i,t}^E, C_{i,t}, C_{-i,t}, R_t) \geq V_i(A_{i,t}, A_{-i,t}^E, C_{i,t}, C_{-i,t}, R_t) \quad \forall\, i, t, \tag{10}$$

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.

This is a game of complete information. Hence, if the state, which includes 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

$$V_i(C_{i,t}, C_{-i,t}, R_t) = V_i(A_{i,t}^E, A_{-i,t}^E, C_{i,t}, C_{-i,t}, R_t). \tag{11}$$

The ex ante value function satisfies the Bellman equation

$$V_i(C_{it}, C_{-i,t}, R_t) \tag{12}$$

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

where $N_t^E$ is the number of firms that enter, which can be computed using equation (1), i.e., $N_t^E = \sum_{i=1}^{I} A_{it}^E$. 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 expected 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. In contrast to equation (9), the expectations operator here is over the conditional distribution of the state variables in time period $t+1$ with the value function evaluated at the best response strategy profile.

Reny (1999) demonstrated 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. A comprehensive discussion of several results for existence of equilibria in Markovian games is provided by Dutta and Sundaram (1998). More results on existence of equilibria in dynamic oligopolistic models are to be found in Doraszelski and Satterthwaite (2007). When the state space can only take on a finite set of values, 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 baseline case 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 that 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 a continuous state space. Theorem 3.1 is of interest to us because its proof relies on a dynamic programming approach that motivates our computational strategy described in detail in Technical Appendix A (see also Rust (1996) 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 automatically satisfy the regularity conditions of Theorem 5.1. While the results described above imply the guaranteed existence of equilibria for a slightly modified version of the game proposed by us, 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. Multiple equilibria usually take the form of a situation where one or another firm can profitably enter but if both enter they both will incur losses whereas if neither enters then one of them would have an incentive to deviate. When we encounter multiple equilibria we adopt an explicit equilibrium selection rule as follows.[16] We pick the equilibrium with the lowest total cost. This idea is similar to that used by Berry (1992) and 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 the equilibrium condition (10) is accepted as the solution. Note that our distributional assumptions (described in the Technical Appendix A) 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

---

[16]In practice we don't find a large number of multiple equilibria, e.g., in the three firm game the frequency of multiple equilibria is about 4%. See Section 6 for additional details.

to be able to compute a pure strategy equilibrium for each parameter value as soon as we compute a sufficient number of value function iterations.

In general, because we use importance sampling and Bayesian MCMC methods, as long as one is willing to assume that the data is generated from a pure strategy equilibrium at the true parameter value, a parameter and state trajectory pair that has no equilibrium can be regarded as an irrelevant portion of the parameter space and rejected at the MCMC likelihood comparison step. That is, given parameters, only particles that produce equilibria are retained. A state trajectory that does not produce an equilibrium could not have occurred.

# 4    Likelihood Computation

In this section we describe how the likelihood is computed using the solution of the dynamic game and the MCMC algorithm. The details of the solution method are provided in Technical Appendix A.

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. To compute a likelihood that depends only on observable variables in the data, the latent state variables are integrated out using sequential importance sampling. Using the likelihood that depends only on observable data, 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) For that parameter value, generate values for the latent variable over the sample period by means of the sequential importance sampler. (3) Solve the dynamic game to compute the equilibrium outcome as function of the observed and unobserved state variables and the parameter value. (4) Use the equilibrium outcome generated from the solution to compute a likelihood that depends on the observed data and latent state variables (at the given 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 (at the given parameter value).[17] (6) Use the likelihood

---

[17]As the name suggests, the sequential importance sampler does this by averaging sequentially as one progresses through the sample rather than storing all latent variable trajectories prior to averaging them.

that depends only on observed variables to make the accept/reject decision of the MCMC algorithm. Cycling through steps (1) to (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. Details follow.

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

$$X_t = (C_{u,1,t}, \ldots, C_{u,I,t}). \tag{13}$$

Also, denote the variables that we can observe by

$$Y_t = (A_{1t}^o, \ldots, A_{It}^o, C_{k,1,t}, \ldots, C_{k,I,t}, R_t). \tag{14}$$

As previously, a lower case variable denotes the logarithm of an upper case variable with the exception that $a_t = A_t$. With these conventions, $x_t = (c_{u,1,t}, \ldots, c_{u,I,t})$, and $y_t = (a_{1t}, \ldots, a_{It}, c_{k,1,t}, \ldots, c_{k,I,t}, r_t)$. Recall that cost evolves as

$$c_{i,t} = c_{u,i,t} + c_{k,i,t} \tag{15}$$

$$c_{u,i,t} = \mu_c + \rho_c \left( c_{u,i,t-1} - \mu_c \right) + \sigma_c e_{it} \tag{16}$$

$$c_{k,i,t} = \rho_c \, c_{k,i,t-1} - \kappa_c A_{i,t-1}^o \tag{17}$$

and revenue evolves as

$$r_t = \mu_r + \sigma_r e_{I+1,t}. \tag{18}$$

We have data for both the pre- and post-scandal periods. The pre-scandal period is indexed by $t = -n_0, \ldots, 0$ and the values of $Y_t$ over the pre-scandal period are denoted by $Y_{pre}$. The post-scandal 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-scandal data unsuitable for general estimation, it can still be used for two purposes: The pre-scandal entries $\{A_{it}^o\}_{t=-n_0}^0$ can be used to compute the last two pre-scandal values $c_{k,i,-1}$ and $c_{k,i,0}$ of the observable part of log cost for each firm. Although we do not have firm fixed effects, i.e., $\mu_c$ are not firm specific, it is through these differences in initial costs that we allow for some form of firm specific heterogeneity. As previously stated, our approach emphasizes path-dependent unobserved heterogeneity in favor of firm fixed effects. As seen in the results

described in Section 6 our approach captures the key features of the data well. The scandal log revenues $\{r_t\}_{t=-n_0}^0$ can be used to help identify the parameters $\mu_r$ and $\sigma_r$.

We compute the last two pre-scandal values $c_{k,i,-1}$ and $c_{k,i,0}$ for each firm by running the recursion (17) started at $-n_0$ over the observed choices $\{A_{it}^o\}_{t=-n_0}^0$. This gives us the vectors $y_{-1}$ and $y_0$ because $(R_{-1}, A_{-1}^o)$ and $(R_0, A_0^o)$ are also in $Y_{pre}$. How we use the pre-scandal log revenue to help identify the parameters $\mu_r$ and $\sigma_r$ is discussed in Subsection 4.3.

## 4.1  Boundedly Rational Likelihood

Since we are estimating a game of pure strategy, a density for the strategy profile $A_t$ that depends only on $(x_t, r_t, y_{t-1})$ and the model parameters introduced thus far[18] would put mass one on a single value of $A_t$. The implication is that a likelihood that depends only on these parameters and $\{(x_t, r_t, y_{t-1})\}_{t=1}^n$ would be one if we predict every entry decision perfectly and zero otherwise. Estimation with such a likelihood would not be feasible.[19]

We resolve this problem by defining a misclassification probability $q_a = 1 - p_a$, $0 < p_a < 1$, and adopting[20] the following likelihood for an observed (ex post) action profile $A_t^o$

$$p(A_t^o \mid r_t, x_t, y_{t-1}, \theta) = \prod_{i=1}^{I} (p_a)^{I(A_{it}^o = A_{it})} (1 - p_a)^{I(A_{it}^o \neq A_{it})} \tag{19}$$

where $A_{it}$ is the predicted entry decision computed from the model given $(x_t, r_t, y_{t-1})$ and

$$\theta = (\mu_c, \rho_c, \sigma_c, \kappa_c, \mu_r, \sigma_r, \gamma, \beta, p_a) \tag{20}$$

using the method described in Technical Appendix A. The full likelihood for the data is, of course, the product $\prod_{t=1}^n p(A_t^o \mid r_t, x_t, y_{t-1}, \theta)$.

There are several ways that one can justify this likelihood of which the most appealing is that there is a small probability $q_a$ that planned entry decisions are not realized. This

---

[18]Recall that these parameters are $\mu_c, \rho_c, \sigma_c, \kappa_c, \mu_r, \sigma_r, \gamma, \beta$.

[19]It is worth noting that such a problem hasn't arisen in the existing literature on estimating static games of complete information with pure strategies, e.g., Berry (1992). One reason is that these models don't allow for any unobserved heterogeneity. Hence conditional on observables the identity of a firm is irrelevant. The implication being that estimation can be based on the total number of entrants rather than which particular firms entered a market. Moreover, often the choice probabilities of entry (or the probability of the number of entrant firms) do not require solving the game numerically and can be computed in part based on expressions describing the optimal decisions of firms.

[20]Our approach is similar to that frequently adopted in likelihood based inference when estimating reservation wages in labor search models, and in estimating models of option pricing and yield curves.

could come about, e.g., because a decision to enter is thwarted by disapproval of an ANDA application or because a decision not to enter is effectively reversed by an acquisition. Less appealing interpretations are that (19) represents measurement error either on the part of the econometrician or the firm due to, e.g., clerical error or firms not being able to observe competitors' entry decisions before a decision must be made on the next drug coming off patent. The various interpretations have different likelihood implications. We shall adopt the first interpretation that there is a small probability that plans are not realized.

A firm can either take the probability $q_a$ that plans will not be realized into account or not in making its entry decision. Because ignoring $q_a$ dramatically reduces the cost of computing a decision and because, as seen later, ignoring $q_a$ for small $p_a$ has little effect on decisions, solving the game with $q_a$ ignored can be viewed as a boundedly rational solution strategy.[21] Conversely, taking $q_a$ into account can be viewed as a fully rational strategy which confers the additional benefit that $q_a$ has a clear structural interpretation.

The boundedly rational likelihood is easy to describe: It is (19) with $A_t$ computed as described in Technical Appendix A. When it is necessary to distinguish a boundely rational equilibrium, we use the notation $A_t^{BR}$.

