A Framework for Dynamic Oligopoly in Concentrated Industries*

Vivek Farias[†]

Bar Ifrach[‡]

Gabriel Y. Weintraub[§]

October, 2011

NEW VERSION COMING SOON

Abstract

We consider dynamic oligopoly models in the spirit of Ericson and Pakes (1995). We introduce a new computationally tractable model for industries with a few dominant firms and many fringe firms. This is a prevalent market structure in consumer and industrial goods. In our model, firms keep track of the detailed state of dominant firms and of few moments of the distribution that describes the states of fringe firms. Based on this idea we introduce a new equilibrium concept that we call *moment-based Markov equilibrium (MME)*. MME is behaviorally appealing and computationally tractable. However, MME can suffer from an important pitfall. Because moments may not summarize all payoff relevant information, MME strategies may not be optimal. We propose different approaches to overcome this difficulty with varying degrees of restrictions on the model primitives and strategies. Our first approach introduces models for which moments summarize all payoff relevant history and therefore for which MME strategies are optimal. The second approach restrict fringe firm strategies so that again moments become sufficient statistics. The third approach does not impose such restrictions, but introduces a computational error bound to assess the degree of sub-optimality of MME strategies. This bound allows to evaluate whether a finer state aggregation is necessary, for example by adding more moments. We provide computational experiments to show that our algorithms and error bound work well in practice for important classes of models. We also show that, cumulatively, fringe firms discipline dominant firms to behave more competitively, and that ignoring fringe firms in counterfactual analysis may lead to incorrect conclusions. Our model significantly extends the class of dynamic oligopoly models that can be studied computationally. In addition, our methods can also be used to improve approximations in other contexts such as dynamic industry models with an infinite number of heterogeneous firms and an aggregate shock; stochastic growth models; and dynamic models with forward-looking consumers.

*We have had very helpful conversations with Lanier Benkard, Allan Collard-Wexler, Dean Corbae, Boyan Jovanovic, Sean Meyn, Ariel Pakes, Ben Van Roy, Daniel Xu, as well as seminar participants at UT Austin, Maryland, Informs, IIOC, MSOM Conference, Econometric Society Summer Meeting, and Behavioral and Quantitative Game Theory Conference.

[†]MIT Sloan School, (vivekf@MIT.edu)

[‡]Columbia Business School, (bifrach14@gsb.columbia.edu)

[§]Columbia Business School, (gweintraub@columbia.edu)

1 Introduction

Dynamic oligopoly models offer a framework to analyzing intertemporal strategic interactions between competing firms. An important such example is the Ericson and Pakes (1995) (hereafter, EP) framework to modeling dynamic industries with heterogeneous firms. EP offers a tool to conduct empirical research and evaluate the effects of policy and environmental changes on market outcomes such as consumer welfare and market shares. The importance of evaluating policy outcomes in a dynamic setting and the broad flexibility and adaptability of the EP framework has spurred many applications in industrial organization, marketing, and operations.¹

Despite the broad interest in dynamic oligopoly models, there remain significant hurdles in applying them to problems of interest. Dynamic oligopoly models are typically analytically intractable, hence numerical methods are necessary to solve for the equilibrium. With recent estimation methods, such as Bajari et al. (2007), it is no longer necessary to solve for the equilibrium in order to structurally estimate a model. However, solving for Markov perfect equilibrium (MPE) is still essential to perform counterfactuals and evaluate environmental and policy changes. The practical applicability of EP-style models is severely limited by the ‘curse of dimensionality’ this computations suffers from. Methods that accelerate these equilibrium computations have been proposed (Judd (1998), Pakes and McGuire (2001) and Doraszelski and Judd (2006)). However, in practice computational concerns have typically limited the analysis to industries with just a few firms (say, two to six) which is far fewer than the real world industries the analysis is directed at. Such limitations have made it difficult to construct realistic empirical models.

Thus motivated, we propose a new computationally tractable model to study dynamic oligopoly. Our framework is suited to industries that have a few *dominant* firms with significant market shares and many *fringe* firms with small market shares. This market structure is prevalent in both consumer goods and intermediate products. Typical examples include industries in which there are few large national firms and many small local firms. Such industries are intractable in the standard EP framework. In these industries, although individual fringe firms have negligible market power, fringe firms may have significant cumulative market shares and may collectively discipline dominant firms’ behavior. Our model and methods capture this type of interactions and therefore significantly expand the set of industries that can be analyzed computationally.

