The Dynamic Spillovers of Entry:
An Application to the Generic Drug Industry*

A. Ronald Gallant
Penn State University

Han Hong
Stanford University

Ahmed Khwaja
Yale University

September 2015

Abstract

This paper examines if experience from entry in one market can potentially enhance profitability at a future market opportunity for a related product. We formulate and estimate a dynamic game of entry in which forward looking firms make decisions not just based on present benefits of past entry but also anticipating potential future benefits of current entry. Dynamic spillovers of entry are incorporated through a firm specific unobservable (to the researcher) cost that depends on past entry decisions. The unobserved costs may also be serially persistent. Thus, the model allows for firm specific unobserved heterogeneity that evolves based on firm actions. The challenge of estimating a dynamic game with serially correlated unobserved state variables subject to endogenous feedback is overcome by embedding a particle filter based technique in a Nested Fixed Point Algorithm. Using an application to a stylized model of entry in the generic pharmaceutical industry we underscore the motivation for the model specification and the methodology developed. Our estimates imply positive spillover effects of entry. Moreover, these spillovers suggest heterogeneity not just across firms but also within firms over time based on their history of entry decisions. Our results illustrate that entry may potentially provide firms with additional strategic advantage in later markets, and that entry spillovers may be an important factor to consider in the equilibrium evolution of the generic drug industry.

Keywords: Dynamic Discrete Games, Particle Filter, Unobserved State Variables, Endogenous Feedback, Serial Correlation, Dynamic Entry, Entry Spillovers, Generic Pharmaceuticals.

JEL Classification: E00, G12, C51, C52

1　Introduction

One of the most significant decisions with long term implications for itself and its rivals that 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 on profitability in related future markets, i.e., experience gained from entering a market could enhance profitability when entering markets for related products in the future. In the presence of dynamic spillovers a firm with a long term perspective may even enter a market when the current opportunity is unprofitable. This can happen if the entry experience leads to greater cumulative profitability from the 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 related product markets.[1]

In this paper we analyze the dynamic spillover effects of entry decisions of firms on future market performance. In order to do this we formulate and estimate a dynamic game that incorporates spillover effects of current entry on future latent (to the researcher) costs, and consequently future entry decisions. Hence, spillovers imply that heterogeneity among firms arises based on the history of their actions.[2] Given the potential spillovers of entry on latent costs, the model needs to accommodate unobserved state variables that are endogenous to firm actions and possibly serially correlated. We address the methodological challenge of estimating such a model using a particle filter 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 (see e.g., Aguirregabiria and Mira (2009)). We illustrate our method with an application to the generic pharmaceutical industry.

In contrast to static entry models, our approach allows for, and distinguishes between the dynamic long run and static short run incentives faced by forward looking strategic firms. It is particularly important to incorporate both these incentives because the the entry threshold in the

---

[1] Consider for example the decision of Rupert Murdoch's (then nascent) Fox network to enter the market for broadcasting NFL football games on TV in 1993. At that time it was estimated by industry analysts that Fox would lose between $500 and $700 million on its bid if it were successful in acquiring NFL broadcasting rights. The conjectured rationale for this entry decision was that it helped Fox build a reputation and loyalty. As an industry insider explained "Fox needed to establish that it was willing and able to compete with the three established networks...Football gave them instant credibility in that area. It was through sports that a lot of viewers got their first glimpses of *Melrose Place* or *The Simpsons* or *90120*. From that standpoint it was worth it. We might not have the *X-Files* if not for football on Fox (Anand and Conneely (2003a), p. 1)." Ex post, back-of-envelope calculations suggest that the bid price was fair and even conservative; in other words, the anticipated cross-product spillovers were large and relevant (Anand and Conneely (2003a; 2003b)). Another related example is in the M&A context of learning in alliances by firms that enter multiple ones (Anand and Khanna (2000)). In this case there are some firms that are serial acquirers with the strategy of entering a new sector by first acquiring a "platform" company with the explicit intention of expanding with "bolt-on" acquisitions.

[2] See Shen and Villas-Boas (2010) for a recent theoretical paper that also explores entry decisions of forward looking firms in dynamically evolving markets.

dynamic context is very different from that in a static context. In a dynamic setting, firms have an incentive to enter even if they make losses in the short run as long as entry raises the present discounted value of the stream of expected future profits in the long run. Such dynamic incentives are ignored by static models which require that a firm enter only if the current profits are non-negative. In the absence of incorporating such dynamic long run incentives it would not be possible to estimate the dynamic spillovers of entry if these existed.

Our research is related to the work of Keane (1994) that used particle filters or 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. A special case of that estimator is the well-known Geweke-Hajivassiliou-Keane (GHK) estimator (see e.g., Hajivassiliou, McFadden, and Ruud (1996), Geweke and Keane (2001)) that arises for a particular choice of importance sampling densities. Another strand of related research is the work on dynamic linear models that has employed Kalman-Filters (e.g., Naik, Raman, and Winer (2005), Sriram, Chintagunta, and Neelamegham (2006), Jap and Naik (2008), Bass, Bruce, Murthi, and Majumdar (2007), Bruce (2008)). The important difference is that Kalman-Filters require the dynamic relationships to be linear and error terms to be Gaussian whereas our paper allows for non-linear relationships and non-Gaussian error terms. The critical distinction is that the particle filter in our research embeds the solution to a dynamic game.[3] Moreover, the unobserved state variable (i.e., cost) allows for endogenous feedback in contrast to most particle filter models in which unobserved variables evolve exogenously.

There is a long history of research on spillovers of experience and broadly speaking learning curves in various contexts and disciplines, e.g., Ebbinghaus (1913), Wright (1936), Arrow (1962), Bass (1980) and Dolan and Jeuland (1981). However, it has been difficult to estimate spillovers in a dynamic oligopolistic context that explicitly accounts for strategic interaction between firms. An exception is the work of Benkard (2004) who incorporates learning by doing in a dynamic oligopoly model to examine such experience spillovers in the manufacture of aircraft, specifically, the Lockheed L-1011.

There is an extensive literature on estimating discrete games. Static games of entry and strategic firm interaction have been estimated among others by Bresnahan and Reiss (1991b), Bresnahan and Reiss (1991a), Berry (1992), Scott-Morton (1999), Mazzeo (2002), Seim (2006), Zhu and Singh (2009), Orhun (2013), Vitorino (2008), Datta and Sudhir (2009) and Ellickson and Misra (2008). In this literature the paper most closely related to ours is the work of Jia (2008) who examines entry spillovers across multiple markets in a static duopoly context.[4] Dynamic models of entry have been analyzed by, e.g., Hitsch (2006), Ching (2010a), and Shen (2010). However, to our knowledge there

---

[3]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.

[4]As Jia (2008) states, her paper, "is not applicable to oligopoly games with three or more chains (pp. 1265-1266)" because the model requires that one firm's gain is the other firm's loss, an assumption which cannot be satisfied in an oligopoly model like ours with more than two firms. The market chains in the context of Jia's model correspond to firms in our model.

is no empirical work on incorporating the dynamic spillovers of entry and the consequent serial correlation in the underlying pay-off relevant state variables in game theoretic oligopoly models.

Analysis of supply side decisions in the generic pharmaceutical firm industry has been a topic of considerable empirical research (e.g., Reiffen and Ward (2005), Ching (2010a)). The benchmark paper in modeling entry in to multiple markets remains Scott-Morton (1999) that estimates 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.[5] Our work extends this research along three significant dimensions. First, we recognize linkages across different markets because entry into one market may provide an advantage in later markets. Second, if such linkages exist then firms may anticipate such advantages and incorporate this into their decision making, i.e., make decisions not just based on present benefits of past entry but also anticipating future benefits of current entry. Third, we allow for firms to be asymmetric and their unobserved type to evolve over time. A recent paper by Amisano and Giorgetti (2013) also analyzes the dynamics of entry by pharmaceutical firms. Using a "reduced-form discrete-choice panel data framework (p. 667)" they estimate a dynamic probit model to examine the determinants of entry such as firm size, intensity of competition, demand conditions, and proxies for sunk cost in multiple markets. Like our model the paper allows for linkages between past and current entry, and across the markets for different products. However, an important distinction is that in contrast with our work the strategic nature of repeated and dynamic oligopolistic competition between forward looking rival firms in not explicitly incorporated in the model.

Our research also contributes to the wider literature on the estimation of games. There is a growing literature on estimating static[6] and dynamic[7] games of *incomplete* information. This literature 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 rules out unobserved dynamics in the latent state variables. It also rules out any private information

---

[5]Although, specific examples are hard to find (presumably due to firms not wanting to publicize these for proprietary reasons), there exists corroborating anecdotal evidence in the empirical literature on the existence of spillovers in the pharmaceutical industry (e.g., Scott-Morton (1999), Amisano and Giorgetti (2013), Conti and Berndt (2013)). Such spillovers may arise on the supply side due to fixed costs of investments that may have to be made as part of the regulatory approval process. For example, some drugs are controlled substances that require a large and sophisticated security infrastructure to be in place before a firm is allowed to manufacture the drug. However, once this infrastructure is in place it can be used in the manufacture of any other drug that is classified as a controlled substance. Another example is liquid active pharmaceutical ingredients (APIs) that have to be sterilized using filtration. The packaging of these into a final "fill and finish" product requires aseptic and automated filling in to vials and syringes by keeping human intervention to a minimum. Such automated aseptic "fill and finish" infrastructure is a fixed cost but can be used across various liquid APIs once in place. "Specialized expertise and automated, aseptic fill and finish procedures likely provide ANDA holders significant cost advantages in injectable drug production due to economies of scope (Conti and Berndt, p. 5, 2013)."

[6]See e.g., Rysman (2004), and Bajari, Hong, Krainer, and Nekipelov (2010).

[7]See e.g., Aguirregabiria and Mira (2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007), Pesendorfer and Schmidt-Dengler (2010), Bajari, Chernozhukov, Hong, and Nekipelov (2007), and Shen (2010).

that a player might have about competing firms that the researcher does not have. Aguirregabiria and Mira (2007) and Arcidiacono and Miller (2011) 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. However, these methods are are tractable only for market level unobserved heterogeneity. In contrast, our method is applicable even when the unobserved variable is continuous, and specific to the agent and not just at the market level. Furthermore, our method allows for the unobserved variable to be endogenous to past actions, i.e., endogenous feedback. As far as we know, games of *complete* information have only been estimated in a static context, e.g., Tamer (2003), Ciliberto and Tamer (2009) and Bajari, Hong, and Ryan (2010). The complete information assumption allows substantial unobserved heterogeneity at the level of the firm. 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. We believe 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 well-developed literature that allows for time invariant unobserved heterogeneity, e.g., Keane and Wolpin (1997), and a smaller literature that allows for serially correlated unobserved *endogenous* state variables (e.g., Erdem and Keane (1996), Khwaja (2010)). A Bayesian approach for single agent discrete dynamic programming models with unobserved state variables that are serially correlated over time has been developed by Imai, Jain, and Ching (2009), and extended by Norets (2009). These papers use MCMC for integrating out the unobserved state variables. In contrast, we use particle filters (or 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 a substantial computational burden in solving and estimating the model.

In related work (Gallant, Hong, and Khwaja (2014)), we address several issues not analyzed in this paper. We build on the current paper to examine theoretical properties of a Bayesian approach using a particle filter to estimate dynamic models with small and large number of players. More specifically, we derive an unbiased estimate of the likelihood within a MCMC algorithm which verifies that the method is exact for Bayesian inference. A key variation in the MCMC algorithm in Gallant, Hong, and Khwaja (2014) is that at each repetition of the chain the initial seed that determines all random draws within the filter is proposed along with the parameter values. This has the computational advantage that it reduces the number of simulation draws that are required in order to evaluate the likelihood. The algorithm in Gallant, Hong, and Khwaja (2014) is closer to the work of Flury and Shephard (2010). In contrast, these aspects are not present in this paper, which uses instead an algorithm similar to that also found in Fernandez-Villaverde and Rubio-Ramirez (2005). In addition, Gallant, Hong, and Khwaja (2014) shows that the algorithm

developed works not just for dynamic entry models but also for Ericson and Pakes (1995) style models. In particular, when used in a setting with a large number of players with the solution concept of Oblivious Equilibrium developed by Weintraub, Benkard, and Van Roy (2008).

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), Amisano and Giorgetti (2013)), (iii) it is an oligopolistic industry with firms that ostensibly take a strategic and long-term perspective in making decisions[8] and (iv) pay-offs are relatively big as are the fixed costs of entry so entry decisions have important consequences. Our estimates imply positive spillover effects of entry. Moreover, these spillovers lead to heterogeneity not just across firms but also within firms over time based on their history of entry decisions. Our spillover estimates underscore the motivation for the model specification and the methodology developed, and illustrate that entry spillovers may be a potentially important factor to consider in the equilibrium evolution of the generic drug industry.

In summary, two important contributions of our paper that we highlight through the empirical application to real data are: (1) developing a framework to estimate dynamic discrete games in which unobserved firm states evolve based on past actions, and (2) bridging the literatures on estimation of linear models with latent states using a Kalman-Filter (which is a special case of particle filters) and dynamic structural models. More generally, our method is applicable to estimating dynamic games in which (i) the choice set is discrete in nature, e.g., entry and exit, 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.

2 Data and Background

The generic drug industry is a very big and important part of the U.S. pharmaceutical market, e.g., it comprised about 65% of all prescriptions and about $58.5 billion in sales in 2007. 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.[9] This time period is particularly interesting because of a big change in regulation, i.e., 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

---

[8]It is possible that there may be exceptions to this, e.g., among fringe firms. However, given that our focus is on dominant firms this framework may be a plausible place to start. Moreover, we have also estimated a version of the model for boundedly rational firms given the alternative possibilities.

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

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.

We refer the reader to Scott-Morton (1999) for details of the data set and the institutional environment and just summarize the facts relevant for our study here. In particular we emphasize the features of the industry 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 costs of submitting an ANDA can range from $250,000 to $20 million while the average revenues for generic firms in one-firm markets are $10 million. Thus, the cost of submitting an ANDA is very heterogeneous and high, even though it is much less than that for a new drug invention. (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 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 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). Also of note is that 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. Hence, 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 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. We order markets according to the date when the first ANDA was received by the FDA for a particular market opportunity. 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.[10] 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. On average 3.3 firms enter a market, with the minimum being one and the maximum being nine firms. More relevant in the context of our model to be described below, on average 1.2 dominant firms enter a market. 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. Our dynamic model represents a sequence of simultaneous entry games that are inter-linked because the costs of firms evolve based on their past entry decisions. 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. In the tradition of the literature on dynamic games, the model we formulate is very stylized and abstracts from a number of rich institutional details described in the previous section.[11]

The actions available to firm i when market t opens are to enter, $A_{it} = 1$, or not enter $A_{it} = 0$.[12]

---

[10]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.