## 4.2   Fully Rational Likelihood

In a fully rational solution, firms account for the potential discrepancy between future equilibrium and ex post action profiles in forming continuation values. This requires redefinition of the choice specific value function (9) as follows:

$$
\begin{aligned}
V_i^f&(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t) \\
&= \sum_{l_1=0}^{1} p_a^{l_1=A_{1t}} q_a^{l_1 \neq A_{1t}} \cdots \sum_{l_I=0}^{1} p_a^{l_I=A_{It}} q_a^{l_I \neq A_{It}} \Bigg\{ l_i \left( \frac{R_t^\gamma}{\sum_{j=1}^I l_j} - C_{it} \right) \\
&\quad + \beta\, \mathcal{E}\Big[ V_i^f(A_{i,t+1}^E, A_{-i,t+1}^E, C_{i,t+1}, C_{-i,t+1}, R_{t+1}) \,|\, L_{i,t}, L_{-i,t}, C_{i,t}, C_{-i,t}, R_t \Big] \Bigg\},
\end{aligned}
\tag{9'}
$$

---

[21]See Che, Seetharaman, and Sudhir (2007) for another example of employing bounded rationality in estimating a model of forward looking firms.

where $L_{i,t} = l_i$ and $L_{-i,t}$ is $(l_1, \ldots, l_I)$ with $l_i$ deleted. Note that $L_{i,t}$ and $L_{-i,t}$ affect the conditioning information via equations (15) through (17). Specifically,

$$
\begin{aligned}
c_{j,t+1} &= c_{u,j,t+1} + c_{k,j,t+1} \\
c_{u,j,t+1} &= \mu_c + \rho_c \left( c_{u,j,t} - \mu_c \right) + \sigma_c e_{j,t+1} \\
c_{k,j,t+1} &= \rho_c \, c_{k,j,t} - \kappa_c l_j
\end{aligned}
$$

both for $j \in L_{i,t}$ and $j \in L_{-i,t}$; $l_j$, like $A_j$, is one for entry and zero otherwise.

A fully rational stationary pure strategy equilibrium is defined by a best response strategy profile $(A_{i,t}^E, A_{-i,t}^E)$ that satisfies (10) with $V^f$ replacing $V$. When it is necessary to distinguish a fully rational equilibrium, we use the notation $A_t^{FR}$.

The ex ante value function is related to the choice specific value function by (11) with $V^f$ replacing $V$ and $A_t^{FR}$ replacing $A^E$. The expression that replaces (12) is more complicated and cannot be described by simple plug-in rules as with the replacements for (10) and (11). However, we have no need for an explicit expression because the ex ante Bellman equation is defined operationally by the algorithm described in Technical Appendix A. That algorithm makes no explicit use of (12). It only uses the choice specific value function defined by (9) and the fact that the left hand sides of (9) and (12) are the same at an equilibrium due to (11). All that is required to compute the fully rational equilibrium is to modify Step 5 of that algorithm by replacing (9) by (9'), $A_{i,t}$ by $L_{i,t}$, and $A_{-i,t}$ by $L_{-i,t}$. The fully rational likelihood is (19) with $A_t = A_t^{FR}$.

The algorithm for parameter estimation described in Section 5 applies to both the boundedly rational likelihood and the fully rational likelihood. The results from implementing both for the data displayed in Table 1 are discussed in Section 6.

## 4.3   Pre-Scandal Information

While the scandal may have affected which firms participated in the market post-scandal, there is no reason to believe that market opportunities were different pre- and post-scandal. Therefore the pre-scandal 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 $p(Y_{pre} \,|\, \theta)$

for convenience.

# 5 Parameter Estimation

Douced, de Freitas, and Gordon (2001) present a concise description of the sequential impor-
tance sampler that we follow in describing our analysis. The densities relevant to a sequential
importance sampler are the transition density of the hidden state vector

$$p(x_t \mid x_{t-1}, \theta), \tag{21}$$

which is defined by recursion equation (16), the initial density

$$p(x_0 \mid \theta), \tag{22}$$

which, from equation (16), is normal with mean $\mu_c$ and standard deviation $\sigma_c / \sqrt{1 - \rho_c^2}$, and
the observation density

$$p(y_t \mid y_{t-1}, x_t, \theta) = p(A_t^o \mid r_t, y_{t-1}, x_t, \theta) \, p(r_t \mid y_{t-1}, x_t, \theta), \tag{23}$$

where, from equation (18), $p(r_t \mid y_{t-1}, x_t, \theta)$ 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 $x_0^{(j)}$ for $j = 1, \ldots, N$ from the initial density equa-
   tion (22).

   (b) Compute

   $$
   \begin{aligned}
   p(y_0 \mid \theta) &= \int p(y_0 \mid y_{-1}, x_0, \theta) \, p(y_{-1}, x_0 \mid \theta) \, dx_0 \\
   &\doteq \frac{1}{N} \sum_{j=1}^{N} p(y_0 \mid y_{-1}, x_0^{(j)}, \theta).
   \end{aligned}
   $$

2. For $t = 1, \ldots, n$

   (a) For each particle, draw $\tilde{x}_t^{(j)}$ from the transition density equation (21) and set

   $$\tilde{x}_{0:t}^{(j)} = (x_{0:t-1}^{(j)}, \tilde{x}_t^{(j)}).$$

(b) For each particle compute the particle weights $\hat{w}_t^{(j)}$ using the observation density equation (23); i.e.

$$\tilde{w}_t^{(j)} = p(y_t \mid y_{t-1}, \tilde{x}_t^{(j)}, \theta).$$

The parametrization in equation (19) eliminates the problem that the weights could all be zero. It is at this stage that the game needs to be solved in order to be able to compute the weights $\hat{w}_t^{(j)}$. This adds to the computational burden considerably. Hence, it is at this step that our work differs significantly from the literature on dynamic models using sequential importance sampling (e.g., Naik, Raman, and Winer (2005), Sriram, Chintagunta, and Neelamegham (2006), Bruce (2008)) or the research using MCMC to incorporate serially correlated unobserved variables in single agent models (e.g., Imai, Jain, and Ching (2005), Norets (2009)).

(c) Normalize the weights so that they sum to one

$$\hat{w}_t^{(j)} = \frac{\tilde{w}_t^{(j)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}}.$$

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

(e) Compute

$$p(y_t \mid y_{1:t-1}, \theta, p_a) = \int p(y_t \mid y_{t-1}, x_t, \theta) \, p(y_{t-1}, x_t \mid y_{1:t-1}, \theta) \, dx_t$$

$$\doteq \frac{1}{N} \sum_{j=1}^N p(y_t \mid y_{t-1}, x_t^{(j)}, \theta).$$

Note that $p(y_t \mid y_{t-1}, x_t^{(j)}, \theta)$ does not have to be recomputed here if the weights $\tilde{w}_t^{(j)}$ 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. The likelihood is

$$\mathcal{L}(\theta) = p(y_{0:t} \mid \theta) = p(Y_{pre} \mid \theta, p_a) p(y_0 \mid \theta) \prod_{t=1}^{n} p(y_t \mid y_{0:t-1}, \theta).$$

We use the Metropolis-Hastings algorithm to estimate the model. Briefly, the method is as follows.[22] 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 $\theta_j^o$ with a draw from the normal distribution with mean $\theta_j^o$ 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 likelihood is hierarchical in that given model parameters and conditional upon the latent cost variables, it can be evaluated by solving the game. Given this structure, estimation can be viewed as a double-layer 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 averaging out the latent cost distribution. At this point the MCMC accept/reject decision is made and the MCMC chain is moved forward. One iterates through the outer loop to obtain the complete MCMC 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.

---

[22] Additional details are provided in a companion Web Appendix B.

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; we experimented with different number of particles till the results were not sensitive to the choice of number of particles. 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 6.

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.[23] 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%." Theoretically, 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)). Adding 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 an annual 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 yield estimates that meander about 0.93. Therefore, based on the plausible lower bound of 0.908 derived in Section 3 and our experience from trying to estimate $\gamma$, we take 0.93 to be a reasonable value. Rounding to a nearby fractional

---

[23]See Kasahara and Shimotsu (2008) and Hu and Shum (2008) for more results on nonparameteric identification of single agent dynamic discrete choice models.

power of two, we set $\gamma = 0.9375$.

For the remaining parameters we use flat, noninformative priors that impose these support conditions:[24] $-1 \leq \rho_c \leq 1$, $0 \leq \kappa_c$, $0 < \sigma_c$, and $0 < \sigma_r$.[25]

# 6    Results

We estimate both the boundedly rational and fully rational models 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 (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 for the boundedly rational model and in Table 3 for the fully rational model. 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.

Table 2 about here

Table 3 about here

As evident from these two tables, the parameter estimates are very similar across the boundedly rational and the fully rational models. Therefore in the rest of paper we will discuss the results using the boundedly rational model. The parameters are tightly estimated[26] and, as seen from the extremely low classification error rates, model predictions are quite

---

[24]There is one exception. A value of $\kappa_c < 0$ would imply negative spillovers or an indication of resource or capacity constraints. We allowed for this when we initially started estimating the model but did not find any evidence of this in our preliminary work. Hence we adopted a prior that restricts spillovers to be non-negative.

[25]Open ended ranges actually should have large upper and lower bounds to assure stationarity of the MCMC chain. We find that bounds do not interfere with the chain. Histograms of the marginal posterior distributions of the parameters providing evidence of this are available from the authors on request.

[26]Despite the small standard deviations shown in Table 2, we find that the profile likelihoods suggest that the MCMC chain adequately explored the posterior density. See Figures 5 and 7 in Web Appendix B. The likelihood is proportional to the posterior because priors are flat.

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) cost process (equation (2)) is 51%. However, $\sigma_c$ is large, so that this reduction can easily be eliminated by a cost shock.[27]

Figure 1 about here

Figure 2 about here

Figure 3 about here

Figure 1 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 2 plots together the log cost of the three dominant firms from both the 3 and 4 firm 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 1. 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 3 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 using a threshold of

---

[27]One source of this large standard deviation may be organizational forgetting (see e.g., Benkard (2000), Besanko, Kryukov, Doraszelski, and Satterthwaite (2007)). However, since we do not explicitly model forgetting in the AR(1) cost process we are cautious in making this interpretation.

0.5 to predict entry (i.e., entry if predicted probability $\geq 0.5$; no entry otherwise) for the average predictions shown in Figure 3.