In an EP-style model, each firm is distinguished by an *individual state* at every point in time. The value of the state could represent a measure of product quality, current productivity level, capacity, or other characteristic that determines the firm’s profitability. The *industry state* is a vector (or “distribution”) encoding the number of firms with each possible value of the individual state variable. Assuming its competitors follow a prescribed strategy, a given firm selects, at each point in time, an action (e.g., an investment level) to maximize its expected discounted profits; its subsequent state is determined by its current individual state, its chosen action, and a random shock. The selected action will depend in general on the firm’s individual state and the industry state. Even if firms were restricted to symmetric strategies, the computation entailed in selecting such an action quickly becomes infeasible as the number of firms and individual states grow.

¹See Doraszelski and Pakes (2007) for an excellent survey.

For example, in a model with 20 firms and 40 individual states more than 22 million gigabytes would be required just to store a strategy function. This renders commonly used dynamic programming algorithms to compute MPE infeasible in many problems of practical interest.

In this work we introduce a new model that is behaviorally appealing and overcomes the computational complexity involved in computing MPE. In a dominant/fringe market structure it is reasonable to expect that firms are more sensitive to variations in the state of dominant firms than those of individual fringe firms. In addition, it is unrealistic to believe that managers have unlimited capacity to monitor the evolution of all rival firms. Therefore, a plausible model of behavior is one where firms closely monitor rivals with the larger impact—the dominant firms—and monitor the remainder of the industry—the fringe firms—in a less detailed way. Thus motivated, we postulate that firms’ strategies depend on (1) the detailed state of dominant firms; and (2) few aggregate statistics of the fringe firms’ distribution (we refer to fringe firms’ state or distribution interchangeable). For example, firms could keep track of the first two moments of the fringe firm distribution, or of the first moment and a fractile related to the tail of that distribution. With some abuse of terminology, we call these strategies, *moment-based strategies*, with the understanding that firms could keep track of moments of the fringe firm distribution, but also of fractiles or other statistics. Based on these strategies, we introduce an equilibrium concept that we call *Moment-Based Markov Equilibrium (MME)*. An MME in which firms keep track of few moments of the fringe firm state is both computationally tractable and behaviorally appealing.

A natural question that arises is whether MME strategies are optimal in some meaningful sense. Our notion of optimality will consider that a MME strategy performs as well as a strategy that keeps track of the full fringe firm state when competitors use MME strategies. Unfortunately, it is simple to observe that MME strategies are not necessarily optimal. The reason is that even if moments of the fringe firm state are sufficient to predict static profits, they may not be sufficient statistics to predict the future evolution of the industry. For example, suppose firms keep track of the first moment of the fringe firm state. For a given value of the first moment, there could be many different fringe firm states consistent with it, from which the future evolution of the industry is very different. More formally, moments do not necessarily induce a sufficient partition of histories and do not summarize all payoff relevant history in the sense of Maskin and Tirole (2001); observing the full histogram of fringe firms provides valuable information for decision making. Technically, the issue that arises is that the stochastic process of moments may not be a Markov process even if the underlying dynamics are.

The previous arguments raise a concern regarding the optimality of moment-based strategies. Even if competitors use moment-based strategies, a moment-based strategy may not be close to a best response. Hence, MME strategies may not be close to a subgame perfect equilibrium. On one hand we have a behaviorally appealing and computationally tractable model, while on the other, the resulting strategies may not be optimal in any meaningful sense. We deal with this tension by proposing three alternative approaches; each places different restrictions on the model primitives and strategies being played and has different theoretical justifications.

First, we introduce classes of models for which equilibrium strategies yield moments that form a Markov process and hence summarize all payoff relevant information in a finite model (or as the number of fringe firms becomes large). In this case, an MME is a subgame perfect Nash equilibrium (or becomes subgame perfect as the number of fringe firms grows). While simple and elegant, the models impose relatively strong restrictions in the model primitives and may be too restrictive for some empirical applications.