[11]For example, we do not model the diffusion of a product in the market over time due to consumer learning (see e.g., Ching (2010b)). Modeling the diffusion of a product over time 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 this 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. Moreover, our data is not conducive to an analysis of such diffusion of a product over time. Similarly, we do not model the decision of branded manufacturers to introduce new drugs and create markets, or to pre-empt entry from generics (see e.g., Ching (2010a)). As is also the convention in this literature, we assume entry decision for each firm are sequential and abstract from multiple simultaneous decisions or a portfolio of decisions made by a firm because that would make the model computationally intractable. We discuss this more below in Section 6. Also, as in Scott-Morton (1999) we don't model the decisions of pioneer brands to pre-empt entry of generics in different ways, e.g., by using legal means such as "patent thickets" or by introducing its own generic version or partnering with another generic manufacturer. Our model is primarily motivated by our focus on estimating the spillover effects of entry on future profitability and we leave enriching it with such additional features to future work.

[12]It should be noted that we are modeling entry in to product categories, not firm entry in to an industry. As stated by Scott-Morton (1999, p. 423), "...entry by a particular firm is a fairly irreversible decision; the costs that can be recovered upon exit from one drug (and not a factory) are close to zero." Moreover, there are no exits in our data, which is a sub-sample of the Scott-Morton (1999) data. Our dynamic model builds on her static set up and she does not model exits either. Furthermore, even if a firm were to exit a category it could presumably still be

Empirically this is determined by whether a firm submits an ANDA or not. There 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 the evolution of (latent) firm costs with endogenous feedback from past actions. The firm specific total costs for the current entry opportunity, C_{it}, are unobserved to the researcher. These are determined in part by past entry decisions and random shocks that may affect each firm's costs. The past entry decisions account for spillovers of past industry experience on costs at 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 to allow for serial persistence in firm level costs. 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, σ_c is a scale parameter, κ_c is the entry spillover or immediate impact on cost at market t if there was entry in market $t-1$, μ_c is a location parameter that represents the overall average of the log cost over a long period of time. It should be noted that a priori the model allows for both positive ($\kappa_c > 0$) or negative spillovers ($\kappa_c < 0$). In contrast with the positive spillovers discussed earlier, negative spillovers may arise 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 or expand the number of teams.

The autoregressive parameter ρ_c represents the degree of persistence between the current cost and its long run stationary level. The cost process exhibits state dependence on lagged entries, and also accumulates heterogeneous random cost shocks $\sigma_c e_{it}$ over time. The model allows for unobserved heterogeneity between firms through the effects of current decisions on future latent costs. Alternatively put, heterogeneity arises in the model based on endogenous feedback of past actions of the firms. Instead of specifying unobserved heterogeneity as a time invariant firm specific fixed effect we employ a more general model that allows for serially persistent unobserved heterogeneity in cost that varies over time based on past actions.[13]

---

operational in other product categories. Also, if modeling another empirical setting in which exits are observed then extending our framework to allow for exits would be potentially feasible and interesting.

[13]In an even richer specification not reported here we estimated a version of the model that allowed for firm specific $\mu_{i,c}$. A statistical test could not reject the equality of the $\mu_{i,c}$. We demonstrate why this may be the case in Section 6 below. Thus, we estimate the more parsimonious version in the paper with μ_c. However, in estimating

We assume, as in Berry (1992) and Scott-Morton (1999), that 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 solely on past actions), $c_{k,i,t}$, and a component unobservable to the researcher, $c_{u,i,t}$, 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 μ_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 $c_{k,i,t} = -\sum_{j=0}^{\infty} \rho^j \kappa_c A_{i,t-j-1}$. The classification of the known (or observed) and unobserved components of the cost process is convenient for the implementation of the particle filtering algorithm described below in Section 5.[14] 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. Furthermore, in view of the excellent fit to the data that we are able to achieve (see Section 7 below) this timing convention appears reasonable a posteriori.

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), μ_r is a location parameter that reflects the average total revenue for all the firms across all market opportunities, and σ_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

---

the model we do allow for heterogeneity in the *initial* costs $c_{k,i,-1}$ of firms based on the history of their actions using a pre-estimation sample. Section 4 provides details. We find firm specific heterogeneity in the initial cost is sufficient to capture the key features of the data as described in Section 7. In addition, we incorporate structural state dependence through the lagged entry decisions. We account for state dependence and unobserved heterogeneity directly by incorporating both in our model explicitly. The difference in capabilities for each firm can arise from either endogenous path-dependent unobserved heterogeneity or time invariant firm fixed effects. Given that we could not reject the equality of firm specific $\mu_{i,c}$ our approach emphasizes the former in favor of the latter. Although, our approach does accommodate some degree of time invariant firm fixed effects through the heterogeneity in the initial conditions (see Equations (3), (4), and (5)).

[14]In modeling the dynamic cost process (equation (2)) we focus on the sequence of market openings and not duration between the openings. An extension would be to model this duration. An implication of our specification in equations (3)-(5) is that irrespective of the calendar time that has elapsed between any two adjacent market openings, the cost changes are of the same magnitude. Such an assumption is implicit and common in much of the literature on structural entry models. Moreover, this is a plausible assumption in our context 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. For the ANDA dates shown in Table 1 the mean interval is 42 days and the standard deviation is 37. In fact, the time interval between each two adjacent entries appears to be close enough to six weeks to use six weeks as an approximation for the true interval between entries.

generic firms, and being proportional to the total discounted value of the revenue flows to generic drugs after patent expiration. Firms have rational expectations about revenues at future market opportunities.[15]

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.[16] Therefore, we consider only the dominant firms. We consider two cases where the potential entrants are either: (1) the top three or (2) top four dominant firms, i.e., $I = 3$ or $I = 4$. In each case the remaining firms are combined into a category referred to as "other" as described below. The function that relates the entry decisions of the dominant firms to the fraction of the market allocated to the other firms is taken as given and is anticipated by the dominant firms when considering entry. The procedure we use to implement this is described in detail next.

In the spirit of the literature on entry games (e.g., Scott-Morton (1999), Mazzeo (2002), Seim (2006), Zhu and Singh (2009)) we define a reduced form profit function for a dominant firm i at time t which takes the following form,

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

[15]Following the Scott-Morton (1999) set up we take the revenue in the last year before patent expiration for a branded product as a proxy of the size of the market for the generic entrants after patent expiration. As discussed below our model allows for the share of revenues to be determined by the number of generic entrants in equilibrium. Our assumption that firms can anticipate how much revenue a given entry opportunity offers is reasonable if the following are valid: (i) firms are not observed exiting in the data, and (ii) the markets tend not to come to an unanticipated end. These conditions are plausibly true in our empirical setting. We do not observe any exits in our data as stated above. Typically, manufacturers of branded products develop the market (e.g., through advertising or detailing to physicians) prior to patent expiration. After the patent has expired the incentives for a branded manufacturer to continue such promotional activities is substantially reduced. Furthermore, generic entrants also do not tend to carry out such promotional activities after patent expiration. It is also worth noting that FDA drug approval process is quite long, taking multiple years, and once drugs are approved it is known when their patent will expire. Additionally, the market for a particular generic product typically (with some exceptions) tends to have a finite lifetime until a more newly developed branded substitute product goes off patent and is available in a generic version. This can also be reasonably anticipated because the date a patent expires is publicly available. Thus, we think it is reasonable to assume that firms can anticipate and form expectations about how much revenue an entry opportunity offers based on the revenue of the branded product before patent expiration. We thank an anonymous referee for pointing out these underlying conditions needed for our assumption about anticipation of revenues at an entry opportunity to be plausible.

[16]See Weintraub, Benkard, and Van Roy (2008) for a discussion of the concept of Oblivious Equilibrium (OE) and the associated method to compute the solution to dynamic games when the number of players is very large. In the OE approach firms ignore the current states of their rivals and make decisions instead based on the long run equilibrium aggregate state of the industry. Since we want to model strategic interaction between firms this makes using the OE approach less appealing for us. Given the trade-off in allowing all 51 firms to interact strategically (which would not be computationally feasible) or adopting the OE approach that abstracts from strategic interaction, we have developed a model of strategic interaction between a subset of dominant firms that were more active and had greatest number of entries, while treating the less active firms with lower number of entries as passive participants. Moreover, as stated below, the top four firms enter 60% of the markets, the top five 65%, and the top ten firms 73%. Also, the average number of firms that entered any market is 3.3, and the maximum number of firms that enter any market is nine. Thus, including additional fringe firms does not seem to provide incremental information commensurate with the additional computational burden that it imposes. See also the recent paper by Benkard, Jeziorski, and Weintraub (2015) that introduces the concept of Partially Oblivious Equilibrium (POE), extending the OE concept to industries with high concentration but many players to allow for strategic interaction between a few dominant firms. The POE approach might be an excellent way to model the generic pharmaceutical industry in future research.

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 β 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. In the above, N_t is used to denote the number of entering dominant firms. This 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 non-dominant firms. 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 a market. The idea of this regression dates back to Bresnahan and Reiss (1991b) 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 with $1 - b = 0.908$ being a reasonable lower bound for γ. The upper bound is one.[17]

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

---

[17]In the reduced form profit function (7), we note that although we are calling the latent variable $(C_{i,t})$ firm "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 entry spillovers could be factors like learning by doing or economies of scope or development of distribution networks. This is similar to the unobservable in the reduced form pay-off function in the literature on discrete games of entry and location (e.g., Scott-Morton (1999), Seim (2006)). The unobservable could result from either demand or supply side factors and is typically attributed to one or the other depending on the context or convenience. 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 forms of spillovers (e.g., in liquid injectable drugs), or disentangling supply or demand side sources of these is beyond the scope of this paper and is left for future research.

t of the game. The state variables include the firm specific latent costs $C_{i,t}$ signifying that the identity of each firm i is relevant.

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

$$
\begin{aligned}
V_i&(C_{it}, C_{-i,t}, R_t) \\
&= A_{it}^E \left(R_t^\gamma / N_t^E - C_{it} \right) + \beta\, \mathcal{E}\left[V_i(C_{i,t+1}, C_{-i,t+1}, R_{t+1}) \,|\, A_{i,t}^E, A_{-i,t}^E, C_{i,t}, C_{-i,t}, R_t \right],
\end{aligned}
\tag{12}
$$

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.

A hurdle we face is multiplicity of equilibria.[18] Multiple equilibria may take the form of a situation where one firm or another can profitably enter but if both enter then each will incur losses. However, if neither enters then one of them would have an incentive to deviate. To deal with multiple equilibria we adopt an explicit equilibrium selection rule that is similar to that used by Berry (1992) and Scott-Morton (1999). In particular, we pick the equilibrium that has the lowest aggregate cost $(C = \sum_{i=1}^{I} A_{it} C_{it})$, i.e., entrants are such that the most profitable production takes place.[19]

---

[18]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 7 for additional details.

[19]One justification for this approach is that a firm with lower costs of entry is also more likely to gain approval from the FDA earlier although this is by no means certain, e.g., Scott-Morton (1999), Ching (2010a). Thus, the equilibrium selection rule seems a plausible way to deal with the multiplicity issue.

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 the Web Appendix. 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. Therefore we apply Bayesian methods in this paper.

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 a particle filter or 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). (6) Use the likelihood 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. We outline the details next.

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 in place, we can write $x_t = (c_{u,1,t}, \ldots, c_{u,I,t})$, and $y_t = (a_{1t}, \ldots, a_{It}, c_{k,1,t}, \ldots, c_{k,I,t}, r_t)$.

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

14

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: First, 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. We compute the last two pre-scandal values $c_{k,i,-1}$ and $c_{k,i,0}$ for each firm by running the recursion (5) 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}. Thus, the pre-scandal data allow for us to control for firm specific heterogeneity in initial costs. This firm specific cost then evolves endogenously conditional on the entry decisions of the firm. Hence, our approach allows for both path-dependent unobserved heterogeneity as well firm fixed effects based on initial costs.

Second, the scandal log revenues $\{r_t\}_{t=-n_0}^0$ can be used to help identify the parameters μ_r and σ_r. 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 (μ_r, σ_r) of θ, we denote it as $p(Y_{pre} \mid \theta)$ for convenience. We discuss estimation of this likelihood below in Section 5.

Since we are estimating a game of pure strategy, a density for the strategy profile A_t is a deterministic function of (x_t, r_t, y_{t-1}) and model parameters. 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 of such a likelihood with a particle filter would be computationally difficult as each particle with positive weight would have to predict all decisions perfectly.[20]

We resolve this problem by introducing a probability that allows for additional uncertainty in the ex-post decision, $q_a = 1 - p_a$, $0 < p_a < 1$. This yields the following likelihood for an observed ex-post action profile A_t^o

$$p(A_t^o \mid r_t, x_t, y_{t-1}, \theta) = \prod_{i=1}^I (p_a)^{I(A_{it}^o = A_{it})} (1 - p_a)^{I(A_{it}^o \neq A_{it})} \tag{15}$$

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

$$\theta = (\mu_c, \rho_c, \sigma_c, \kappa_c, \mu_r, \sigma_r, \gamma, \beta, p_a) . \tag{16}$$

The full likelihood for the data is the product $\prod_{t=1}^n p(A_t^o \mid r_t, x_t, y_{t-1}, \theta)$.

Although, this statistical approach is computationally useful it may not be very appealing from the perspective of economic interpretation. In our view the most plausible interpretation is that there is a small probability q_a that planned entry decisions are not realized. This could

---

[20]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.

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 or management at the parent company of a subsidiary or by a change in firm leadership. Less plausible interpretations are that (15) 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] The boundedly rational likelihood is given by (15) with A_t computed as described in the Web Appendix which provides additional details on computation and estimation. 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. Although, the fully rational model is computationally much more burdensome we estimated both boundedly rational and fully rational versions of the model. The results, however, turned out to be qualitatively similar so we only report and discuss the results for the boundedly rational model.[22] In a related paper Gallant, Hong, and Khwaja (2014), we analyze the fully rational model as part of a Monte Carlo exercise examining the statistical properties of the estimator.