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 analogous to the previous boundedly rational model that might be played instead of the game with payoffs (9). They could play a game with payoffs

$$V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t) = A_{it} \left( R_t^\gamma / N_t - C_{u,i,t} \right), \tag{24}$$

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

$$V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t) = A_{it} \left( R_t^\gamma / N_t - C_{it} \right) \tag{25}$$

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, we adopt an explicit equilibrium selection rule, i.e., we pick the equilibrium with the lowest total cost.

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 1, 2, and 3 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) (i.e., the "unknown" component of costs) and not enough on (5) (i.e., "known" component of costs). Stated differently, one might surmise that the effect 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.

# 7    Conclusions

In this paper we provide evidence on the dynamic spillover effects of past industry experience in one product market on performance in other product markets for the generic pharmaceutical industry. Our stylized model fits the data well, i.e., the classification error rates are small. Our results also enhance understanding of entry decisions in the pharmaceutical industry. We find evidence that past entry affects current costs and that the dynamic evolution of the production cost plays an important role in the equilibrium path of the generic pharmaceutical industry structure. In order to estimate our model we develop a procedure based on sequential importance sampling to estimate a dynamic discrete game that includes serially correlated unobserved endogenous state variables. Our method is more generally applicable to estimating dynamic models in which heterogeneity between agents arises from serially

correlated unobserved endogenous state variables, e.g., when there is state dependence in consumer demand or accumulation or depreciation of goodwill of brands in strategic advertising games. Our method is feasible whenever there is available an algorithm for solving the game and a Markovian representation of the latent dynamics.

To summarize, our paper contributes to the literature on estimating spillovers and also extends the literature on estimating dynamic discrete games to allow for endogenously arising unobserved heterogeneity between players. However, several limitations remain which provide opportunity for future research. The spillover effect estimated in this paper is related to experience gained in producing drugs in the form of oral solids. It may be worthwhile in future work to quantify other sources of spillovers. This is beyond the scope of this paper due to the prohibitive computational burden it would impose by expanding the state space. Another important extension would be to allow for estimation of dynamic games where the strategy set is mixed discrete-continuous, e.g., introduction of a new brand *and* the associated decision about advertising expenditure. Furthermore, the methods we exposit may provide an alternative procedure to incorporate unobserved heterogeneity in estimating dynamic games of incomplete information. We hope our paper will provide a foundation to address some of these issues in future.

# References

ABRAMOWITZ, M., AND I. A. STEGUN (1964): *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover.

ACKERBERG, D. (2009): "A new use of importance sampling to reduce computational burden in simulation estimation," *Quantitative Marketing and Economics*, 7(4), 343–376.

AGUIRREGABIRIA, V., AND P. MIRA (2007): "Sequential estimation of dynamic discrete games," *Econometrica*, 75(1), 1–53.

——— (2009): "Dynamic discrete choice structural models: A survey," *Journal of Econometrics.*

ARADILLAS-LOPEZ, A. (2005): "Semiparametric estimation of a simultaneous game with incomplete information," working paper, University of Wisconsin, Madison.

ARCIDIACONO, P., AND R. MILLER (2008): "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," Working Paper, Duke University.

ARROW, K. (1962): "The Economic Implications of Learning by Doing," *The Review of Economic Studies*, 29(3), 155–173.

AUGEREAU, A., S. GREENSTEIN, AND M. RYSMAN (2006): "Coordination versus differentiation in a standards war: 56k modems," *RAND Journal of Economics*, 37(4), 887–909.

BAJARI, P., L. BENKARD, AND J. LEVIN (2004): "Estimating Dynamic Models of Imperfect Competition," working paper, forthcoming Econometrica.

BAJARI, P., V. CHERNOZHUKOV, H. HONG, AND D. NEKIPELOV (2007): "Semiparametric Estimation of a Dynamic Game of Incomplete Information," working paper, University of Michigan and Stanford University.

BAJARI, P., H. HONG, J. KRAINER, AND D. NEKIPELOV (2006): "Estimating Static Models of Strategic Interactions," working paper, Duke University, Stanford University and University of Minnesota.

BAJARI, P., H. HONG, AND S. RYAN (2004): "Identification and Estimation of Discrete Games of Complete Information," Working paper, Department of Economics, Duke University.

BASS, F. (1980): "The Relationship Between Diffusion Rates, Experience Curves, and Demand Elasticities for Consumer Durable Technological Innovations," *The Journal of Business, Part 2: Interfaces Between Marketing and Economics*, 53(3), S51–S67.

BASS, F., N. BRUCE, B. P. S. MURTHI, AND S. MAJUMDAR (2007): "Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of the Advertising Sales Relationship," *Marketing Science*, 26(2), 179–195.

BENKARD, L. (2000): "Learning and Forgetting: The Dynamics of Aircraft Production," *American Economic Review*, 90, 1034–1054.

——— (2004): "A Dynamic Analysis of the Market for Wide-bodied Commercial Aircraft," *Review of Economic Studies*, 71, 581–611.

BENKARD, L., G. WEINTRAUB, AND B. V. ROY (2007): "Markov Perfect Industry Dynamics with Many Firms," working paper, Stanford University.

BERRY, S. (1992): "Estimation of a model of entry in the airline industry," *Econometrica*, 60(4), 889–917.

BERRY, S., AND P. REISS (2007): "Empirical models of entry and market structure," *Handbook of industrial organization*, 3, 1845–1886.

BERRY, S., AND E. TAMER (2006): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics, Vol 2, Ninth World Congress, 46-85*, ed. by Blundell, Newey, and Persson. Cambridge University Press.

BESANKO, D., Y. KRYUKOV, U. DORASZELSKI, AND M. SATTERTHWAITE (2007): "Learning-by-Doing, Organizational Forgetting, and Industry Dynamics," Working Paper, Harvard University.

BJORN, P. A., AND Q. VUONG (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," SSWP No. 537, Caltech.

BRESNAHAN, T., AND P. REISS (1990): "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531–553.

——— (1991a): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57—81.

——— (1991b): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, 57—81.

———— (1991c): "Entry and competition in concentrated markets," *Journal of Political Economy*, 99, 977–1009.

BRUCE, N. (2008): "Pooling and Dynamic Forgetting Effects in Multi-Theme Advertising: Tracking the Advertising Sales Relationship with Particle Filters," *Marketing Science*, 27(4), 659–673.

CABRAL, L., AND M. RIORDAN (1994): "The Learning Curve, Market Dominance, and Predatory Pricing," *Econometrica*, 62, 1115–1140.

CHE, H., P. B. SEETHARAMAN, AND K. SUDHIR (2007): "Bounded Rationality in Pricing under State Dependent Demand: Do Firms Look Ahead? How Far Ahead?," *Journal of Marketing Research*, 44(3), 434–449.

CHERNOZHUKOV, V., AND H. HONG (2003): "A MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115(2), 293–346.

CHING, A. (2009): "A Dynamic Oligopoly Structural Model for the Prescription Drug Market After Patent Expiration," forthcoming in International Economic Review.

CHING, A. (2010): "Consumer learning and heterogeneity: Dynamics of demand for prescription drugs after patent expiration," *International Journal of Industrial Organization*.

CILIBERTO, F., AND E. TAMER (2009): "Market structure and multiple equilibria in airline markets," *Econometrica*, 77(6), 1791–1828.

COLLARD-WEXLER, A. (2006): "Demand Fluctuations and Plant Turnover in the Ready-Mix Concrete Industry," Working paper, NYU.

CRAWFORD, G., AND M. SHUM (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73, 1135–1174.

DATTA, S., AND K. SUDHIR (2009): "Sleeping with the "Frenemy": The Agglomeration-Differentiation Tradeoff in Spatial Location Choice," Working paper, School of Management, Yale University.

DOLAN, R. J., AND A. P. JEULAND (1981): "Experience Curves and Dynamic Demand Models: Implications for Optimal Pricing Strategies," *The Journal of Marketing*, 45(1), 52–62.

DORASZELSKI, U., AND M. SATTERTHWAITE (2007): "Computable Markov-Perfect Industry Dynamics: Existence, Purification, and Multiplicity," Working Paper, Harvard University.

DOUCED, A., N. DE FREITAS, AND N. GORDON (2001): "An Introduction to Sequential Monte Carlo Methods," in *Sequential Monte Carlo Methods in Practice*, ed. by A. Douced, N. de Freitas, and N. Gordon, pp. 3–13. Springer.

DUTTA, P. K., AND R. K. SUNDARAM (1998): "The Equilibrium Existence Problem in General Markovian Games," in *Organizations with Incomplete Information*, ed. by M. Majumdar, pp. 159–207. Cambridge University Press.

EBBINGHAUS, H. (1913): *Memory. A Contribution to Experimental Psychology.* New York: Teachers College, Columbia University.

ELLICKSON, P., S. HOUGHTON, AND C. TIMMINS (2009): "Estimating Network Economies in Retail Chains: A Revealed Preference Approach," Working paper, Simon School of Business Administration, University of Rochester.

ELLICKSON, P., AND S. MISRA (2008): "Supermarket pricing strategies," *Marketing Science*, 27(5), 811–828.

ERDEM, T., AND M. KEANE (1996): "Decision-making under uncertainty : Capturing dynamic brand choice processes in turbulent consumer goods markets," *Marketing Science*, pp. 1–20.

FERNANDEZ-VILLAVERDE, J., AND J. F. RUBIO-RAMIREZ (2005): "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood," *Journal of Applied Econometrics*, 20, 891–910.

FUDENBERG, D., AND J. TIROLE (1983): "Learning by Doing and Market Performance," *Bell Journal of Economics*, 14, 522–530.

GAMERMAN, D., AND H. F. LOPES (2006): *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd Edition.* Chapman & Hall.

GEBHARDT, W. R., C. M. C. LEE, AND B. SWAMINATHAN (2001): "Toward and Implied Cost of Capital," *Journal of Accounting Research*, 39, 135–176.

GELMAN, A., G. O. ROBERTS, AND W. GILKS (1996): "Efficient Metropolis Jumping Rules," in *Bayesian Statistics V*, ed. by J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, pp. 599–607. Oxford University Press.

GEWEKE, J., AND M. KEANE (2001): "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics, Vol 5, chapter 5*, ed. by J.J.Heckman, and E. Leamer, pp. 3465–3564. Elsevier Science.

GOLUB, G. H. (1973): "Some modified matrix eigenvalue problems," *SIAM Review*, 15, 318–334.