In the second approach we impose less severe restrictions on the model, but instead restrict the strategies fringe firms can use. Under this restriction, we show that moments again form a Markov process as the number of fringe firms grows so that they become sufficient statistics to predict the industry evolution. We also provide a method to ex-post test how severe is our restriction on fringe firms' strategies. Note that we do not restrict the strategies of dominant firms.

In the third approach we do not restrict the model nor the strategies of fringe firms. Instead we assume that firms wrongly assume that moments form a Markov process that summarizes all payoff relevant information. In these models one postulates a Markov transition process for moments that approximates the (non-Markov) stochastic process of the industry moments; this can be done for example by using the empirical transition probabilities. We introduce a stochastic algorithm similar to Pakes and McGuire (2001) and Fershtman and Pakes (2010) to compute MME based on this idea.

In the third approach MME strategies will not be optimal, because moments are not sufficient statistics for the future evolution of the industry. To address this limitation, using ideas from robust dynamic programming, we propose a computationally tractable error bound that measures the extent of the sub-optimality of MME strategies in terms of a unilateral deviation. This bound is useful because it allows one to evaluate whether the state aggregation is appropriate or whether a finer state aggregation is necessary, for example by adding more moments.

We propose computationally efficient algorithms to compute our equilibrium concepts and show that they work well in important classes of models. We also conduct numerical experiments to show the practical applicability of our approach. We show how MME generate interesting strategic interactions between dominant and fringe firms. Notably, our results suggest that dominant firms could make investment decisions to deter growth and entry from fringe firms. For this reason, our results show that a common practice used by researchers to overcome the curse of dimensionality, namely to simply ignore the fringe, may lead to incorrect conclusions in a counterfactual analysis. Finally, we show how our robust error bound can be used to guide researchers in finding appropriate state aggregation architectures.

To further illustrate the applicability of our model and methods we show how they can be used to endogenize the market structure in a fully dynamic model. More specifically, we perform numerical experiments that are motivated by the long concentration trend in the beer industry in the US during the years 1960-1990. In the course of those years, the number of active firms dropped from about 150 to 30, and three industry leaders emerged. One common explanation of this trend is the emergence of national TV advertising as an "endogenous sunk cost" (Sutton, 1991). We build and calibrate a dynamic advertising model and use our methods to determine how a single parameter related to the returns to advertising expenditures critically

affects the resulting market structure and the level of concentration in the industry.

In summary, our approach offers a computationally tractable model for industries with a dominant/fringe market structure, capturing important and novel strategic interactions in industry dynamics. As such, our model greatly increases the applicability of dynamic oligopoly models and opens the door to studying novel issues in industry dynamics. Finally, while our main focus has been on dynamic oligopoly models, our methods can also be used in dynamic industry models with an infinite number of heterogeneous firms and an aggregate shock that are used in macroeconomics (see, for example, Khan and Thomas (2008) and Clementi and Palazzo (2010)). We specifically discuss the applicability of our results to this setting. In addition, we hope that our methods will be useful in other dynamic settings with heterogeneous agents such as dynamic models with forward-looking consumers and stochastic growth models in the spirit of Krusell and Smith (1998).

The rest of this paper is organized as follows. We discuss related literature in Section 2. Section 3 describes our basic industry model. Section 4 introduces our new equilibrium concept. Then, Sections 5, 6, 7, and 8 describe our three approaches to deal with the optimality of MME strategies. Section 9 concludes. All proofs are presented in the Appendix.

2 Related Literature

In this section we review related literature. First, we discuss different approaches and heuristics researchers have used in various applications to deal with the computational complexity involved in the equilibrium computation, and how our methods relate to those approaches. Second, we discuss other methods that are being developed to alleviate the computational burden involved in the equilibrium computation.

Researchers have used different approaches to practically deal with the computational burden involved in the equilibrium computation in applications. First, some papers empirically study industries that only hold few firms in which exact MPE computation is feasible (e.g., Benkard (2004), Ryan (2010), Collard-Wexler (2010a), and Collard-Wexler (2011)). Other researchers structurally estimate models in industries with many firms using approaches that do not require MPE computation, and do not perform counterfactuals that require computing equilibrium (e.g., Benkard et al. (2010), Sweeting (2007)). We hope that our work will provide a method to perform counterfactuals in concentrated industries with many firms.