5   Parameter Estimation

Douced, de Freitas, and Gordon (2001) present a concise description of the particle filter or sequential importance 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

$$p(x_t \,|\, x_{t-1}, \theta), \tag{17}$$

which is defined by recursion equation (4), the initial density

$$p(x_0 \,|\, \theta), \tag{18}$$

which, from equation (4), is normal with mean μ_c and standard deviation $\sigma_c/\sqrt{1-\rho_c^2}$, and the observation density

$$p(y_t \,|\, y_{t-1}, x_t, \theta) = p(A_t^o \,|\, r_t, y_{t-1}, x_t, \theta)\, p(r_t \,|\, y_{t-1}, x_t, \theta), \tag{19}$$

where, from equation (6), $p(r_t \,|\, y_{t-1}, x_t, \theta)$ is normal with mean μ_r and standard deviation σ_r.

The sequential importance sampler is as follows:

---

[21]See Che, Sudhir, and Seetharaman (2007) for another example of a model of forward looking firms that are boundedly rational.

[22]Results for the fully rational model are available from the authors on request.

1. For $t = 0$

   (a) Start N particles by drawing $x_0^{(j)}$ for $j = 1, \ldots, N$ from the initial density equation (18).

   (b) Compute

   $$p(y_0 \,|\, \theta) \;=\; \int p(y_0 \,|\, y_{-1}, x_0, \theta)\, p(y_{-1}, x_0 \,|\, \theta)\, dx_0$$

   $$\doteq \;\; \frac{1}{N} \sum_{j=1}^{N} p(y_0 \,|\, y_{-1}, x_0^{(j)}, \theta).$$

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

   (a) For each particle, draw $\tilde{x}_t^{(j)}$ from the transition density equation (17) 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 (19); i.e.

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

   The parametrization in equation (15) 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 (2009), 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

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

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

17

Note that $p(y_t \,|\, y_{t-1}, x_t^{(j)}, \theta)$ 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

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

We use the Metropolis-Hastings algorithm to estimate the model. Briefly, the method is as follows.[23] The proposal density $q(\theta^o, \theta^*)$ defines a distribution of potential new values θ^* given an old value θ^o. Denote the likelihood by $\mathcal{L}(\theta)$ and the prior by $\pi(\theta)$. Given the value θ^o at the end of the MCMC chain, one moves the chain forward one step to θ' as follows:

1. Draw θ^* 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 α, set $\theta' = \theta^*$, otherwise set $\theta' = \theta^o$.

For our particular q, one randomly chooses an element j of θ^o to move and then proposes a new value by replacing θ_j^o with a draw from the normal distribution with mean θ_j^o and scale σ_j, where σ_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 θ in the outer loop. The sequential importance sampler generates a sequence of cost draws within the inner loop. Solving the game both evaluates the conditional likelihood along this trajectory of cost draws 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.

In estimating parameters we use flat, noninformative priors that impose these support conditions: $-1 \le \rho_c \le 1$, $0 \le \kappa_c$, $0 < \sigma_c$, and $0 < \sigma_r$. 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.

---

[23] Additional details are provided in a companion Web Appendix that is not meant for publication and is available online and from the authors directly.

The firm's discount factor β is extremely difficult to estimate in studies of this sort (see e.g., Rust (1994), Magnac and Thesmar (2002) and Ching, Erdem, and Keane (2013)). Hence, the literature on structural estimation often does not estimate such parameters but rather fixes these.[24] Similarly, we do not attempt to estimate the discount factor but set $\beta = 0.96875$ after rounding it to a nearby fractional power of two (in order to speed computation). 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. 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, we arrive at the value of the discount factor reported above.

Just as for the discount factor we do not attempt to estimate the parameters γ and p_a for the reasons discussed below. Examination of equation (9) indicates that were γ to enter as a linear factor then γ 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 γ yield estimates that are about 0.93. Therefore, based on the plausible lower bound of 0.908 derived in Section 3 and our experience from trying to estimate γ, we take 0.93 to be a reasonable value. Rounding to a nearby fractional power of two (in order to speed computation), we set $\gamma = 0.9375$.

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 θ are hardly affected. What we do find is that varying p_a affects the rate at which particles "die out" at Step (2d) in the sequential importance sampler. A particle is considered to "die out" if its importance weight $\{\hat{w}_t^{(j)}\}$ is zero or very close to zero. This happens if the observation density $p(y_t \,|\, y_{t-1}, \tilde{x}_t^{(j)}, \theta)$ for a particular particle trajectory tends to zero. 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 7.

---

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

6 Identification Analysis

Non-parametric identification of dynamic games of complete information is an area of ongoing research with few conclusive results. The main hurdles for establishing identification come from the potential for multiplicity of equilibria and because the information set of the agent is much richer than what is typically observed by the researcher. In this spirit it is beyond the scope of this paper to establish non-parametric identification of our model. However, in this section we discuss the basis for identification in our model that relies on a mix of data variation, economic modeling and statistical assumptions.[25]

There are three groups of parameters in our model. The first group consists of the discount factor β, the market share parameter for dominant firms γ and the probability parameter p_a for uncertainty in ex-post decisions. These parameters are difficult to identify (see e.g., Rust (1994), Magnac and Thesmar (2002)). As discussed above in Section 5 we follow the literature on structural estimation and do not attempt to estimate these parameters but rather calibrate these based on preliminary analysis of the data. Our preliminary analysis shows that the estimation results do not seem to be sensitive to small differences in these parameters. The second group of parameters consists of the mean and the variance of the normal distribution of revenues. Since we observe revenue data, they are identified directly by the sample average and the sample variance of the revenue data.

The key parameters in our model that the estimation procedure aims to identify are the third group of parameters, which include the stationary mean μ_c of the latent cost distribution, the variance σ_c^2 of the innovation of the cost process, and most importantly, the autoregressive coefficient ρ_c of the cost process and the coefficient of lagged entry or spillover effect κ_c. These parameters jointly determine the cost evolution of firms and hence their entry behavior. In the data set, however, we observe the entry pattern of the firms over time, revenues and the number of entrants, and do not observe the latent cost directly. At a very intuitive level identification of cost parameters

---

[25]We estimate the model using Bayesian methods and the particle filter algorithm is also a Bayesian procedure that computes the posterior of the unobservable costs conditional on the data. Hence, the notion of classical "non-parametric" identification is not directly applicable in our case. Although, ideas related to (non-parametric) identification of distribution of unobservables are broadly relevant (e.g., Heckman and Singer (1984)). Simlarly the notions related to the identification of hyper-parameters in hierarchical Bayes models are relevant in our case. Briefly, in Bayesian settings the focus is on (parametric) identification (Hurwicz (1950), Zellner (1971)), i.e., a parameter value θ_1 is identified if there is no other value θ_2 such that $p(y|\theta_1) = p(y|\theta_2) \; \forall \; y \in \mathcal{Y}$. If all parameter values are identified then the model $p(y, \theta)$ is identified, and the parameter θ is identified (Lancaster (2004)). Furthermore, as stated by Lancaster (2004), "Flat spots at the top of the likelihood pose a problem for maximum likelihood inference since there will never be unique maxima and second derivative matrices will typically be singular at non-identified points. It is of no special significance from the Bayesian point of view because Bayesians do not maximize likelihoods–they combine them with priors and integrate them. A qualification to this is that if all values of a parameter on, say the, real line are unidentified then an (improper) flat prior distribution on that line would lead to a flat posterior and this is not allowed (p. 26)." The main identification concern in Bayesian settings is that (1) MCMC algorithms can be constructed even when a posterior does not exist or (2) a posterior distribution may exist for a parametrically unidentified model given a proper prior (Geweke (2005)). Alternatively put, with a properly specified prior it is always possible to conduct Bayesian analysis of a parametrically unidentified model, i.e., the prior and posterior may be identical for non-identified parameters. For more on identification in Bayesian settings see Poirier (1998).

may be thought of in a way similar to the identification of such paramters in Berry, Levinsohn, and Pakes (1995). By inverting the first order condition for profit maximization Berry, Levinsohn, and Pakes (1995) are able to relate the parameters describing the unobservable costs to the observables in their data, e.g., product characteristics (see Section 3, pp. 853-854). The analogous optimality condition in our case comes from the Bellman equation (9). This provides a way to relate the unobservable cost parameters to the observables, i.e., revenues, entry decisions and total number of entrants. The main difference is that because we solve a highly non-linear dynamic programming problem there is no simple way to invert this optimality condition to and represent it analytically as in Berry, Levinsohn, and Pakes (1995) (see Equation 3.6, p. 854). However, it is through exploiting the relationship generated by the optimality condition between the observed data and unobserved cost that we are able to pin down the parameters of the cost process. Alternatively put, these parameters are identified based on (i) variation in the choices of a given firm over time and (ii) variation in choices in a given market across firms, with these being used to infer the unobserved cost. Since, we are dealing with unobservable costs, we need to make some parameteric distributional assumptions in order to infer these from the optimality conditions and then project these on the observables in our data to estimate the cost parameters. It should be noted however that the distributional assumptions we need to make to implement the particle filter to integrate our the unobserved costs from the likelihood are sufficient to guarantee the identification of cost parameters, and no extra parameteric assumptions are required. Next we discuss in detail how this is done.

The (unobserved) cost process incorporates dependence on one period lag cost and lagged entry that have to be inferred from the observed data on entry and revenue. This is related to the issue of unobserved heterogeneity in pay-offs versus state dependence. We address this in multiple ways. In a pioneering paper Heckman (1981) pointed out that one can use the panel structure of data to exploit *within* agent variation in states to identify state dependence from unobserved heterogeneity. In particular for us this requires that we observe each of the four possible transitions $A_{t-1} \in \{0, 1\}$ to $A_t \in \{0, 1\}$ for every (dominant) firm in our data set. As can be seen in Table 1 this is indeed the case. Identification is aided by this source of variation with the particular exclusion restriction that the total size of the market or revenue affects entry decisions but conversely the entry decisions of generic manufacturers don't affect the total market size. This would be plausible if the market size for a drug is primarily developed by the pioneer branded product and generic manufacturers take this market size to be pre-determined in making their decisions. Alternatively stated, the total market revenue is determined prior to the entry of the generic firms and does not depend on the costs of entry for generic firms. This helps to identify the parameters associated with the costs of generic manufacturers.

Another approach to this problem is to directly incorporate firm fixed effects. As stated above in Section 3, we estimated an alternative version which directly allowed for firm specific fixed effects $\mu_{i,c}$ in the cost process (Equation 2) along with state dependence from lagged entry. However, we

could not reject the statistical equality of these firm specific $\mu_{i,c}$. This would be the case if the time varying specification that we adopt soaks up most of the unobserved heterogeneity leaving little for the firm fixed effects to capture. We demonstrate below how this might happen. The difference in costs for each firm can arise from either endogenous path-dependent unobserved heterogeneity or time invariant firm fixed effects and our approach emphasizes the former over the latter.[26] To our knowledge, we estimate one of the most general specifications of a game with asymmetric firms. For example, the literature on estimating games largely treats players as symmetric and without firm specific unobserved heterogeneity, but allowing for market level heterogeneity.[27] In contrast, our approach goes beyond an ad-hoc time invariant firm specific fixed effect by allowing for even ex-ante identical firms to become heterogeneous over time as their cost evolves based on endogenous feedback from past actions (Equation 2).[28]

The model assumes that firms make entry decisions sequentially across markets over time and know the market size, i.e., the revenue of the patented drug at each market opening. The firms also form expectations about the market size at future entry opportunities. These assumptions play crucial roles in the identification of the model. An alternative specification might be that firms make a portfolio of decisions for a set of entry opportunities at a time. However, we do not adopt this specification for a number of reasons. Our analysis is at the level of subsidiary firms which is similar to the primary specification in Scott-Morton (1999), where there is no co-ordination over a set of entry decisions. If the analysis was at the level of parent firms it could be that the parent firm might co-ordinate entry decisions for multiple subsidiary firms simultaneously. Scott-Morton (1999), however, shows that "the choice of organizational level from which to analyze entry does not drive the results (p. 434)" and so co-ordination of entry decisions over a set of opportunities is not a concern. Furthermore, the assumption about entry opportunities being sequential is common in the literature on estimating entry games.[29] In our context assuming entry opportunities are sequential is more plausible than assuming that firms make entry decisions on future market openings simultaneously, despite the fact that the patent expiration dates of drugs are well known in advance. This is because market conditions and latent firm heterogeneity variables typically vary over time, and a decision that was ex ante optimal prior to the market opening may no longer be optimal closer to when the market opens and new information becomes available to firms. In addition, a model of simultaneous entry in to a group of markets will induce a much larger set of endogenous variables due to a potential portfolio of choices, and there will not be sufficient

---

[26]Moreover, a specification with unobserved cost as a time invariant fixed effect seems less interesting to us than a story of the dynamic evolution in unobserved cost based on past actions.

[27]For example, Ellickson and Misra (2008) control for unobserved heterogeneity at the market level but *not* at the level of the firm. As they state on pp. 820-821, they have a two-fold approach, "The first incorporates a fixed effect at the MSA level, while the second incorporates a random effect at the level of the cluster."

[28]A structural learning model can also generate a similar implication by endogenously generating heterogeneity across consumers in perceived quality even though if they started from identical priors. This is because the perceptions of agents evolve based on their heterogeneous experiences, e.g., Ching, Erdem, and Keane (2013).

[29]As an example, Hitsch (2006) models entry and exit in the breakfast cereal industry where 78 new products were launched between 1985 and 1992 at the rate of 10 products per year (p. 28). The paper models launch of each product as a sequence of entry decisions.

exclusion restrictions to identify the coefficients on all the endogenous variables.[30]

Another assumption we make for identification is that the revenue process (Equation (6)) evolves with an i.i.d. error term and firms do not know the exact realization of revenues at future entry opportunities. If firms do have more information about the future realization of revenues then the dimension of the state space of the model will necessarily become larger and the computation will be substantially more costly. Typically, while firms might have some information about future market profitability than implied by the i.i.d assumption, if revenue evolves according to a first order Markov process, then the current revenue will still be a sufficient statistic for the observed state variable of the Bellman equation. The i.i.d assumption, can be a cause for concern of model misspecification. Although, this assumption is common in the literature on estimating games (see e.g., Aguirregabiria and Mira (2009)). In our case, given that the misclassification rate is small in the current model,[31] the amount of model misspecification is likely to be small.