GOLUB, G. H., AND J. H. WELSCH (1969): "Calculation of Gaussian quadrature rules," *Mathematics of Computation*, 23, 221–230.

GOWRISANKARAN, G., C. LUCARELLI, P. SCHMIDT-DENGLER, AND R. TOWN (2008): "Government Policy and the Dynamics of Market Structure: Evidence from Critical Access Hospitals," Working paper, University of Arizona.

GOWRISANKARAN, G., AND J. STAVINS (2004): "Network Externalities and Technology Adoption: Lessons from Electronic Payments," *RAND Journal of Economics*, 35(2), 260–276.

GRABOWSKI, H., J. VERNON, AND J. A. DIMASI (2002): "Returns on Research and Developement for 1990s New Drug Introductions," *Pharmacoeconomics*, 20, 11–29.

HAILE, P., A. HORTAÇSU, AND G. KOSENOK (2008): "On the empirical content of quantal response equilibrium," *American Economic Review*, 98(1), 180–200.

HAJIVASSILIOU, V., D. MCFADDEN, AND P. RUUD (1996): "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives, Theorectical and Computational Results," *Journal of Econometrics*, 72, 85–134.

HITSCH, G. (2006): "An Empirical Model of Optimal Dynamic Product Launch and Exit Under Demand Uncertainty," *Marketing Science*, 25(1), 25–50.

HO, K. (2009): "Insurer-provider networks in the medical care market," *American Economic Review*, 99(1), 393–430.

HOTZ, J., AND R. MILLER (1993): "Conditional Choice Probabilties and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

HOWARD, R. A. (1960): *Dynamic Programming and Markov Processes.* Wiley.

HU, Y., AND M. SHUM (2008): "Nonparametric Identification of Dynamic Models with Unobserved State Variables," Working paper, Johns Hopkins University.

IMAI, S., N. JAIN, AND A. CHING (2005): "Bayesian Estimation of Dynamic Discrete Choice Models," working paper, Concordia University, Northern Illinois University and the University of Toronto.

ISHII, J. (2005): "Interconnection Pricing, Compatibility, and Investment in Network Industries: ATM Networks in the Banking Industry," working paper, Stanford University.

JAP, S., AND P. NAIK (2008): "Bid Analyzer: A Method for Price Discovery in Online Reverse Auctions," *Marketing Science*, 27(6), 949–960.

JIA, P. (2008): "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry," *Econometrica*, 76(6), 1263–1316.

KASAHARA, H., AND K. SHIMOTSU (2008): "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 77(1), 135–176.

KEANE, M., AND K. WOLPIN (1997): "The Career Decisions of Young Men," *The Journal of Political Economy*, 105(3), 473–522.

Keane, M. P. (1994): "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116.

Khwaja, A. (2001): "Health Insurance, Habits and Health Outcomes: A Dynamic Stochastic Model of Investment in Health," Doctoral dissertation, University of Minnesota.

Kuhn, M. (2006): "Notes on Numerical Dynamic Programming in Economic Applications," CDSEM, University of Mannheim, Denmark.

Magnac, T., and D. Thesmar (2002): "Identifying dynamic discrete decision processes," *Econometrica*, 70(2), 801–816.

Manuszak, M. D., and A. Cohen (2004): "Endogenous Market Structure with Discrete Product Differentiation and Multiple Equilibria: An Empirical Analysis of Competition Between Banks and Thrifts," Working paper.

Mazzeo, M. (2002): "Product Choice and Oligopoly Market Structure," *Rand Journal of Economics*, 33(2).

Naik, P., M. Mantrala, and A. Sawyer (1998): "Planning Media Schedules in the Presence of Dynamic Advertising Quality," *Marketing Science*, 17(3), 214–235.

Naik, P., K. Raman, and R. Winer (2005): "Planning Marketing-Mix Strategies in the Presence of Interactions," *Marketing Science*, 24(1), 25–34.

Norets, A. (2009): "Inference in dynamic discrete choice models with serially correlated unobserved state variables," *Econometrica*, 77(5), 1665–1682.

Orhun, Y. (2006): "Spatial Differentiation in the Supermarket Industry," Working paper, University of Chicago.

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)," *RAND Journal of Economics*, pp. 373–399.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2005): "Moment Inequalities and Their Application," working paper, Harvard University.

PARTHASARATHY, T. (1973): "Discounted, Positive, and Noncooperative Stochastic Games," *International Journal of Game Theory*, 2.

PESENDORFER, M., AND P. SCHMIDT-DENGLER (2003): "Identification and Estimation of Dynamic Games," NBER working paper No. w9726.

REIFFEN, D., AND M. WARD (2005): "Generic Drug Industry Dynamics," *Review of Economics and Statistics*, 87, 37–49.

RENY, P. J. (1999): "On the Existence of Pure and Mixed Strategy Nash Equilibria in Discountinuous Games," *Econometrica*, 67, 1029–1056.

ROBERTS, G. O., AND J. S. ROSENTHAL (2001): "Optimal Scaling for Various Metropolis-Hastings Algorithms," *Statistical Science*, 16, 351–367.

RUST, J. (1994): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle, and D. McFadden, pp. 3081–3143. North Holland.

———— (1996): "Numerical dynamic programming in economics," *Handbook of computational economics*, pp. 619–729.

RYAN, S. (2005): "The Costs of Environmental Regulation in a Concentrated Industry," Working paper, MIT.

RYSMAN, M. (2004): "Competition Between Networks: A Study of the Market for Yellow Pages," *The Review of Economic Studies*, 71(2), 483–512.

SCOTT-MORTON, F. (1999): "Entry Decisions into Generic Pharmaceutical Markets," *RAND Journal of Economics*.

SEIM, K. (2006): "An empirical model of firm entry with endogenous product-type choices," *RAND Journal of Economics*, pp. 619–640.

SHEN, Q. (2010): "A Dynamic Model of Entry and Exit in a Growing Industry," Working paper, Wharton School, University of Pennsylvania.

SHEN, Q., AND M. VILLAS-BOAS (2010): "Strategic Entry in Dynamic Markets," Working paper, Wharton School, University of Pennsylvania.

SPENCE, A. M. (1981): "The Learning Curve and Competition," *Bell Journal of Economics*, 12, 49–70.

SRIRAM, S., P. CHINTAGUNTA, AND R. NEELAMEGHAM (2006): "Effects of Brand Preference, Product Attributes, and Marketing Mix Variables in Technology Product Markets," *Marketing Science*, 25(5), 440–456.

SWEETING, A. (2006): "Coordination, Differentiation, and the Timing of Radio Commercials," *Journal of Economics and Management Strategy*, 15(4), 909–942.

TAMER, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70(1).

VITORINO, M. A. (2008): "Empirical Entry Games with Complementarities: An Application to the Shopping Center Industry," Working paper, University of Chicago.

WRIGHT, T. (1936): "Factors Affecting the Cost of Airplanes," *Journal of Aeronautical Sciences*, 3(4), 122–128.

XU, D. (2008): "A Structural Empirical Model of R&D, Firm Heterogeneity and Industry Evolution," Working Paper, New York University.

ZHU, T., AND V. SINGH (2006): "Spatial Competition with Endogenous Location Choices: An Application to Discount Retailing," forthcoming in Quantitative Marketing and Economics.

**Table 1.** Data

| Drug / Active Ingredient | ANDA Date | Dominant Firms (enter = 1, not enter = 0) | | | | Total Entrants | Revenue ($'000s) |
|---|---|---|---|---|---|---|---|
| | | Mylan | Novopharm | Lemmon | Geneva | | |
| Sulindac | 03 Apr. 90 | 1 | 0 | 1 | 1 | 7 | 189010 |
| Erythromycin Stearate | 15 May 90 | 0 | 0 | 0 | 0 | 1 | 13997 |
| Atenolol | 31 May 90 | 1 | 0 | 0 | 0 | 4 | 69802 |
| Nifedipine | 04 Jul. 90 | 0 | 1 | 0 | 0 | 5 | 302983 |
| Minocycline Hydrochloride | 14 Aug. 90 | 0 | 0 | 0 | 0 | 3 | 55491 |
| Methotrexate Sodium | 15 Oct. 90 | 1 | 0 | 0 | 0 | 3 | 24848 |
| Pyridostigmine Bromide | 27 Nov. 90 | 0 | 0 | 0 | 0 | 1 | 2113 |
| Estropipate | 27 Feb. 91 | 0 | 0 | 0 | 0 | 2 | 6820 |
| Loperamide Hydrochloride | 30 Aug. 91 | 1 | 1 | 1 | 1 | 5 | 31713 |
| Phendimetrazine | 30 Oct. 91 | 0 | 0 | 0 | 0 | 1 | 1269 |
| Tolmetin Sodium | 27 Nov. 91 | 1 | 1 | 1 | 1 | 7 | 59108 |
| Clemastine Fumarate | 31 Jan. 92 | 0 | 0 | 1 | 0 | 1 | 9077 |
| Cinoxacin | 28 Feb. 92 | 0 | 0 | 0 | 0 | 1 | 6281 |
| Diltiazem Hydrochloride | 30 Mar. 92 | 1 | 1 | 0 | 0 | 5 | 439125 |
| Nortriptyline Hydrochloride | 30 Mar. 92 | 1 | 0 | 0 | 1 | 3 | 187683 |
| Triamterene | 30 Apr. 92 | 0 | 0 | 0 | 1 | 2 | 22092 |
| Piroxicam | 29 May 92 | 1 | 1 | 1 | 0 | 9 | 309756 |
| Griseofulvin Ultramicrocrystalline | 30 Jun. 92 | 0 | 0 | 0 | 0 | 1 | 11727 |
| Pyrazinamide | 30 Jun. 92 | 0 | 0 | 0 | 0 | 1 | 306 |
| Diflunisal | 31 Jul. 92 | 0 | 0 | 1 | 0 | 2 | 96488 |
| Carbidopa | 28 Aug. 92 | 0 | 0 | 1 | 0 | 4 | 117233 |
| Pindolol | 03 Sep. 92 | 1 | 1 | 0 | 1 | 7 | 37648 |
| Ketoprofen | 22 Dec. 92 | 0 | 0 | 0 | 0 | 2 | 107047 |
| Gemfibrozil | 25 Jan. 93 | 1 | 0 | 1 | 0 | 5 | 330539 |
| Benzonatate | 29 Jan. 93 | 0 | 0 | 0 | 0 | 1 | 2597 |
| Methadone Hydrochloride | 15 Apr. 93 | 0 | 0 | 0 | 0 | 1 | 1858 |
| Methazolamide | 30 Jun. 93 | 0 | 0 | 0 | 1 | 3 | 4792 |
| Alprazolam | 19 Oct. 93 | 1 | 1 | 0 | 0 | 7 | 614593 |
| Nadolol | 31 Oct. 93 | 1 | 0 | 0 | 0 | 2 | 125379 |
| Levonorgestrel | 13 Dec. 93 | 0 | 0 | 0 | 0 | 1 | 47836 |
| Metoprolol Tartrate | 21 Dec. 93 | 1 | 1 | 0 | 1 | 9 | 235625 |
| Naproxen | 21 Dec. 93 | 1 | 1 | 1 | 1 | 8 | 456191 |
| Naproxen Sodium | 21 Dec. 93 | 1 | 1 | 1 | 1 | 7 | 164771 |
| Guanabenz Acetate | 28 Feb. 94 | 0 | 0 | 0 | 0 | 2 | 18120 |
| Triazolam | 25 Mar. 94 | 0 | 0 | 0 | 0 | 2 | 71282 |
| Glipizide | 10 May 94 | 1 | 0 | 0 | 0 | 1 | 189717 |
| Cimetidine | 17 May 94 | 1 | 1 | 0 | 0 | 3 | 547218 |
| Flurbiprofen | 20 Jun. 94 | 1 | 0 | 0 | 0 | 1 | 155329 |
| Sulfadiazine | 29 Jul. 94 | 0 | 0 | 0 | 0 | 1 | 72 |
| Hydroxychloroquine Sulfate | 30 Sep. 94 | 0 | 0 | 0 | 0 | 1 | 8492 |
| Mean | | 0.45 | 0.28 | 0.25 | 0.25 | 3.3 | 126901 |