Next, we demonstrate how with our time varying specification of path dependent unobserved heterogeneity the fixed effects might be negligible. We have 163 market entries from 1984 through 1989 that can be used to provide additional information on certain parameters (see Section 4) but not others due to a structural break in the data. In particular we recurse the observable part of cost (Equation 2) through these 163 observations for each MCMC draw of θ to generate the initial heterogeneity in firms in 1990 for that θ. Suppose, instead, that we introduce initial homogeneity by estimating an initial condition $c_{\mu,i,1984}$ for each firm rather than starting these recursions in 1984 at zero, which is what we actually do. The dominant firm enters 45% of the markets; ρ_c and κ_c of Equation (2) are about 0.99 and 0.07, respectively, which implies that the long run steady state value for the dominant firm would be about $0.07 \times 0.45/0.01 = 3.15$. Thus, draws of $c_{\mu,i,1984}$ for this firm would be about 3.15. By the time the recursion reaches 1989 the magnitude has been reduced to 0.99^{163} times that value or 0.658, which is about the most that $c_{\mu,i,1990}$ can differ from the value that we compute. Particles are started by adding a draw to $c_{\mu,i,1990}$ from the unobserved cost component that has stationary mean μ_c of about 10 and stationary variance $\sigma_c/(1 - \rho^2)$ of about $0.37/0.02 = 18.6$. The value of 0.658 is going to be negligible with respect to this draw. Similar arguments apply to all firms. For instance, the smallest firm enters 25% of the markets so that its possible magnitude at 1989 can differ from what we use by at most 0.145. One can conclude from the foregoing that attempting to introduce initial heterogeneity and estimate it would add three or four poorly identified parameters without changing results substantively. This is exactly what we find in our alternative specification as discussed above. An intuitive way to think about this is that since the estimated serial correlation is lower than 1 (see Section 7 below) a model with time invariant unobserved heterogeneity would fit the data worse than one that allows for time varying unobserved

---

[30]Hypothetically speaking, if one attempts to model the 40 entry decisions as a simultaneous portfolio choice then this would become a combinatorial problem of dimension 2^{40} for each firm. With 4 such firms, this problem has a dimensionality of 2^{160}. A model that allows for strategic interaction among firms that simultaneously make entry decisions about 40 markets and moreover allows for spillovers across time from previous entry is both empirically infeasible and computationally intractable.

[31]See discussion of classification error rates (CER) in Table 2 in Section 7 below.

heterogeneity. Alternatively put, the unobserved cost variable plays the role of a serially correlated firm specific "error term" which soaks up most of the unobserved heterogeneity. We return to address this issue in yet another way in Section 7 below in our discussion about potential reverse causality.

Additionally, we consider a numerical verification of identification local to our parameter estimates as an analytical demonstration of identification for a model of this complexity is not available. Figure 5 of the Web appendix plots the profile likelihoods for the three firm model and Figure 7 is the same for the four firm model. In each panel of these plots, the indicated parameter is moved and all others are fixed at the values that maximize the likelihood. The sharp peaks suggest local identification in that a change to the parameter must reduce the likelihood. Profile likelihoods do not eliminate the possibility that two or more parameters can be moved together in such a way as to leave the likelihood invariant. A check on this is to compute the condition number of the Hessian of the log likelihood or the condition number of its inverse, these being the same number. If the log of the condition number is a good deal smaller than the log of machine precision, that indicates a well conditioned Hessian and suggests that a simultaneous movement of the parameters that leaves the likelihood invariant is impossible. The variance matrix of the MCMC chain is a consistent estimator of the inverse of the Hessian (Chernozhukov and Hong (2003)). For the three firm model the log (base 10) of the condition number of the variance matrix is 3.75 and for the four firm model it is 3.85. The log of the precision of a "double" on the computing equipment we use is 15.7. Hence, the Hessians are well conditioned.

7   Results and Robustness

We estimate the mode for the cases of three or four dominant firms, i.e., the top three (four) dominant[32] firms are the only potential entrants that are strategic competitors and the actions of the remaining firms are accounted for by the parameter γ. The mode and standard deviations of the posterior distribution are reported in Table 2. We focus on the mode of the multivariate posterior distribution because it actually corresponds to a value at which the model has been evaluated. Other measures of central tendency of the posterior distribution can be misleading when studying the behavior of a structural model because they may have never appeared in the MCMC chain and could give a distorted view of the model were it to be evaluated at such a point. We next discuss the estimates with the caveat that our model is a stylized representation of oligopolistic competition in the generic pharmaceutical industry, and therefore the results should be assessed with appropriate prudence.

Table 2 about here

---

[32]Recall that Mylan, Novopharm, and Lemmon are the potential entrants in the three firm case. Geneva is the additional potential entrant in the four firm case. There are 51 firms in total, so 48 firms are put in the "other" non-dominant category in the three firm case and 47 in the four firm case.

The parameters are tightly estimated[33] and, as seen from the extremely low classification error rates (CER), model fit is quite accurate.[34] We begin by discussing the revenue parameters. The mean of (log) revenue is estimated to be about 10 and the standard deviation about 1.6. As a reality check we find that these numbers are very similar to the numbers reported in Scott-Morton (1999). The long run mean of cost μ_c is also about 10. Given the caveat above we provide some interpretation next. Since the means of revenue and long run cost are similar it implies that in a static sense on average each market will not be able to support too many entrants.[35] Interestingly, in our sample the number of dominant firms that enter a market on average is 1.2. As a consequence firms that enter will need to decrease their cost based on experience (Equation 2) in order to stay profitable in the long run (Equation 8). Modeling entry decisions in a static one shot entry game framework would suggest that in the generic drug industry firms are not very profitable on average or that only low cost firms would enter a market. However, such a model would treat cost as an exogenous firm characteristic. In contrast, our model by connecting the markets through the spillovers of entry and modeling an inter-linked series of entry games is suggestive of an additional insight that firms that do enter and lower their cost through past experience can gain additional strategic advantage to survive these tight markets. Indeed, two firms with initially similar costs might have very different long run trajectories if one enters a very tight market and becomes successful based on past experience whereas the other stays out and loses the opportunity to learn from the experience of entering a market. Put an another way, our model indicates an additional long run reward from experience gained from entry over and above the short run profits from entry. We further compare our work with static entry models below.

We next discuss the parameters of the cost process with the same due caution as above. The autoregressive parameter ρ_c represents the degree of persistence in the cost process. The value of ρ_c is quite large. This indicates that unobserved firm specific costs can be persistent and highly path dependent. A potential implication is that the strategic effects of entry spillovers on firm heterogeneity will be magnified in the long run. Moreover, it suggests that firm heterogeneity will not be a fixed characteristic but will evolve over time given the history of entry decisions. The spillover effect of entry is κ_c which is estimated to be 0.07. This implies a potential immediate cost improvement of 7% going into the next market opening conditional on entry in to the current market. Recall, in our data there are on average eight entry opportunities a year or about one every six weeks. To gain a sense of the probable long run implications of this spillover effect we do a thought experiment and compute the average annual cumulative improvement using the AR(1)

---

[33]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 the Web Appendix. The likelihood is proportional to the posterior because priors are flat.

[34]The cost spillover κ_c is estimated to be a little larger in the case with the larger number of dominant firms. On the other hand, as might be expected, the classification error rate is also a little larger in the game with the larger number of dominant players.

[35]It should be noted that in our context this statement is really about the "dominant" firms, which are the potential entrants in the model.

cost process (Equation (2)). It is seen that a firm that enters all possible markets in a year will decrease its cost by 51% (relative to a firm that did not enter at all that year). Thus, the dynamic spillovers of entry could possibly be very large cumulatively. However, the standard deviation of the cost process σ_c is also large, i.e., about 0.37. This means that although spillovers of entry can persistent and path dependent they could potentially also be eliminated by a cost shock.Modeling the sources of such shocks is beyond the scope of this paper and we leave that to future work.

Figure 1 about here

Figure 2 about here

Figure 3 about here

We further examine the implications of the dynamics of the cost process by plotting the log cost of the dominant firms in the three firm model. This is done in the upper three panels of Figure 1. The circles indicate that the firm entered that market.[36] It is seen that the top firm, Mylan, has a clear cost advantage over its competitors but the broad trends in cost are about the same for all firms. These plots also suggest heterogeneity in cost across firms and additionally heterogeneity *within* firms over time. Cost evolves, waxing and waning with entry or lack of it. It is seen that a sequence of entries usually leads to cost decreasing. On the other hand periods of inactivity lead to cost increasing. It is notable the firms enter the market both when their cost is relatively low and high. One reason for this could be spillovers of current entry on future costs which may help overcome the adverse current payoff from entry that high current cost implies. This is in contrast with what a static model would predict. In the static case firms would not enter a market when their cost was high. 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. This panel also shows that entry tends to coincide with markets that have larger revenues but not always. There are instances when market revenue is very high yet not all three dominant firms enter (e.g., markets 37, 38, 39 as indicated by the x-axis of the bottom panel). Again, this is a feature that would be difficult to rationalize using a static entry model.

We next compare our estimates of cost and the implied cost trajectories for the three and four firm models. Figure 2 plots together the log cost of the three dominant firms from both the three and four 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. The recovered cost trends are consistent with entry patterns. Mylan has the most entries and the lowest cost. It's cost trajectory is also smoothest

---

[36]The logarithm of cost is computed by averaging at Step (2e) of the importance sampler.

because it has the least fluctuations between periods of activity and inactivity. This suggests that its sequence of entries also manages its cost trajectory.

Using Figure 3 we examine the fit of our model. This figure 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.[37] 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 0.5 to predict entry (i.e., entry if predicted probability ≥ 0.5; no entry otherwise) for the average predictions shown in Figure 3. We find that although our entry model is highly stylized and parsimonious it does a fairly good job of predicting entry patterns and fitting the data.

So far, we have tried to directly infer the importance of the estimate of the entry spillover κ_c by examining the implications of this for cost and entry dynamics. Another way to assess our results is to directly explore the possibility that the firms play a different game than the game we propose. Consider two other games that might be played instead of the game with payoffs (9).They could play a game with payoffs

$$V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t) = A_{it} \left(R_t^\gamma / N_t - C_{u,i,t} \right), \tag{20}$$

where no attention at all is paid to the dynamic spillovers of entry ($\kappa_c = 0$) and firms ignore the continuation value of the game ($\beta = 0$). We call this the *myopic* game ($\beta = 0, \kappa_c = 0$). Or they could play a game with payoffs

$$V_i(A_{i,t}, A_{-i,t}, C_{i,t}, C_{-i,t}, R_t) = A_{it} \left(R_t^\gamma / N_t - C_{it} \right) \tag{21}$$

where they take cognizance of the dynamic spillovers of entry 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 cost trajectories 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 aggregate cost.

These computations suggest that the myopic and static games would do a poor job of rationalizing the data.[38] To check, we use our parameter estimates, impose $\beta = \kappa_c = 0$, and find that the

---

[37]The average prediction is computed by averaging game solutions at Step (2e) of the importance sampler at the mode of the posterior density.

[38]In performing these calculations we did not completely re-estimate the model with the parameters constrained. Instead we re-solved the model at the estimated parameter values with the parameters constrained as specified above.

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.

Given our stylized and parsimonious model, it is worth asking the question whether what is recovered is the dynamic spillover effect of entry, i.e., entry makes firms more capable or whether the causality is reversed and it is more capable firms that enter. In the latter case one could think of a situation where there is persistent heterogeneity in costs across firms and the more capable firms always enter and the less capable firms stay out. Recall that in our model heterogeneity in costs arises endogenously based in part on past actions. Stated differently, one might surmise that the effect of the random shocks on costs (operating through σ_c) is larger than that of the dynamic spillovers of entry (through κ_c), or that σ_c and κ_c are correlated in some way. If this was the case then it would be difficult to identify each separately from the other. One way to check this is to set σ_c to different values and re-run the MCMC chain to see whether it affects the values of κ_c. Setting σ_c to 0.25, 0.125, and 0.0625 has very little effect on κ_c although it does dramatically reduce the likelihood evaluated at the mode. Thus, we have statistical evidence that we are able to identify σ_c from κ_c. Hence, we conclude that we are most likely estimating the effect of entry on costs and not vice versa.

8   Conclusions

Our paper contributes to the literature on estimating entry spillovers and also extends the literature on estimating dynamic discrete games to allow for endogenously arising unobserved heterogeneity between players. We illustrate our method with an application to the generic pharmaceutical industry. Our stylized model fits the data well with small classification error rates. Our results are suggestive of positive spillovers of past entry on future profitability and indicate that dynamic spillovers may play a role in the equilibrium path of the industry structure. Moreover, our results indicate possibility of heterogeneity in the cumulative effects of these spillovers not only across firms but also within firms over time based on the entry decisions. A potential implication is that firms may be able to manage their success through sequencing their entry in to new markets in order to maximize the potential for learning from experience. We develop a procedure based on particle filters to estimate a dynamic discrete game that includes firm specific unobserved state variables that evolve based on past actions, i.e., endogenous feedback. Our method is 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 brands, technology adoption, relocation or shut down of stores etc., (ii) when there are serially correlated unobserved endogenous state variables (either discrete or continuous), e.g., models which allow for

---

Thus, these computations are suggestive of our preferred specification being better at rationalizing the data. A more thorough test would be to re-estimate the model for the three model specifications.

accumulation or depreciation of goodwill of brands, and (iii) an algorithm to solve the game is available.

Before concluding, we outline some limitations 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 spillovers for other kinds of drugs and the sources (e.g., demand or supply side) of such 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. We leave such analysis for future research. Yet another interesting and challenging issue to examine in future might be the role of price competition in the generic drug market especially because institutional features such as health plan co-pays and formulary lists at hospitals etc., may lead to ambiguous effects on revenues.[39] Future research could also examine the role of multi-market contact in these entry decisions, e.g., it could analyze whether any collusive practices exist due to this repeated contact in different product categories. Similarly, another rich area for future research might be to analyze marketing mix activities (see e.g., Manchanda and Chintagunta (2004), Manchanda, Rossi, and Chintagunta (2004), Narayanan, Manchanda, and Chintagunta (2005)) in the generic pharmaceutical industry that complement these entry decisions, and their joint effects on entry spillovers. Furthermore, even though we focus on a game of complete information, the methods we exposit may also provide a procedure to incorporate time varying unobserved heterogeneity with endogenous feedback in estimating dynamic games of incomplete information. We hope our paper will provide a foundation to address some of these issues in future.

Finally, even though our dynamic framework has limitations as discussed we believe it enriches the literature. Our focus has been to extend the literature on dynamic games to allow for time varying unobserved heterogeneity that accounts for the evolution of a firm based on its strategic actions. The goal has been to account for the the dynamic implications of the nature of such competition. This is important for two reasons: (a) firms change over time in terms of getting stronger or weaker based on their experience, and (b) similarly, firms recognize that their rivals' competitiveness depends on the experience of their rivals. The existing literature largely abstracts from this kind of heterogeneity. In the interest of making headway in this direction we have abstracted away from observed heterogeneity in other dimensions that the previous literature has considered. Our hope, of course, is to ultimately allow for both kinds of heterogeneity in future work.