Shown is 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: Boundedly Rational Model

| Parameter | Number of Potential Entrants (excluding "other" firms) | |
| | 3 firms | 4 firms |
|---|---|---|
| $\mu_c$ | 10.05 (0.017) | 10.07 (0.0014) |
| $\rho_c$ | 0.9866 (0.00086) | 0.9873 (5.6e-05) |
| $\sigma_c$ | 0.3721 (0.026) | 0.3675 (3.0e-04) |
| $\kappa_c$ | 0.06655 (0.0015) | 0.07067 (1.1e-04) |
| $\mu_r$ | 9.906 (0.083) | 10.008 (0.0037) |
| $\sigma_r$ | 1.591 (0.060) | 1.682 (0.0023) |
| $\gamma$ | 0.9375 | 0.9375 |
| $\beta$ | 0.9688 | 0.9688 |
| $p_a$ | 0.9375 | 0.9375 |
| CER firm 1 | 0.0857 | 0.1208 |
| CER firm 2 | 0.0788 | 0.0876 |
| CER firm 3 | 0.1038 | 0.1061 |
| CER firm 4 | | 0.1374 |
| CER all firms | 0.0894 | 0.1130 |
| 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 tendency 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 $A_{i,t,j}$ are known for each firm $i$ at each time $t$ for each particle $j$ and can be compared to the observed actions $A_{it}^o$. The CER is the the proportion of the cases where $A_{it}^o \neq A_{i,t,j}$ computed both by firm and overall.

**Table 3.** Posterior Distribution: Fully Rational Model

| Parameter | Number of Potential Entrants (excluding "other" firms) | |
| | 3 firms | 4 firms |
| --- | --- | --- |
| $\mu_c$ | 10.06 (0.016) | 10.07 (0.0012) |
| $\rho_c$ | 0.9877 (0.00095) | 0.9873 (4.5e-05) |
| $\sigma_c$ | 0.3714 (0.037) | 0.3676 (0.00035) |
| $\kappa_c$ | 0.06634 (0.0012) | 0.0704 (0.00010) |
| $\mu_r$ | 9.941 (0.064) | 10.009 (0.0032) |
| $\sigma_r$ | 1.531 (0.059) | 1.682 (0.0016) |
| $\gamma$ | 0.9375 | 0.9375 |
| $\beta$ | 0.9688 | 0.9688 |
| $p_a$ | 0.9375 | 0.9375 |
| CER firm 1 | 0.0861 | 0.1045 |
| CER firm 2 | 0.0786 | 0.0837 |
| CER firm 3 | 0.0971 | 0.0860 |
| CER firm 4 | | 0.1351 |
| CER all firms | 0.0873 | 0.1023 |
| 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 tendency 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 $A_{i,t,j}$ are known for each firm $i$ at each time $t$ for each particle $j$ and can be compared to the observed actions $A_{it}^o$. The CER is the the proportion of the cases where $A_{it}^o \neq A_{i,t,j}$ computed both by firm and overall.
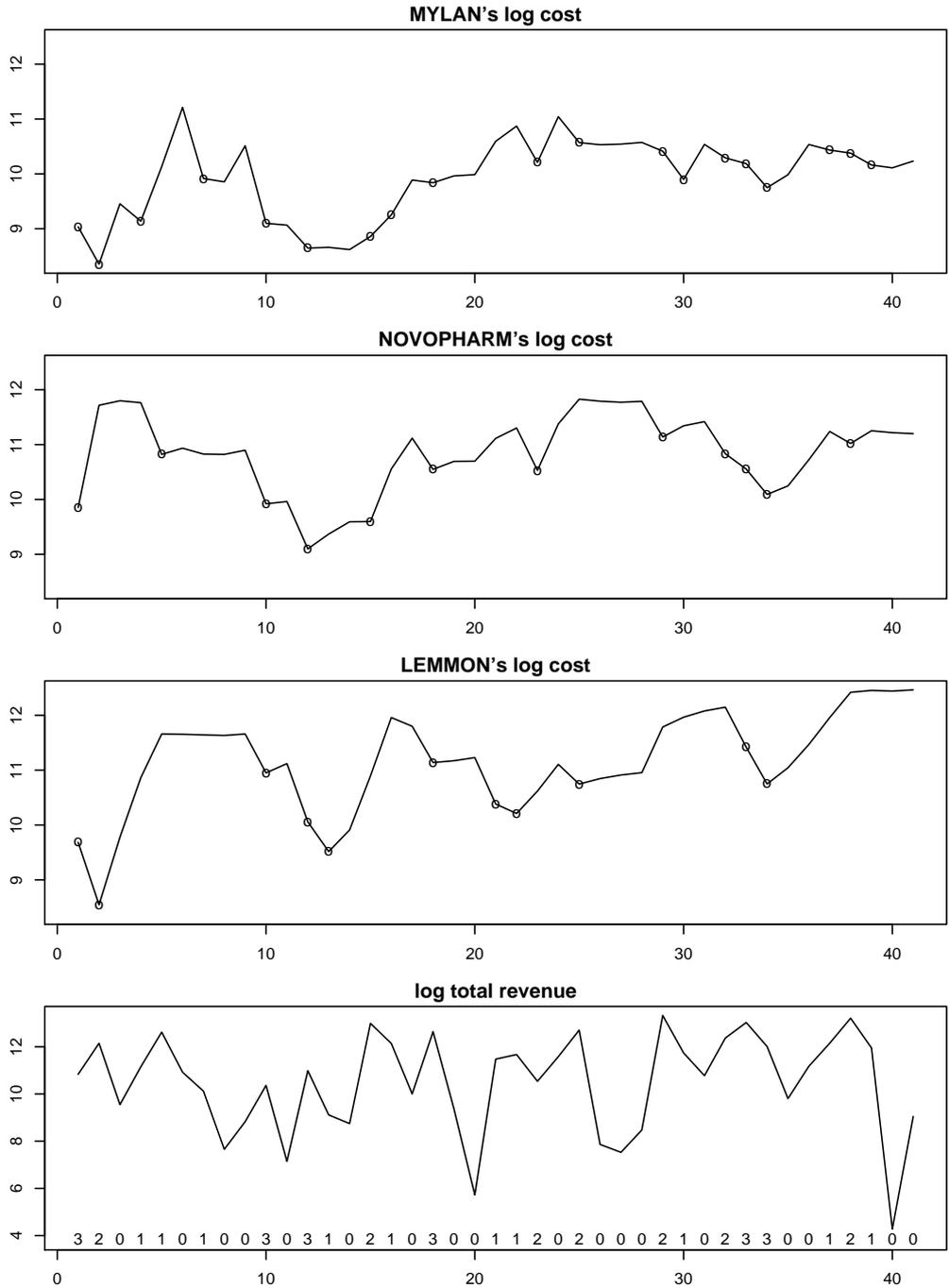
**Figure 1. Cost, Revenue, and Entry Decisions.** Plotted as a solid line in the first three panels is the logarithm of cost for the three dominant firms in the three firm model. The logarithm of cost is computed by averaging at Step 2e of the importance sampler at the maximum likelihood estimate. The 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.
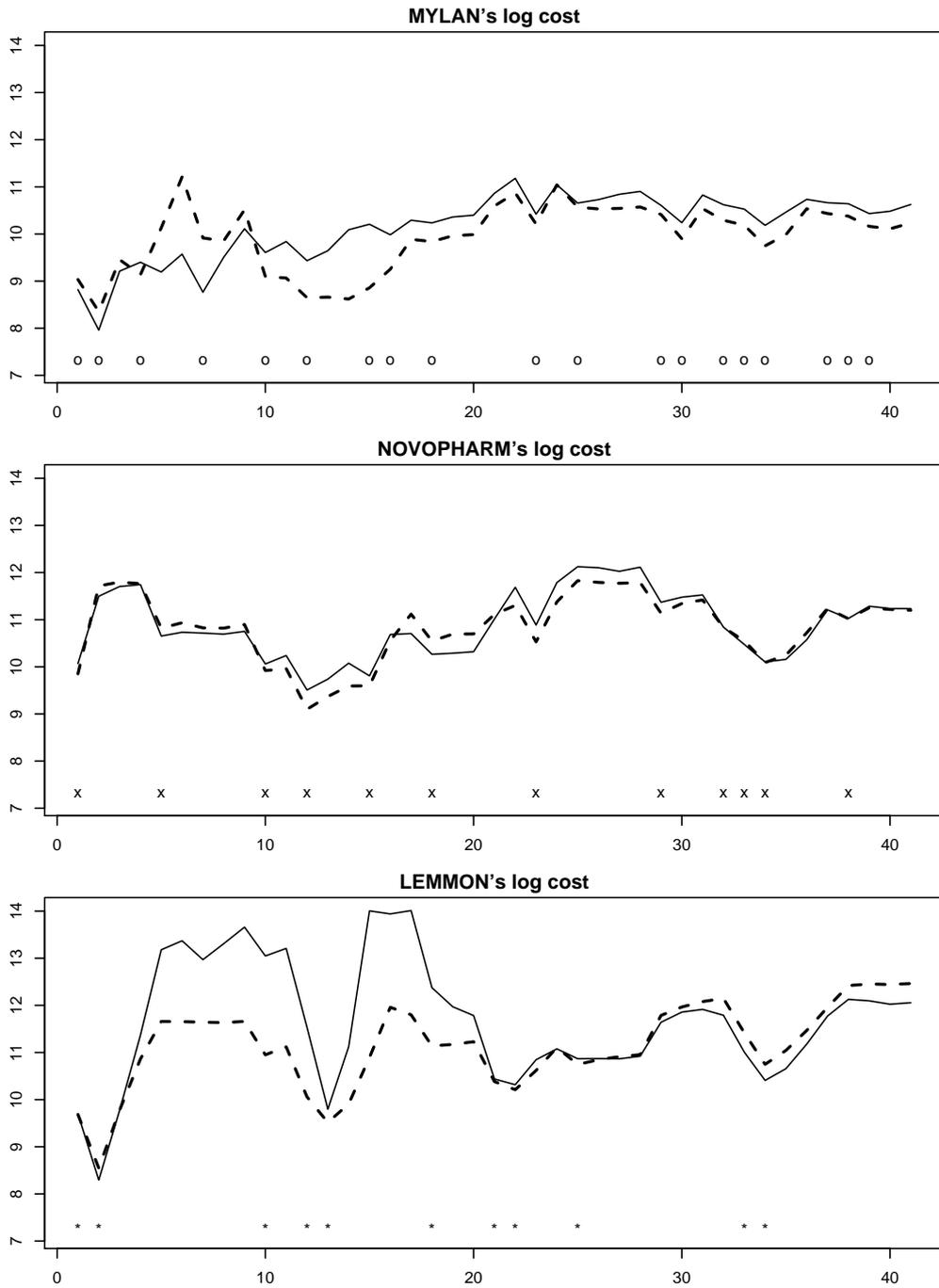
**Figure 2. Cost and Entry Decisions of the Dominant Firms.** Plotted is the logarithm of cost for the three dominant firms. The dashed line is under the three firm model, and the solid under the four firm model. The circles indicate the markets that Mylan entered, crosses the same for Novopharm, and the asterisks for Lemmon. The logarithm of cost as described in the legend of Figure 1.
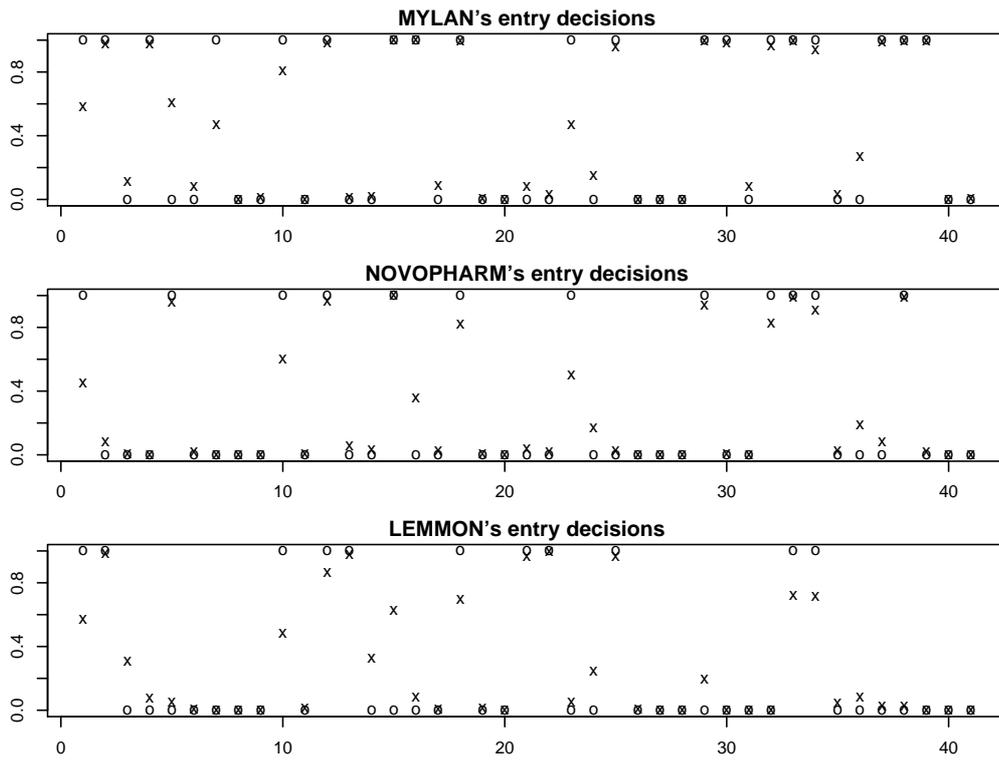
44