---

[39]We thank an anonymous referee for this suggestion.

References

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

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

AMISANO, G., AND M. L. GIORGETTI (2013): "Entry Into Pharmaceutical Submarkets: A Bayesian Panel Probit Analysis," *Journal of Applied Econometrics*, 28, 667–701.

ANAND, B., AND C. CONNEELY (2003a): "Fox and the NFL-1998," *Harvard Business School Case*, 9-704-444.

——— (2003b): "Fox Bids for the NFL-1993," *Harvard Business School Case*, 9-704-443.

ANAND, B., AND T. KHANNA (2000): "Do Firms Learn to Create Value? The Case of Alliances," *Strategic Management Journal*, 21, 295–315.

ARCIDIACONO, P., AND R. MILLER (2011): "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 79, 1823–1868.

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

BAJARI, P., C. BENKARD, AND J. LEVIN (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

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 (2010): "Estimating Static Models of Strategic Interactions," *Journal of Business and Economic Statistics*, 28, 469–482.

BAJARI, P., H. HONG, AND S. RYAN (2010): "Identification and estimation of a discrete game of complete information," *Econometrica*, 78, 1529–1568.

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, 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, 179–195.

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

BENKARD, L., P. JEZIORSKI, AND G. WEINTRAUB (2015): "Oblivious Equilibrium for Concentrated Industries," working paper, forthcoming RAND Journal of Economics.

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

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, pp. 841–890.

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

——— (1991b): "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, 659–673.

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

CHING, A. (2010a): "A Dynamic Oligopoly Structural Model for the Prescription Drug Market After Patent Expiration," *International Economic Review*, 51, 1175–1207.

CHING, A. (2010b): "Consumer learning and heterogeneity: Dynamics of demand for prescription drugs after patent expiration," *International Journal of Industrial Organization*, 28, 619–638.

CHING, A., T. ERDEM, AND M. KEANE (2013): "Learning Models: An Assessment of Progress, Challenges and New Developments," *Marketing Science*, 32, 913–938.

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

CONTI, R., AND E. BERNDT (2013): "Anatomy of U.S. Cancer Drug Shortages: Technology, Market Structure and Price Competition," working paper, MIT Sloan School of Management.

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, 52–62.

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.

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

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

ERICSON, R., AND A. PAKES (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *The Review of Economic Studies*, 62, 53–82.

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

FLURY, T., AND N. SHEPHARD (2010): "Bayesian Inference Based Only on Simulated Likelihood: Particle Filter Analysis of Dynamic Economic Models," *Econometric Theory*, 27, 933–956.

GALLANT, A. R., H. HONG, AND A. KHWAJA (2014): "A Bayesian Approach to Estimation of Dynamic Models with Small and Large Number of Heterogeneous Players and Latent Serially Correlated States," working paper, Yale School of Management.

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. (2005): *Contemporary Bayesian Econometrics and Statistics.* New Jersey: John Wiley.

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.

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

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.

HECKMAN, J. (1981): "Heterogeneity and state dependence," in *Studies in Labor Markets*, ed. by S. Rosen, pp. 91–139. University of Chicago Press.

HECKMAN, J., AND B. SINGER (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52(2), 271–320.

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

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

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

HURWICZ, L. (1950): "Generalization of the Concept of Identification," in *Statistical Inference in Dynamic Economic Models*, ed. by T. Koopmans, pp. 245–257. New York: John Wiley.

IMAI, S., N. JAIN, AND A. CHING (2009): "Bayesian estimation of dynamic discrete choice models," *Econometrica*, 77(6), 1865–1899.

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

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

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

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

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

KHWAJA, A. (2010): "Estimating Willingness to Pay for Medicare Using a Dynamic Life-Cycle Model of Demand for Health Insurance," *Journal of Econometrics*, 156, 130–147.

LANCASTER, T. (2004): *An Introduction to Modern Bayesian Econometrics.* Oxford, UK: Blackwell Publishing.

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

Manchanda, P., and P. Chintagunta (2004): "Responsiveness of Physician Prescription Behavior to Salesforce Effort: An Individual Level Analysis," *Marketing Letters*, 15, 129–145.

Manchanda, P., P. Rossi, and P. Chintagunta (2004): "Response Modeling with Non-Random Marketing Mix Variables," *Journal of Marketing Research*, 41, 467–478.

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

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

Narayanan, S., P. Manchanda, and P. Chintagunta (2005): "Temporal Differences in the Role of Marketing Communication in New Product Categories," *Journal of Marketing Research*, 42, 278–290.

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

Orhun, Y. (2013): "Spatial Differentiation in the Supermarket Industry: The Role of Common Information," *Quantitative Marketing and Economics*, 11, 3–37.

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.

Pesendorfer, M., and P. Schmidt-Dengler (2010): "Sequential estimation of dynamic discrete games: A comment," *Econometrica*, 78, 833–842.

Poirier, D. (1998): "Revising Beliefs in Nonidentified Models," *Econometric Theory*, 14(4), 483–509.

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

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.

Rysman, M. (2004): "Competition Between Networks: A Study of the Market for Yellow Pages," *The Review of Economic Studies*, 71, 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 Before Demand Takes Off," *Management Science*, 56, 1259–1271.

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, 440–456.

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

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

WEINTRAUB, G., C. BENKARD, AND B. VAN ROY (2008): "Markov Perfect Industry Dynamics with Many Firms," *Econometrica*, 76, 1375–1411.

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

ZELLNER, A. (1971): *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley.

ZHU, T., AND V. SINGH (2009): "Spatial Competition with Endogenous Location Choices: An Application to Discount Retailing," *Quantitative Marketing and Economics*, 7, 1–35.

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. Model Estimates: Posterior Distribution

Parameter	Number of Potential Entrants (excluding "other" firms)	
	3 firms	4 firms
Cost: long run average (μ_c)	10.05	10.07
	(0.017)	(0.0014)
Cost: persistence (ρ_c)	0.9866	0.9873
	(0.00086)	(5.6e-05)
Cost: std. dev. (σ_c)	0.3721	0.3675
	(0.026)	(3.0e-04)
Cost spillover (κ_c)	0.06655	0.07067
	(0.0015)	(1.1e-04)
Revenue: mean (μ_r)	9.906	10.008
	(0.083)	(0.0037)
Revenue: std. dev. (σ_r)	1.591	1.682
	(0.060)	(0.0023)
Market share parameter (γ)	0.9375	0.9375
Discount factor (β)	0.9688	0.9688
Misclassification prob. (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 proportion of the cases where $A_{it}^o \neq A_{i,t,j}$ computed both by firm and overall.
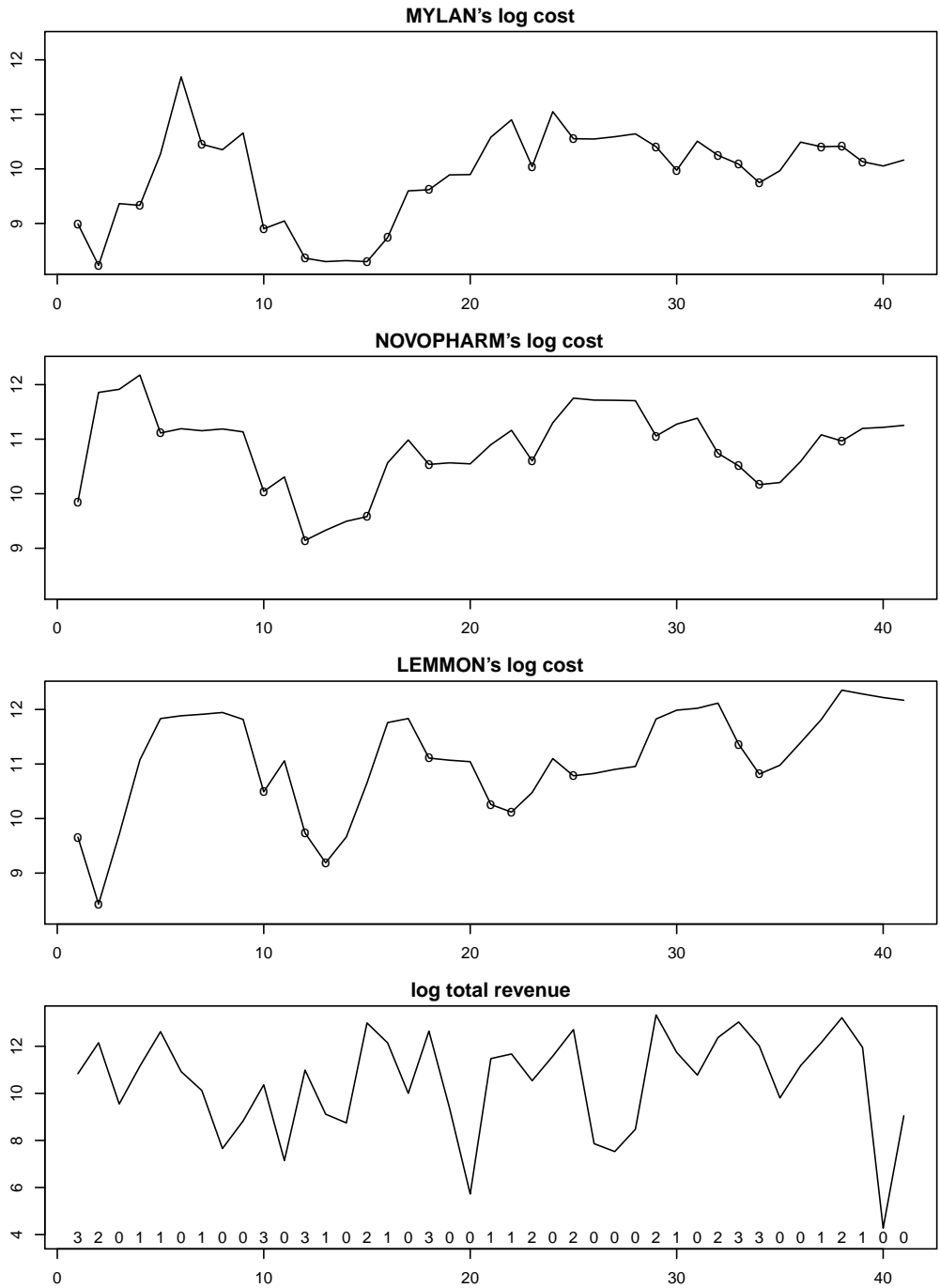
Figure 1. Cost, Revenue, and Entry Decisions. Plotted as a solid line in the first three panels is the logarithm of cost for the three dominant firms in the three firm model. The logarithm of cost is computed by averaging at Step 2e of the importance sampler at the maximum likelihood estimate. The circles 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.

37
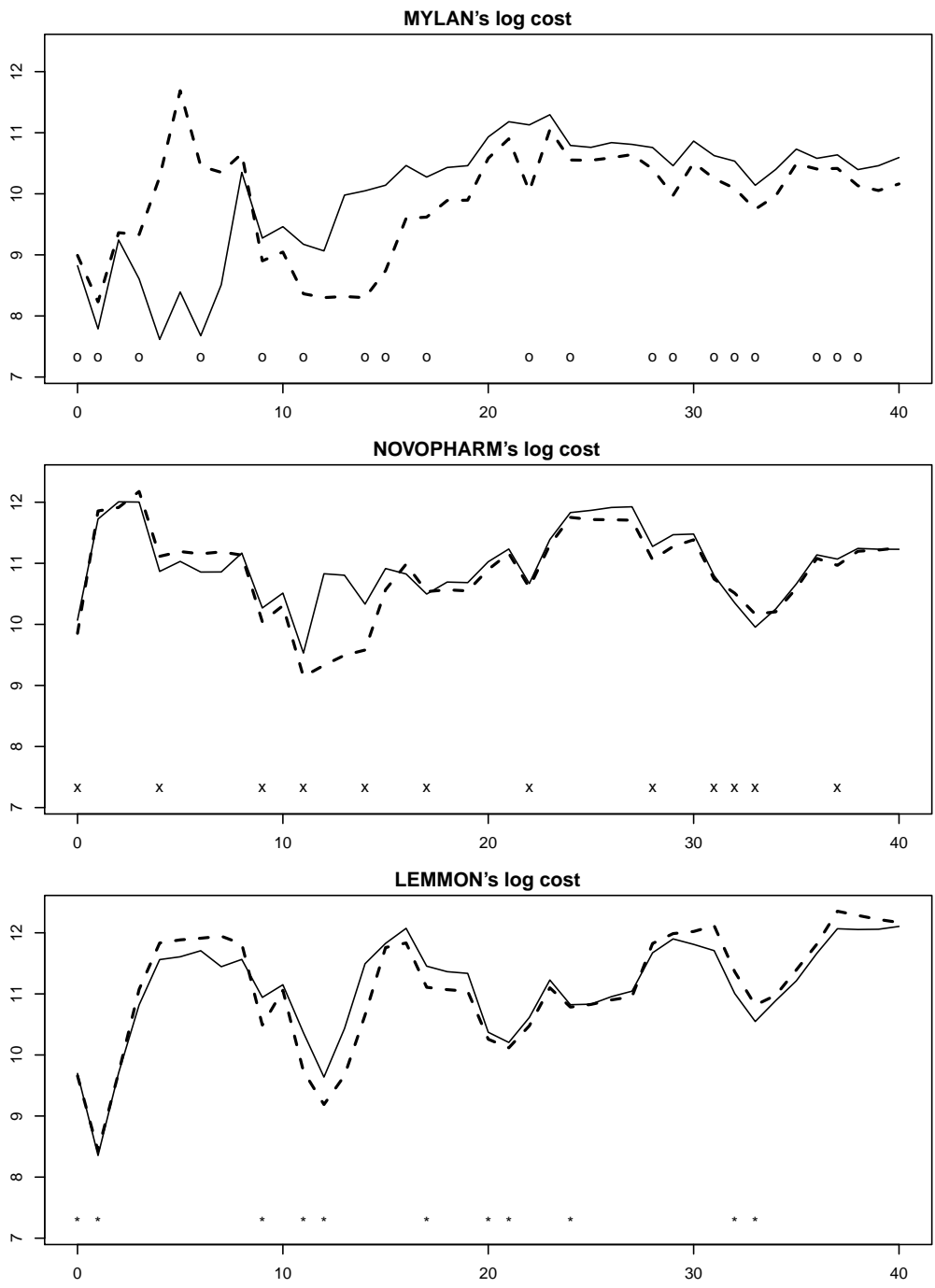
Figure 2. Cost and Entry Decisions of the Dominant Firms. Plotted is the logarithm of cost for the three dominant firms. The dashed line is under the three firm model, and the solid under the four firm model. The circles indicate the markets that Mylan entered, crosses the same for Novopharm, and the asterisks for Lemmon. The logarithm of cost as described in the legend of Figure 1.
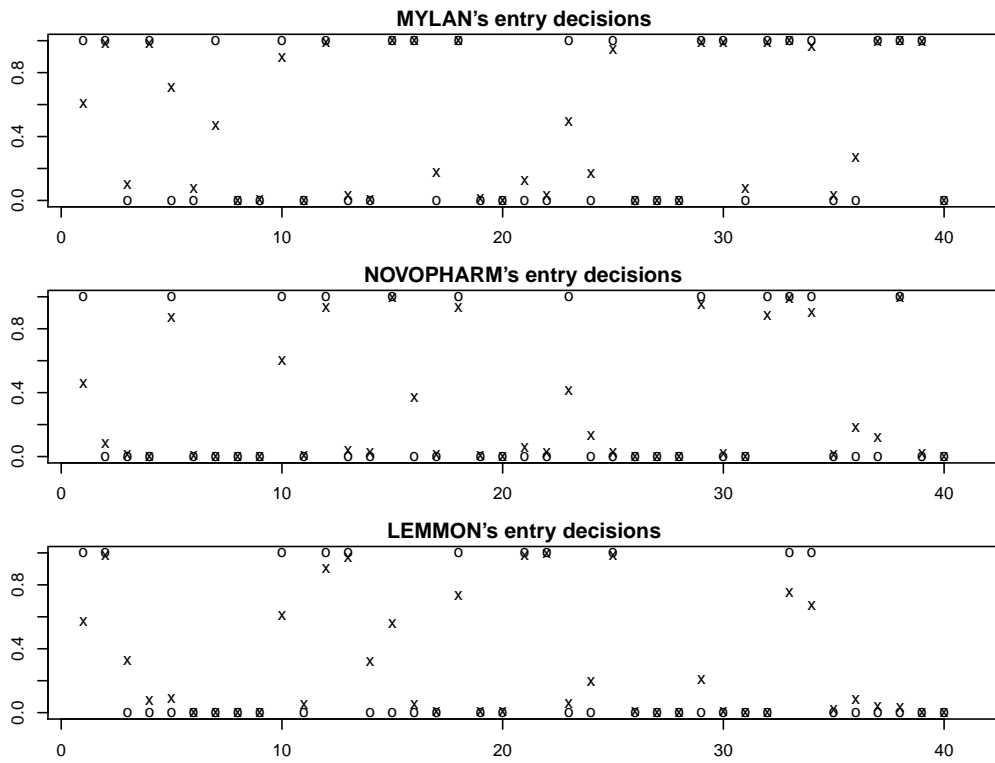
Figure 3. Actual and Predicted Entry Decisions. Plotted as circles are the entry decisions of the three dominant firms in the three firm model. The crosses are the average predictions of the three firm model computed by averaging game solutions at Step 2e of the importance sampler at the maximum likelihood estimate.

A Web Appendix

This is a Web Appendix to

> **"The Dynamic Spillovers of Entry: An Application to the Generic Drug Industry,"**

It is not meant for publication and is provided for review purposes. It will also be made available online and from the authors directly.

A.1 Solving the Model

When the state space can only take on a finite set of values, Theorem 3.1 of Dutta and Sundaram (1998) implies that the game described in Section 3 has a stationary Markov perfect equilibrium in mixed strategies. An intermediate step of the proof shows that the Bellman equation (12) has a fixed point. This result motivates our computational strategy because it implies that a solution to the game can be found by value function iteration. Solution methods that rely on computing a value function are discussed in Rust (1996).

Parthasarathy (1973) showed that Theorem 3.1 of Dutta and Sundaram (1998) holds for a countably discrete state space. Theorem 5.1 of Dutta and Sundaram (1998) is the extension to a continuous, bounded, state space. The equilibrium strategy profiles provided by Theorem 5.1 may depend on periods t and $t-1$ of the state vector. That does not exclude the possibility of a solution that depends only on the state at period t but does, in principle, obligate one to look over a larger domain if one cannot be found.

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 by iterating the Bellman equation (12) using a locally affine, step function to approximate the value function. This type of approximation can approximate any L_2 function to within arbitrary accuracy by taking the grid fine enough.

Strictly speaking, we do not have a finite state space that Theorem 3.1 of Dutta and Sundaram (1998) requires but we have never failed to compute an equilibrium due to nonconvergence. This is because the consequence of failure is that a particle gets killed. Based on the properties of the sequential importance sampler that particle is replenished at the resampling step of the particle filter algorithm described in Section 4. It is extremely rare that any of the N particles used to integrate the likelihood are killed. When it happens, there are no more than a few others, if any. We have never encountered a case where all N particles are killed. For a discussion of similar issues related to the complexity of guaranteeing existence and uniqueness of equilibrium in static games see Berry, Levinsohn, and Pakes (1995).[40]

---

[40]On p. 853, footnote 12 they state, "We assume that a Nash equilibrium to this pricing game exists, and that the

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

$$A_t = (A_{1t}, ..., A_{It}). \tag{22}$$

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 $(C_{1t}, ..., C_{It})$ and R_t. The vector of the log of the state variables at time t is

$$s_t = (c_{1t}, ..., c_{It}, r_t). \tag{23}$$

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. The grid increments are chosen to be (fractional) powers of two.[41] The centroid of the hyper-cube that contains a state vector s can be computed, element by element, by dividing by the increment, rounding to an integer, multiplying by the increment, and dividing by two. We use the centroid of each hyper-cube as the index K to the affine function whose domain is that hyper-cube. The rounding rules of the machine resolve which centroid 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.

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 centroid K and the value function at state s_t is represented by the affine function $V_K(s_t) = b_K + (B_K)s_t$. A value function $V_K(s_t)$ whose elements satisfy equation (12) is denoted $V_K^*(s_t) = b_K^* + (B_K^*)s_t$.

The game is solved as follows:

1. Given a state point s, get the centroid K that corresponds to it (we suppress the subscript t for notational convenience).[42]

2. Check whether the fixed point $V_K^*(s)$ of the Bellman equations (12) at this centroid 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.

---

equilibrium prices are in the interior of the firms' strategy sets (the positive orthant). While Caplin and Nalebuff (1991) provide a set of conditions for the existence of equilibrium for related models of single product firms, their theorems do not easily generalize to the multiproduct case. However, we are able to check numerically whether our final estimates are consistent with the existence of an equilibrium. Note that none of the properties of the estimates require uniqueness of equilibrium, although without uniqueess it is not clear how to use our estimates to examine the effects of policy and environmental changes."

[41] As described below, this is done so that the centroid has an exact machine representation. This facilitates efficient computation through compact storage of objects indexed by the centroid.

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

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 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).

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).[43]

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 centroid 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

---

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

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 $\mu_{I+1} = \mu_R$ 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 $\pi = 3.142$.[44] Then, using a $2L + 1$ rule,

$$\mathcal{E}[f(s_{t+1})\,|\,A_t, s_t] \approx \sum_{j_1=-L}^{L} \cdots \sum_{j_I=-L}^{L} \sum_{j_{I+1}=-L}^{L} f(\tilde{s}_{t+1,j_1}, \cdots, \tilde{s}_{t+1,j_I}, \tilde{s}_{t+1,j_{I+1}})\tilde{w}_{j_1}\cdots\tilde{w}_{j_I}\tilde{w}_{j_{I+1}}. \quad (24)$$

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.[45]

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

---

[44]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.

[45]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.

of two over the state space.[46] 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 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.[47] 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.[48] The easiest way to get a cluster of points $\{s_j\}$ about a centroid 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. 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).

A.2 Estimation Details

In this subsection we provide additional details about the estimation procedure that were not provided in the main paper. 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.

---

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

[47]The set of centroids 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.

[48]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.

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://www.aronaldg.org/webfiles/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 θ visited by the chain are restricted to (fractional) powers of two. We impose this restriction on the chain.

Histograms of the marginal posterior distributions are displayed in Figures 8 and 9 for the three and four firm cases, respectively.

Figure 8 about here

Figure 9 about here

Figures 10, 11, and 12 replicate Figures 1, 2, and 3 for the fully rational model, respectively.

Figure 10 about here

Figure 11 about here

Figure 12 about here

6

Appendix References

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

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

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.

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

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.