**Figure 3. 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.

# A    Technical Appendix

In this Technical Appendix 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. These details were not provided in the main paper in the interest of brevity. Since we use an infinite horizon model we look 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 for a market opening at time $t$, i.e., the strategy profile of the dynamic game, be denoted by

$$A_t = (A_{1t}, \ldots, A_{It}). \tag{26}$$

As discussed in Section 3, 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

$$s_t = (c_{1t}, \ldots, c_{It}, r_t). \tag{27}$$

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 $R_t = \exp(s_{I+1,t})$. We describe the solution algorithm for a given parameter vector 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 $K$. A state $s_t$ within hyper-cube can be mapped to its key $K$.[28] 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 represented by the affine function $V_K(s_t) = b_K + (B_K)s_t$.[29] A value function $V_K(s_t)$ whose elements satisfy equation (12) is denoted $V_K^*(s_t) = b_K^* + (B_K^*)s_t$.

---

[28] Grid increments are chosen to be fractional powers of two so that the key has an exact machine representation. This facilitates efficient computation through 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 entire grid itself is never computed because all we require is the mapping $s \mapsto K$, which is determined by the increments. The end points of a hyper-cube on the grid in order to find the appropriate key $K$ are computed as needed.

[29] Keane and Wolpin (1997) adopt a similar approach for a single agent dynamic programming model. Our approach differs from Keane and Wolpin (1997) in that we let the coefficients of the regression depend on

The game is 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).[30]

2. Check whether the fixed point $V_K^*(s)$ of the Bellman equations (12) at this key has already been computed, i.e., whether the $(b_K^*, B_K^*)$ for the $K$ that corresponds to $s$ has 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 is represented by the coefficients $(b_K^{(0)}, B_K^{(0)})$ being set to 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 the equilibrium condition in equation (10) is satisfied. In this process we need to take expectations to compute the continuation value $\beta \mathcal{E}\left[V_{K,i}^{(0)}(s_{t+1}) \mid A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t, \right]$ that appears in equation (9), where we have used equation (11) to express equation (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 centered around each $s_j$, i.e., $s_{j_l}$, $l = 1, \ldots, L$. 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 value function evaluated at the abscissae (more details are provided below).

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 local linear function.

[30]In fact, because it is a stationary game, the subscript $t$ does not really matter

6. We can now compute the continuation value at $s_j$ for each candidate strategy $A$. We compute the best response strategy profile $A_j^E$ corresponding to $s_j$ by checking the Nash equilibrium condition (equation 10). As just described, the choice specific value function evaluated at $(A_i^E, s_j)$ is computed using $V_K^{(0)}(s)$ and equation (9), and denoted by $V_K^{(1)}(A^E, s_j) = (V_{1,K}^{(1)}(A^E, s_j), \ldots, V_{I,K}^{(1)}(A^E, s_j))$.

7. Next we use the "data" $(V_K^{(1)}(A^E, s_j), s_j)_{j=1}^J$ to update the ex ante value function to $V_K^{(1)}(s_j)$. This is done 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).[31]

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 $A^E$ for each $s_j$ until convergence is achieved for the coefficients $(b_K^{(0)}, B_K^{(0)})$, $(b_K^{(1)}, B_K^{(1)})$, $\ldots$, $(b_K^{(*)}, B_K^{(*)})$. 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 a conjecture $(b_K^{(l)} = 0, B_K^{(l)} = 0)$ for the linear approximation of the value functions at a given state at iteration $l = 0$. These guesses are then used in computing the choice specific value functions at iteration $l + 1$ using equation (9). This computation involves taking expectations over the conditional distribution of the future state variables, which is accomplished using Gaussian-Hermite quadrature. Once we have the choice specific value functions we compute the equilibrium strategy profile at iteration $l+1$ using equation (10). The best response strategy profile at iteration $l + 1$ is then used to compute the iteration $l + 1$ ex ante value functions via a regression that can be viewed as iterating equation (12). The iteration $l + 1$ ex ante value functions are then used to compute the iteration $l + 2$ choice specific value functions using equation (9), and the entire procedure is repeated till a fixed point of equation (12) is obtained. This iterative procedure solves the dynamic game. 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 $\mu_{I+1} =$

---

[31]$V_K^{(1)}(A^E, s_j)$ will not equal $V_K^{(1)}(s_j)$ because the former is "data" and the later is a regression prediction.

$\mu_R$ for the last element (see equation 6), standard deviations $\sigma_i = \sigma_c$ for the first $I$ elements, and standard deviation $\sigma_{I+1} = \sigma_R$ 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 done by a Gauss-Hermite quadrature rule that has been subjected to location and scale transformations. The weights $w_j$ and abscissae $x_j$ 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 $s_{j,t+1}$ conditional upon $A_t$ and $s_t$ the abscissae are transformed to $\tilde{s}_{t+1,j} = \mu_j + \sqrt{2}\sigma_j x_j$, and the weights are transformed to $\tilde{w}_j = w_j/\sqrt{\pi}$, where $\pi = 3.142$.[32] Then, using a $2L + 1$ rule,

$$\mathcal{E}[f(s_{t+1}) \,|\, A_t, s_t] \approx \sum_{j_1=-L}^{L} \cdots \sum_{j_I=-L}^{L} \sum_{j_{I+1}=-L}^{L} f(\tilde{s}_{t+1,j_1}, \cdots, \tilde{s}_{t+1,j_I}, \tilde{s}_{t+1,j_{I+1}}) \tilde{w}_{j_1} \cdots \tilde{w}_{j_I} \tilde{w}_{j_{I+1}}. \quad (28)$$

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

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

We use three point rules throughout. A three point rule will integrate a polynomial in $s_{t+1}$ up to degree five exactly.[33]

Step 7 involves updating the ex ante value function using a regression. We next describe how we do this. As stated above, we have a grid over the state space whose boundaries are fractional powers of two over the state space.[34] We approximate the value function $V(s_t)$ by a locally indexed affine representation as described above. For the grid increments that determine the index of hyper-cubes we tried a range of values from 4 to 16 times the

---

[32]These transformations arise because a Hermite rule integrates $\int_{-\infty}^{\infty} f(x) \exp(-x^2) dx$. Hence we need to do a change of variables to get our integral $\int_{-\infty}^{\infty} g(\sigma z + \mu)(1/\sqrt{2\pi}) \exp(-0.5z^2) dz$ to be of that form. A change of variables puts the equation in the line above in the form $\int_{-\infty}^{\infty} g(\sqrt{2}\sigma x + \mu)(1/\sqrt{\pi}) \exp(-x^2) dx$, which is where the expressions for $\tilde{s}_{t+1,i}$ and $\tilde{w}_i$ come from.

[33]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 abscissae of the quadrature rule so that crossing a boundary will be a rare event, if it happens at all.

[34]Recall that grid increments are chosen to be fractional powers of two so that the key has an exact machine representation. This facilitates efficient computation through compact storage of objects indexed by the key.

standard deviation of the state variables rounded to a nearby fractional power of two to scale the grid appropriately. The results are effectively the same. Hence in estimating the model we set the grid increments at 16 times the standard deviation of the state variables.[35] We compute the coefficients $b_K$ and $B_K$ as follows. They are first initialized to zero. We then generate a set of abscissae $\{s_j\}$ clustered about $K$ and solve the game with payoffs (9) to get corresponding equilibria $\{A_j^E\}$. We substitute the $(A_j^E, s_j)$ pairs into equation (9) to get $\{V(A_j^E, s_j)\}_{j=1}^J$. Using the pairs $\{(V(A_j^E, s_j), s_j\}$ 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 approximately twenty iterations suffice for three firms and thirty for four firms.[36] 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 jiggle the points so that no two firms have exactly the same cost (see next paragraph for the reason for this). Of importance in reducing computational effort is to avoid recomputing the payoff (equation (9)) when checking equilibrium condition (10). Our strategy is to (temporarily) store payoff vectors indexed by $A$ and check for previously computed payoffs before computing new ones in checking condition (10).

---

[35] The set of keys that actually get visited in any MCMC repetition is about the same for grid increments ranging from 4 to 16 times the standard deviation of the state variables in our data. For a three firm game the number of hyper-cubes that actually are visited in any one repetition is about six.

[36] An alternative is to apply a modified Howard acceleration strategy as described in Kuhn (2006); see also Rust (1996) and Howard (1960). The idea is simple: The solution $\{A_t^E\}$ 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.

# B    Web Appendix: Additional Details on Estimation

**This is a companion Web Appendix to "Dynamic Entry with Cross Product Spillovers: An Application to the Generic Drug Industry."**

In this appendix we provide additional details about the estimation procedure that were not provided in the main paper to keep the paper of a reasonable length. We begin by showing the log likelihood surface plotted on a fine grid in Figure 4 and on a coarse grid in Figure 5 for the three firm model. Figures 6 and 7 are for the four firm model. The endpoints of the horizontal axes are tenth of a standard deviation to the left and right of the maximum in Figure 4 and 24 standard deviations to the right and left in Figure 5. For Figures 6 and 7 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 4 and 6, 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 do not 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.

Figure 4 about here

Figure 5 about here

Figure 6 about here

Figure 7 about here

However, Figures 5 and 7 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 5 and 7; see Gamerman and Lopes (2006). However this comes at a cost 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. The vertical lines in Figures 5 and 7 indicate the range of the MCMC chain's excursions after the transient elements of the chain have died out.

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 Chernozhukov 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. 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.
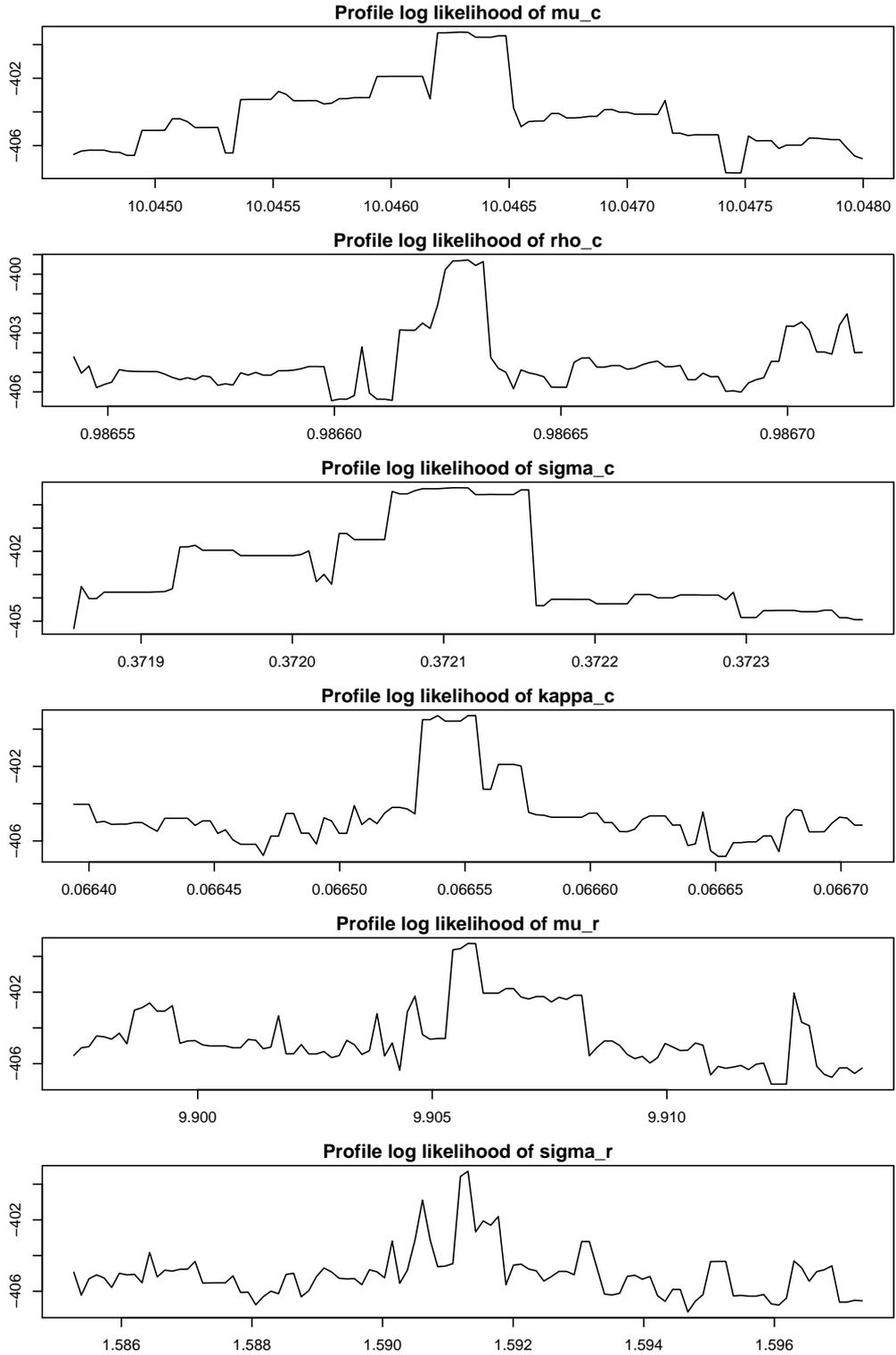
**Figure 4. Profile Log Likelihood, Three Firm Model.** Shown is the logarithm of the profile likelihood plotted for a tenth of the posterior standard deviation to the left and right of the maximum of the likelihood.
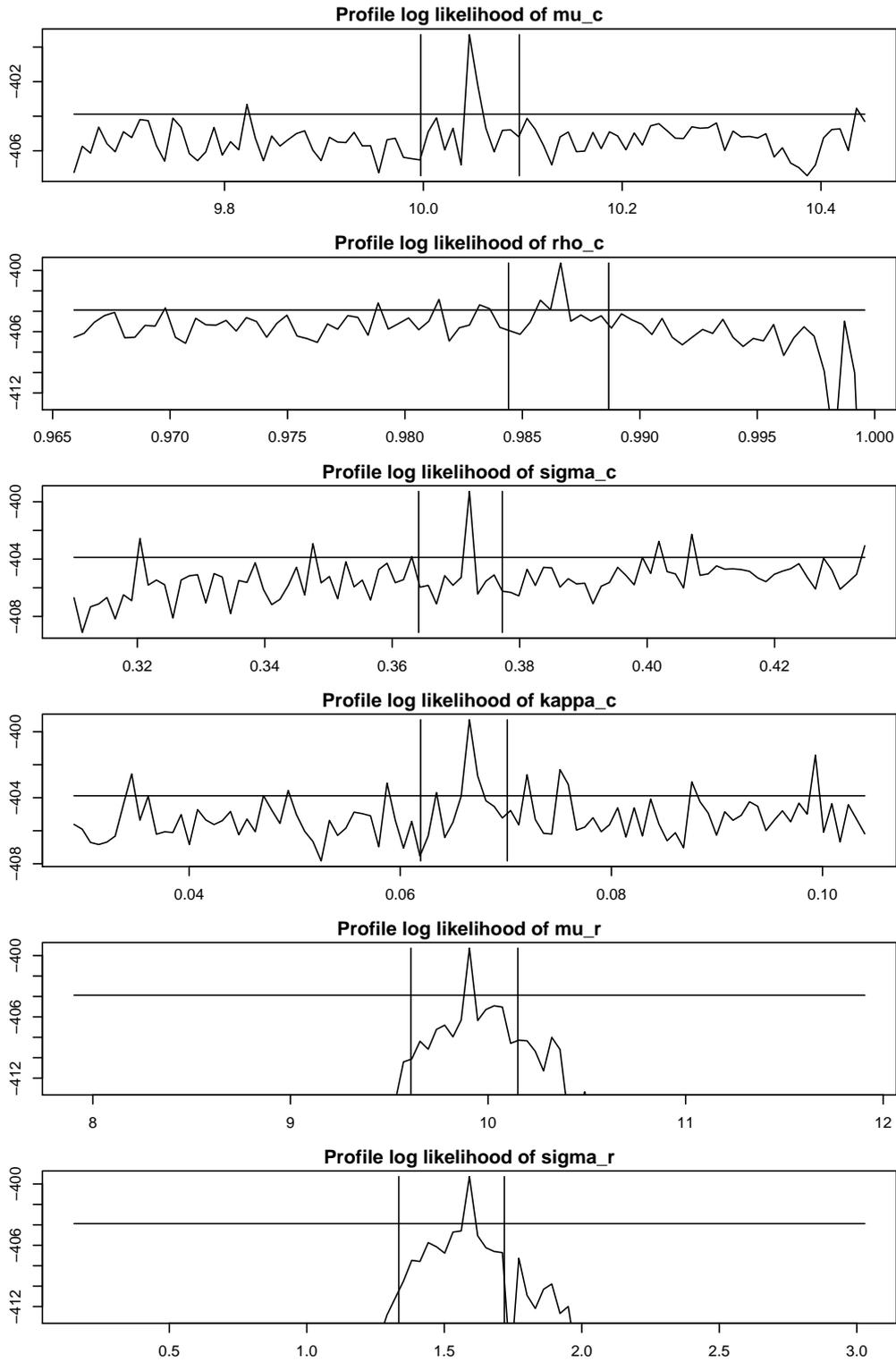
53