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

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

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

RUST, J. (1996): "Numerical dynamic programming in economics," *Handbook of computational economics*, pp. 619–729.
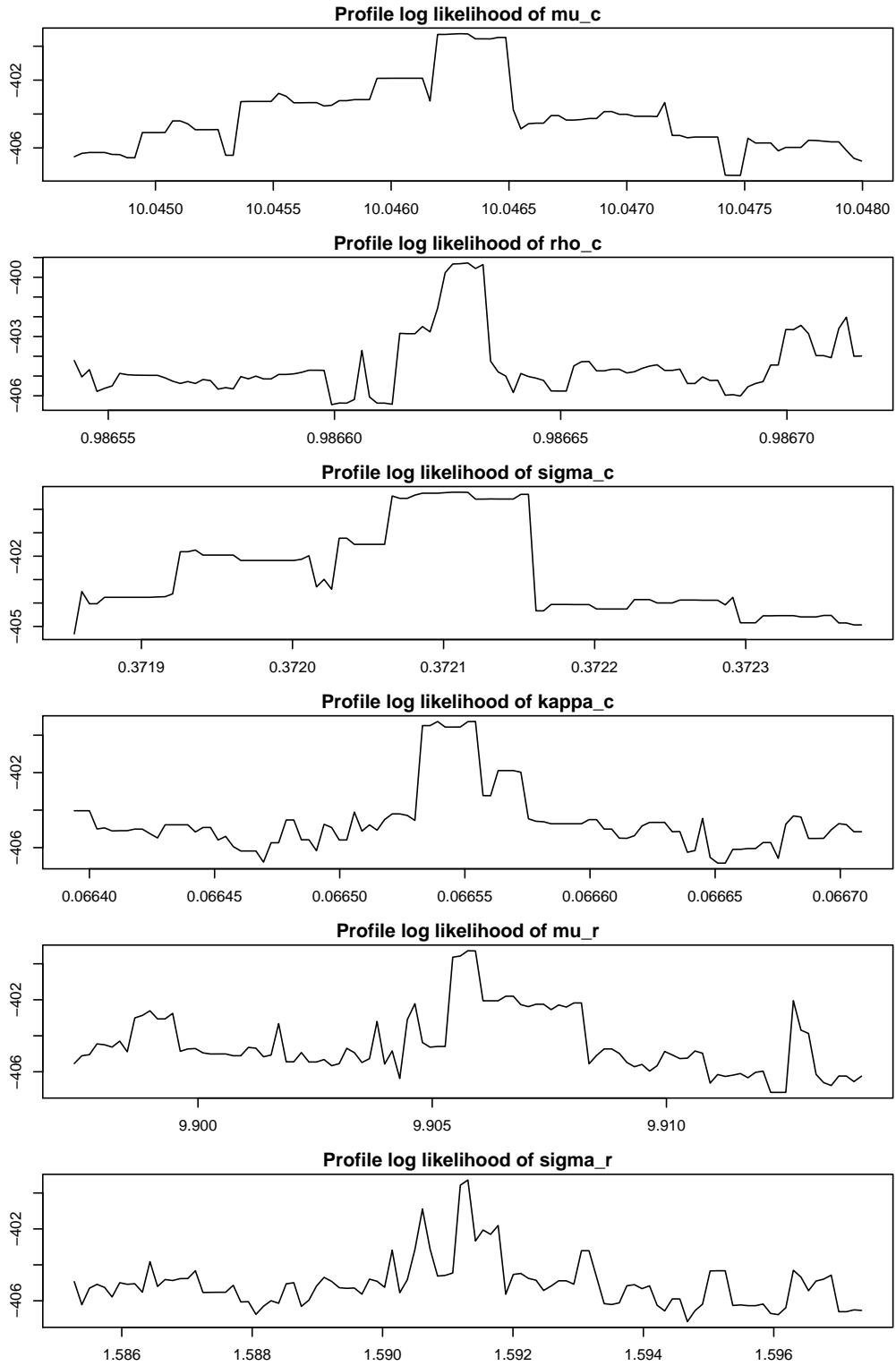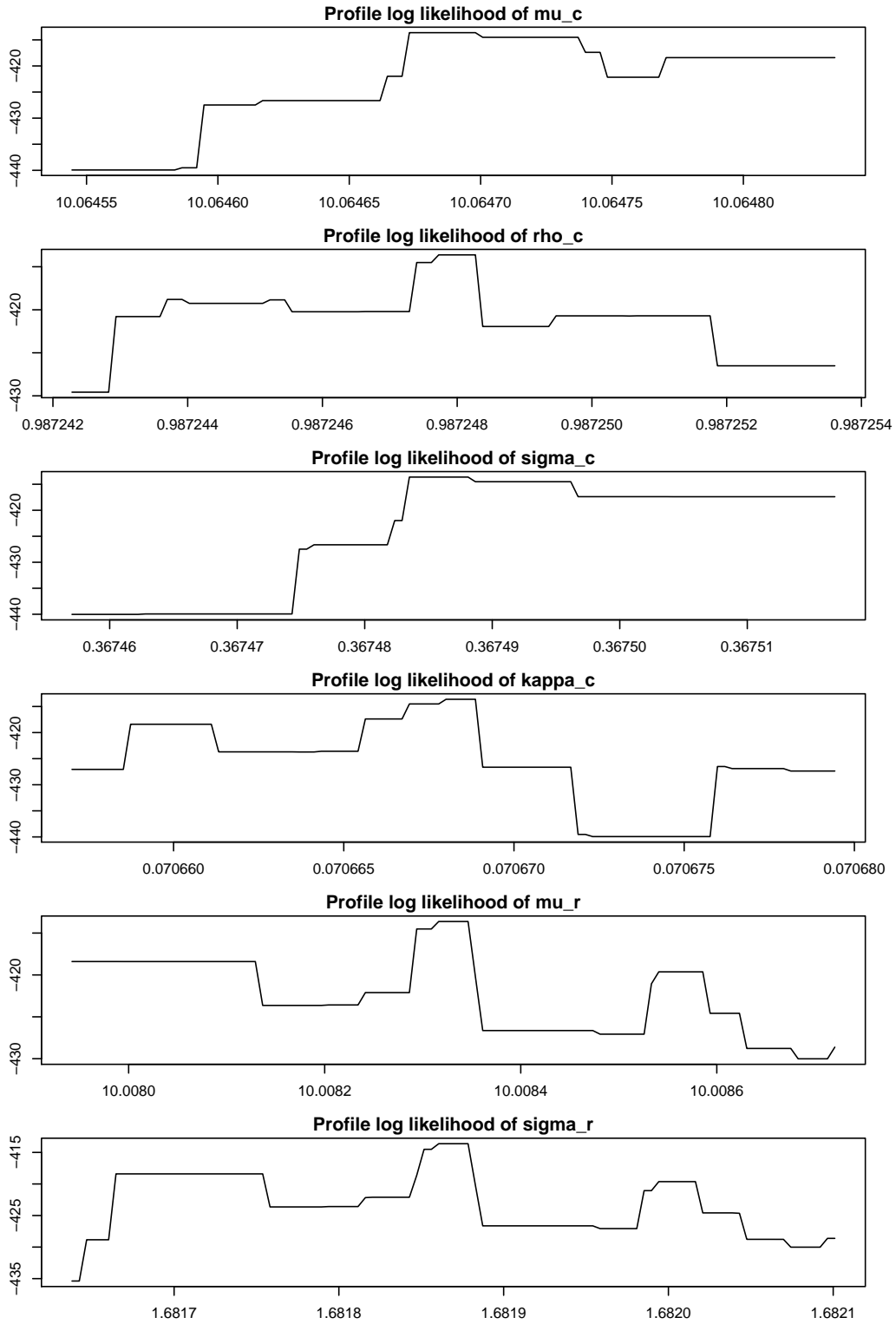
Figure 4. Profile Log Likelihood, Three Firm Model. Shown is the logarithm of the profile likelihood plotted for a tenth of the posterior standard deviation to the left and right of the maximum of the likelihood.
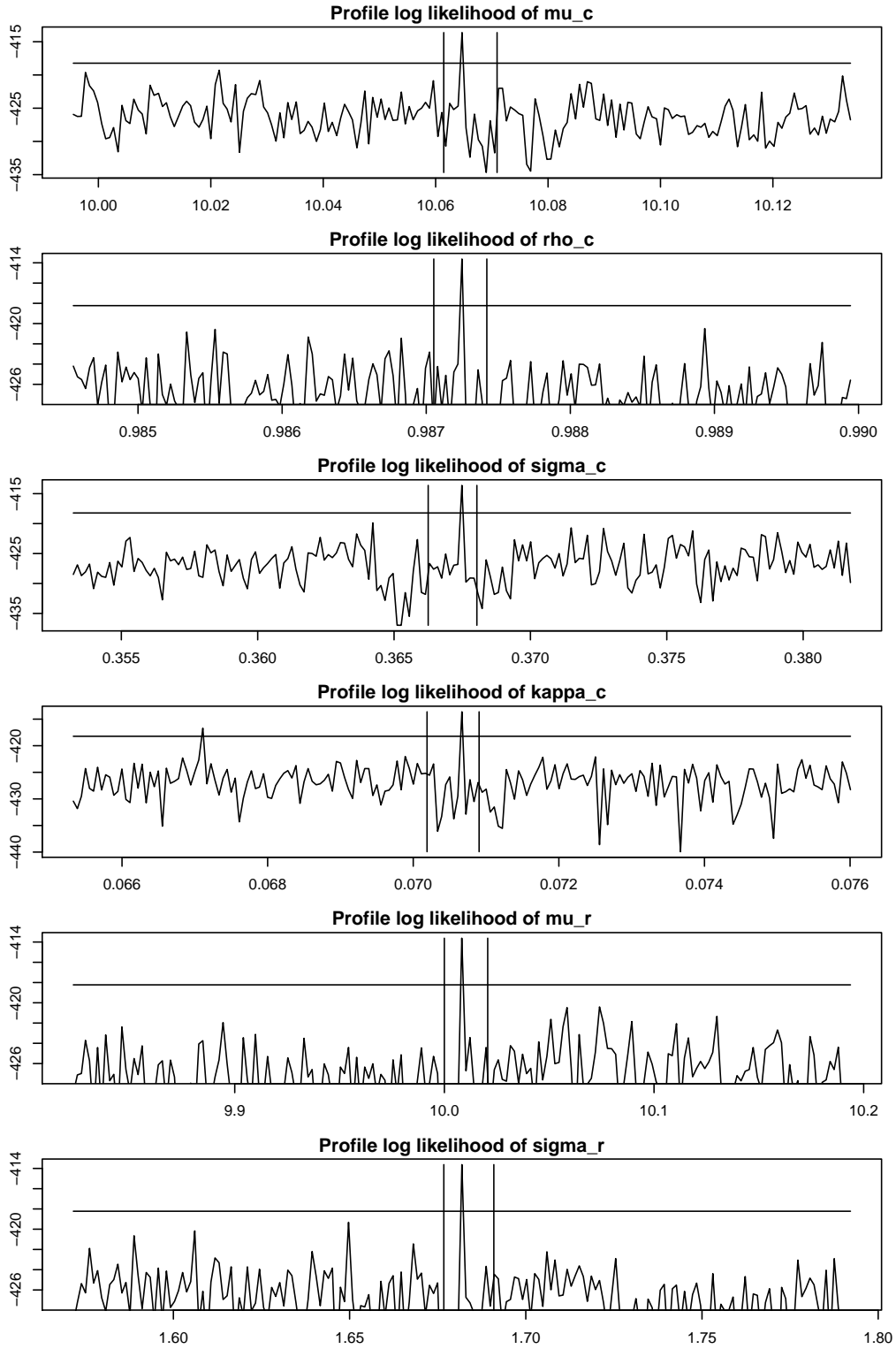
8
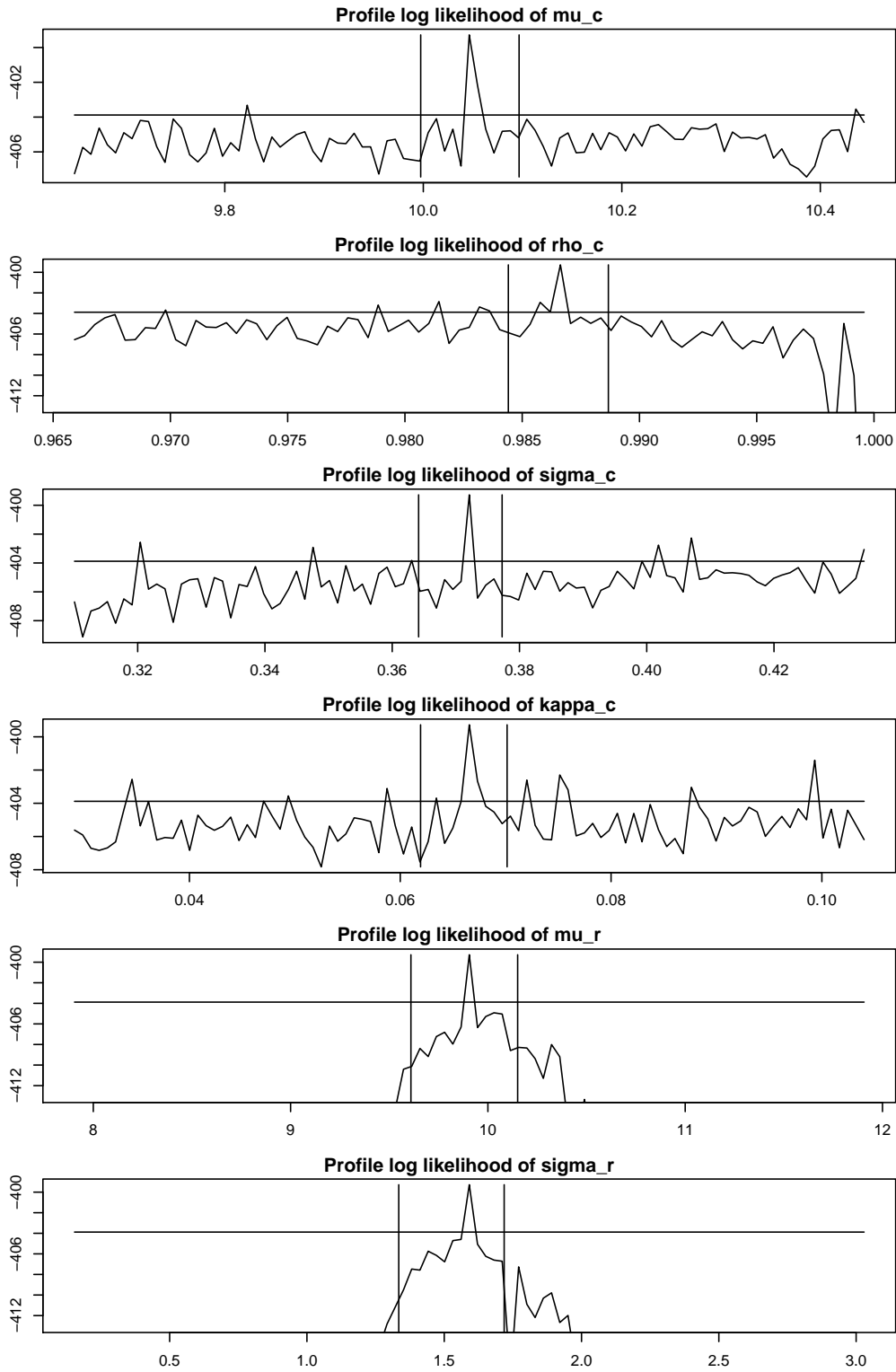
Figure 5. Profile Log Likelihood, Three Firm Model. Shown is the logarithm of the profile likelihood plotted for 24 posterior standard deviations to the left and right of the maximum of the likelihood. Points that violate support conditions and points below 10^{-6} of the maximum of the likelihood are 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.

9

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.

10

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.

11
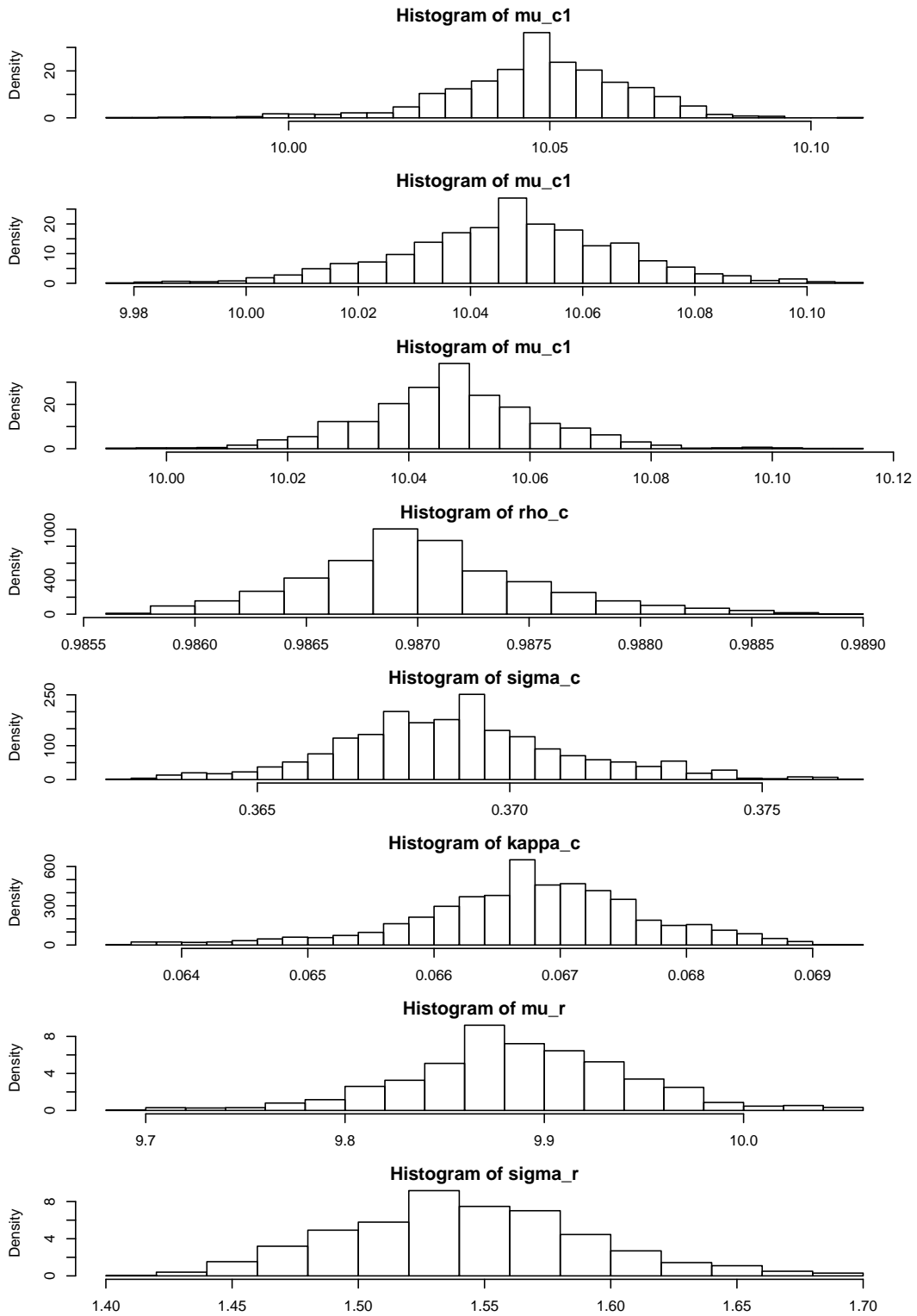
Figure 8. Marginal Posterior Distributions, Three Firm Model. Shown are histograms constructed from an MCMC chain for the three firm model with 3,000,000 repetitions at a stride of 375 for 8000 net. The salient feature of this graphic is the contrast of the histogram for the parameter κ_c compared to that shown in Figure 9.
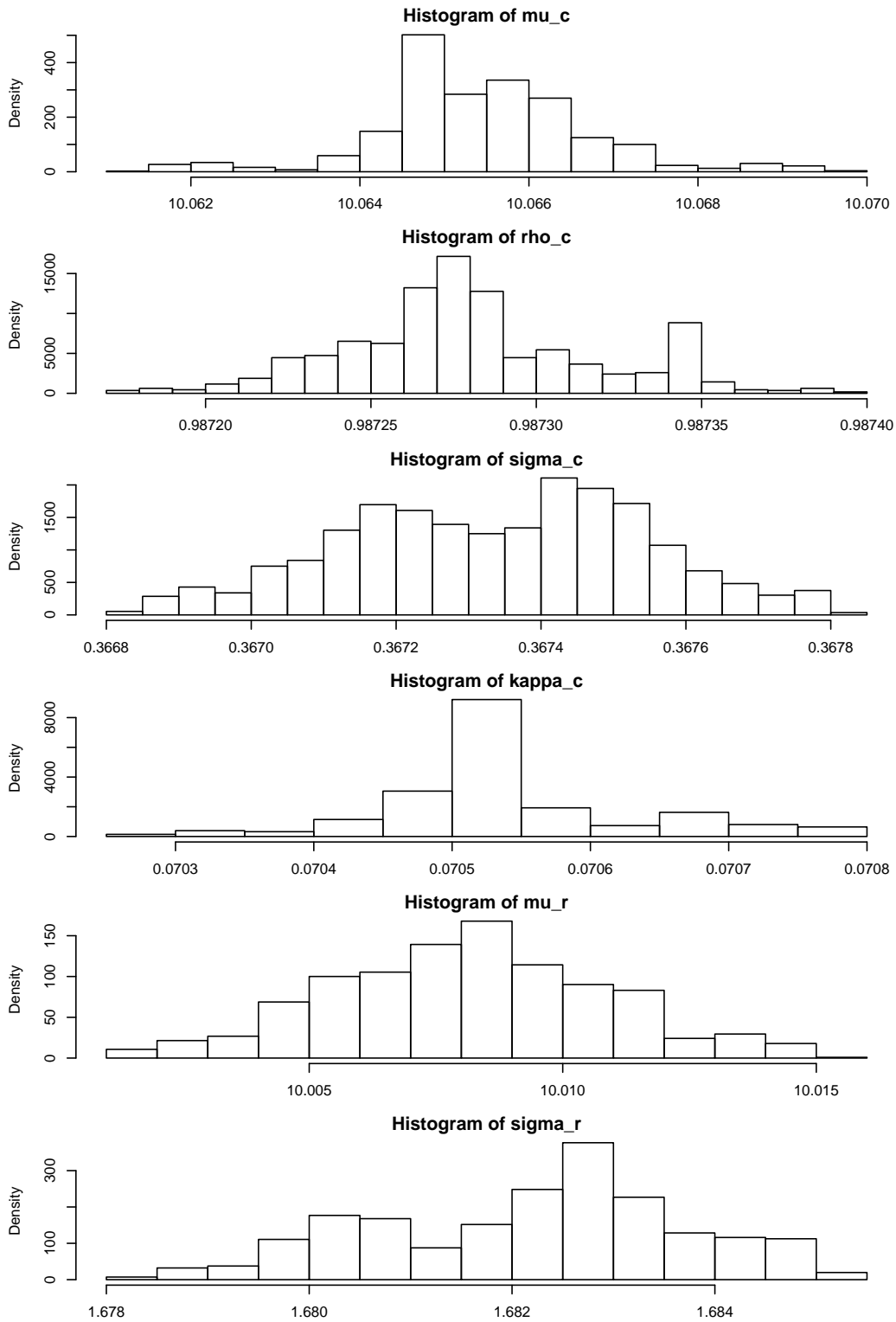
Figure 9. Marginal Posterior Distributions, Four Firm Model. Shown are histograms constructed from an MCMC chain for the four firm model with 3,000,000 repetitions at a stride of 375 for 8000 net. The salient feature of this graphic is the contrast of the histogram for the parameter κ_c compared to that shown in Figure 8.