**Figure 5. 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 the transient elements of the chain have died out.
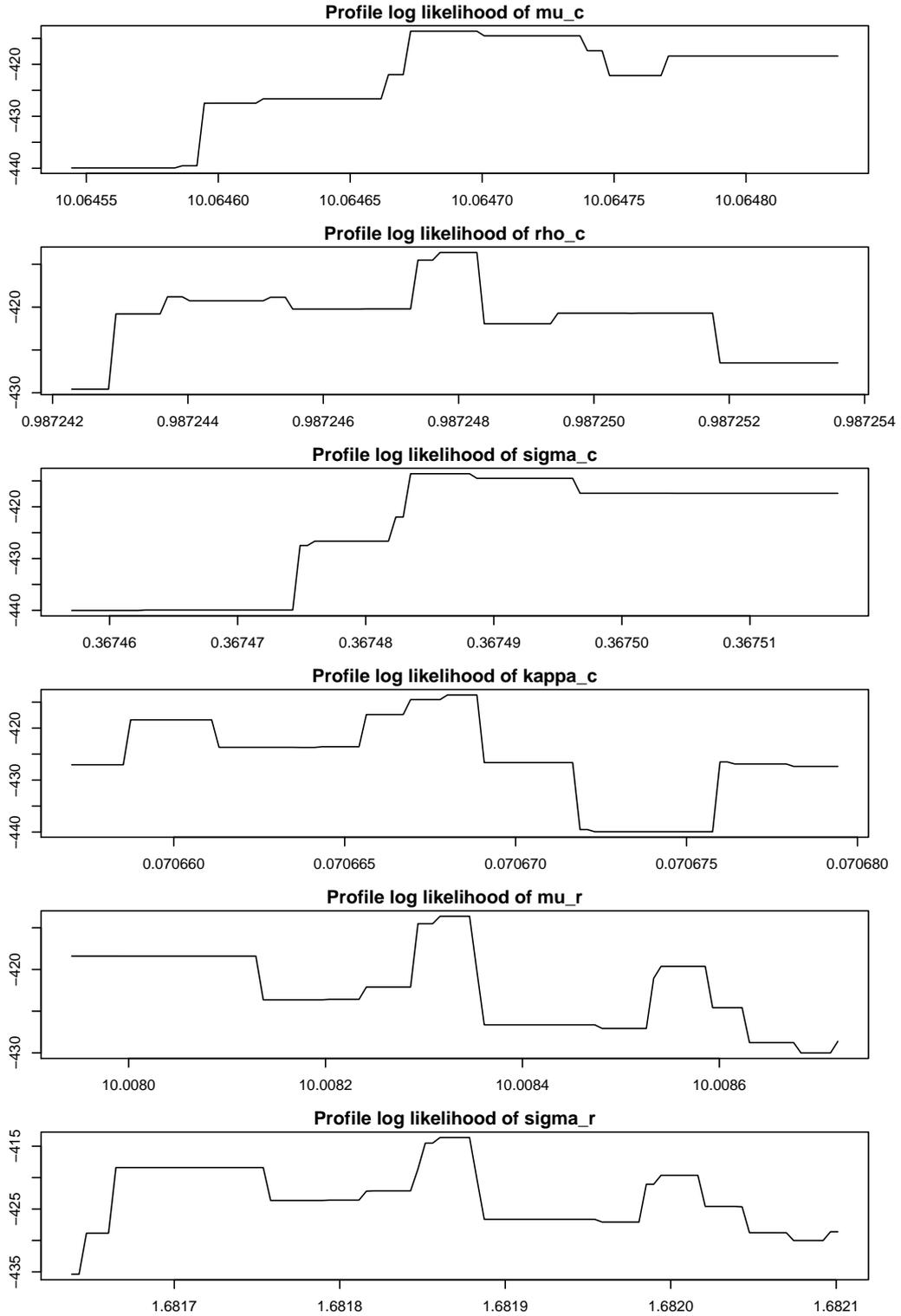
**Figure 6. Profile Log Likelihood, Four Firm Model.** Shown is the logarithm of the profile likelihood plotted for a tenth of a posterior standard deviations to the left and right of the maximum of the likelihood.
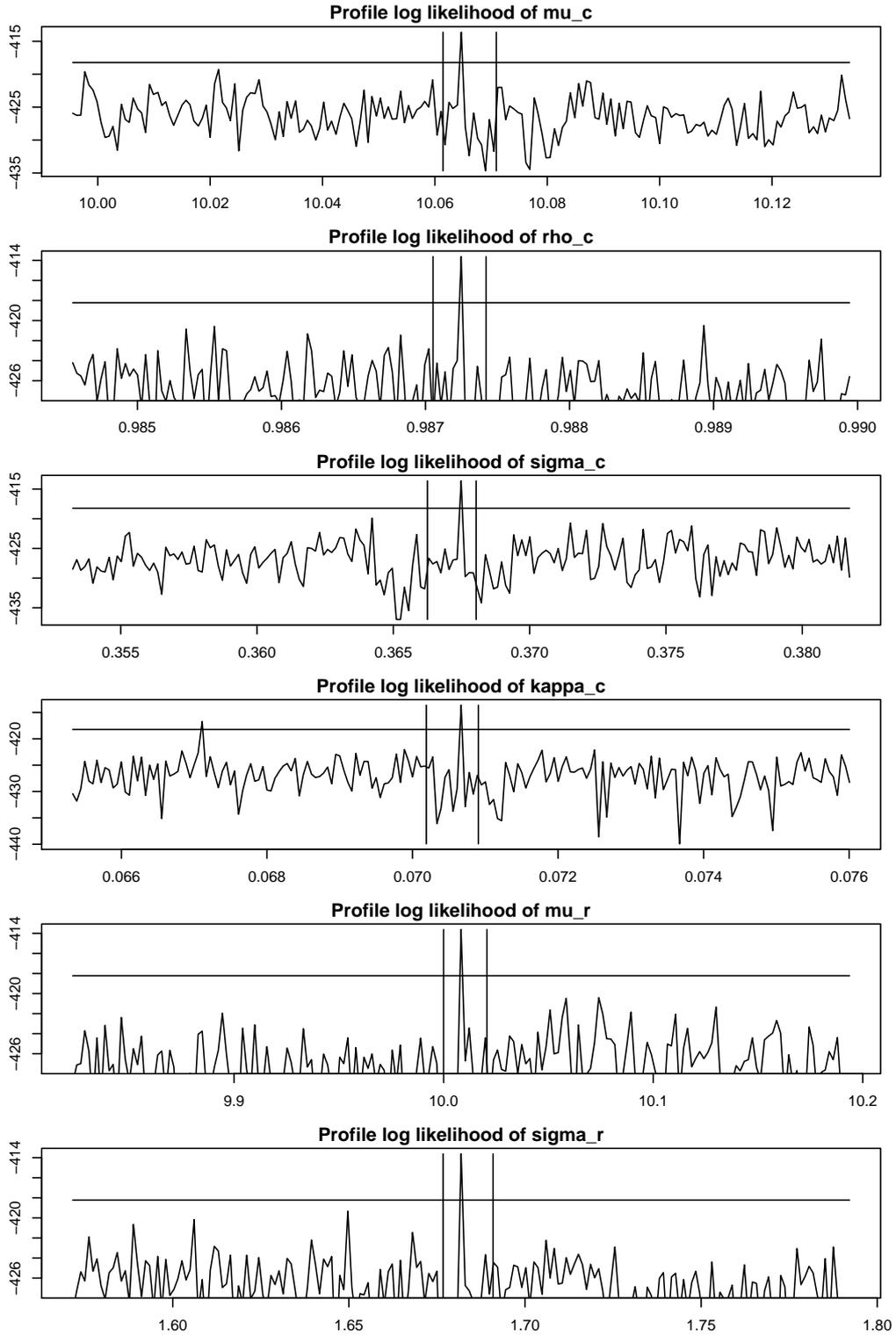
55

**Figure 7. 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 the transient elements of the chain have died out.

56