13
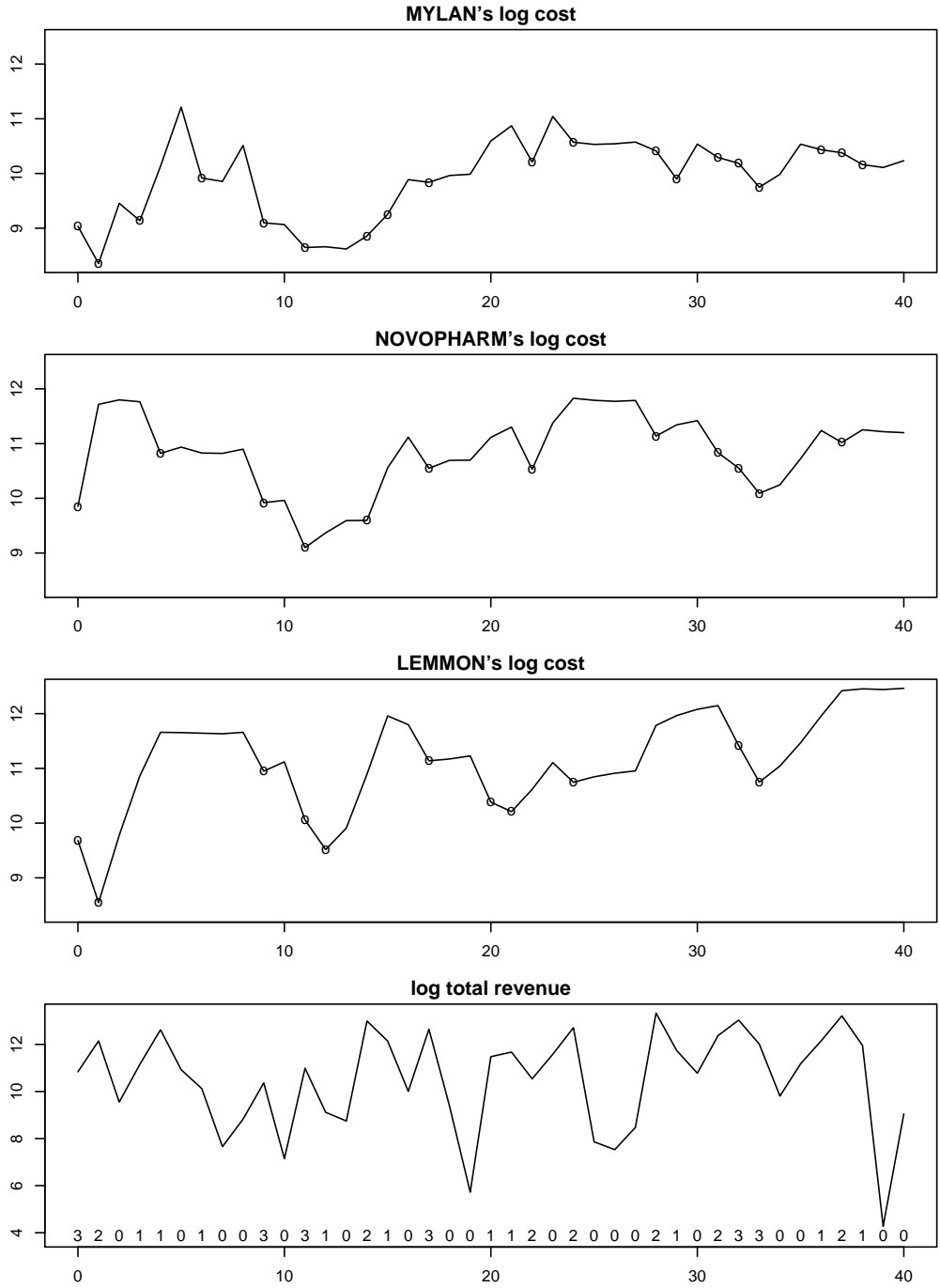
Figure 10. Cost, Revenue, and Entry Decisions. Plotted as a solid line in the first three panels is the logarithm of cost for the three dominant firms in the boundedly rational three firm model. The logarithm of cost is computed by averaging at Step 2e of the importance sampler at the maximum likelihood 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.

14
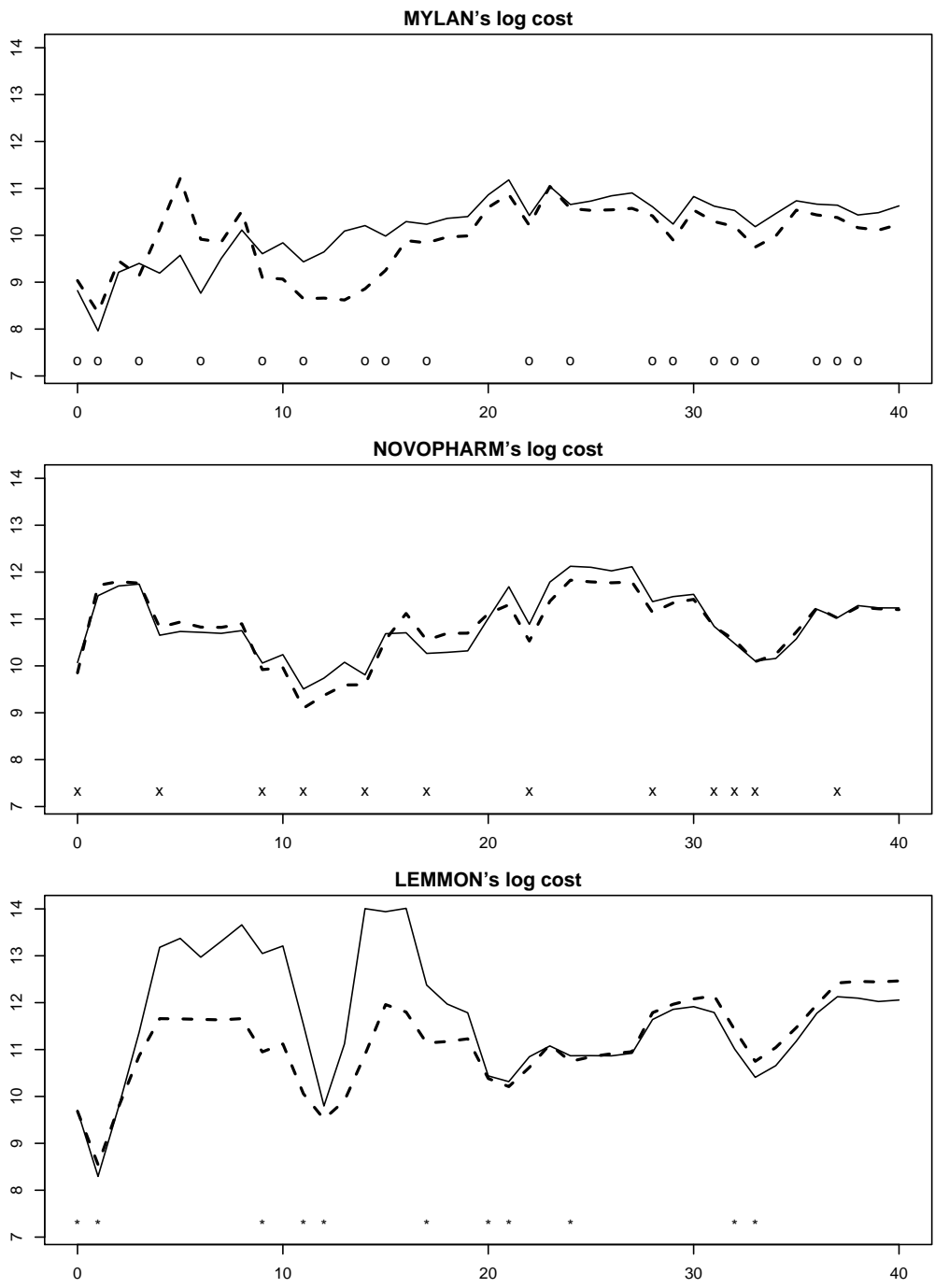
Figure 11. Cost and Entry Decisions of the Dominant Firms. Plotted is the logarithm of cost for the three dominant firms. The dashed line is under the boundedly rational three firm model, and the solid under the four firm model. The circles indicate the markets that Mylan entered, crosses the same for Novopharm, and the asterisks for Lemmon. The logarithm of cost as described in the legend of Figure 10.
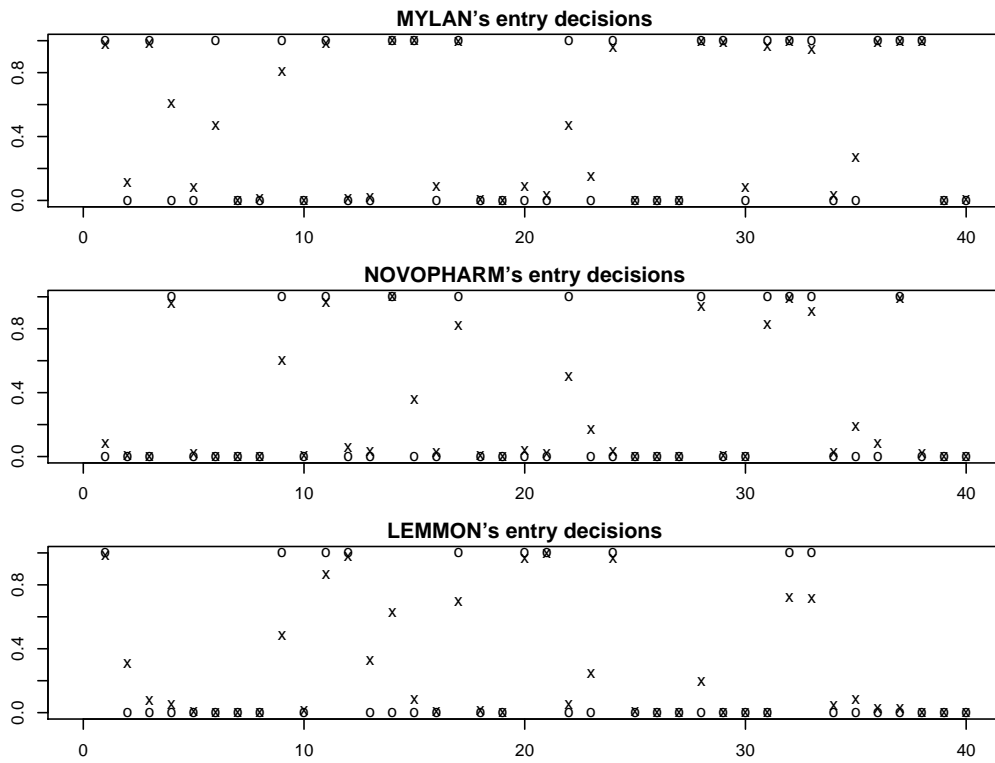
Figure 12. Actual and Predicted Entry Decisions. Plotted as circles are the entry decisions of the three dominant firms in the boundedly rational three firm model. The crosses are the average predictions of the three firm model computed by averaging game solutions at Step 2e of the importance sampler at the maximum likelihood estimate.

16