In other applications, authors do perform counterfactuals computing MPE but in reduced size models compared to the actual industry. These models include few dominant firms and ignore the rest of the (fringe) firms (for example, see Ryan (2010) for an empirical application). Computational applied theory papers often also limit the industry to hold few firms (e.g., Besanko et al. (2010) and Doraszelski and Markovich (2007)). Other papers make different simplifications to reduce the state space. For example, Collard-Wexler (2010b) and Corbae and D’Erasmus (2011) assume firms are homogeneous so that the only relevant state variable is the number of active firms in the industry. Finally, some authors explicitly model heterogeneity but assume a simplified model of dynamics, in which certain process of “moments” that summarize industry

state information, is assumed to be Markov (e.g., see Jia and Pathak (2011), Santos (2010), and Tomlin (2008); Lee (2010) use a similar approach in a dynamic model of demand with forward looking consumers). We hope that our methods will help researchers determine the validity of these simplifications.

A stream of empirical literature related to our work uses simplified notions of equilibrium for estimation and to perform counterfactuals. In particular, Xu (2008), Qi (2008), Iacovone et al. (2009), and Thurk (2009) among others use the notion of oblivious equilibrium introduced by Weintraub et al. (2008), in which firms assume the average industry state holds at any time. OE can be shown to approximate MPE in industries with many firms by a law of large numbers, provided that the industry is not too concentrated. We hope that our methods extend this type of analysis for industries that are more concentrated.

Several methods to alleviate the computational burden in the EP framework have been proposed in the literature (Pakes and McGuire (2001), Judd (1998), and Doraszelski and Judd (2006)). Our work is particularly related to Benkard et al. (2011) that extends the notion of oblivious equilibrium to include dominant firms. In that paper, firms assume that at every point in time the fringe firm state is equal to the expected state conditional on the state of dominant firms. Hence, that paper offers a different way of keeping track of the fringe firm state compared to this paper, where firms keep track of moments. Our paper is also related to Farias et al. (2010) that uses approximate dynamic programming with value function approximation to approximate MPE.

Our moment based strategies are similar to Krusell and Smith (1998) that replace the distribution of wealth over agents in the economy by its moments when computing stationary stochastic equilibrium in a stochastic growth model. We hope that some of our ideas, in particular, our error bound based on robust dynamic programming, could guide researchers to find better approximations in other models with heterogeneous agents and aggregate shocks, such as stochastic growth models.

3 Dynamic Industry Model

In this section we formulate a model of industry dynamics with aggregate shocks in the spirit of Ericson and Pakes (1995). Similar models have been applied to numerous applied settings in industrial organization such as advertising, auctions, R&D, collusion, consumer learning, learning-by-doing, and network effects (see Doraszelski and Pakes (2007) for a survey). These models capture interactions among a finite number of firms, so we refer to them as *finite* models. At the end of the section we also comment how our basic model can be modified to incorporate an infinite number of firms in the spirit of Hopenhayn (1992). This model has been applied to several settings in macroeconomics such as business cycles and international trade policy. Because these models assume an infinite number of firms so we refer to them as *infinite* models.

Time Horizon. The industry evolves over discrete time periods and an infinite horizon. We index time periods with nonnegative integers $t \in \mathbb{N}$ ($\mathbb{N} = \{0, 1, 2, \dots\}$). All random variables are defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ equipped with a filtration $\{\mathcal{F}_t : t \geq 0\}$. We adopt a convention of indexing by t variables that are $\mathcal{F}_t$ -measurable.

Firms' Indices. Each firm that enters the industry is assigned a unique positive integer-valued index. The set of indices of incumbent firms at time t is denoted by S_t .

State Space. The heterogeneity of firms is captured by their states. Firms states may reflect quality level, productivity, capacity, the size of its consumer network, or any other aspect of the firm that affects its profits. At time t , the individual state of firm i is denoted by $x_{it} \in \mathcal{X} \subseteq \mathbb{R}^q$, $q \geq 1$. We define the *industry state* $\bar{s}_t$ to be a vector that encodes the individual states of all incumbent firms: $\bar{s}_t = (x_{it})_{i \in S_t}$. We define the state space $\bar{\mathcal{S}} = \{\bar{s} \in \mathcal{X}^n | n < \infty\}$.

Exit process. In each period, each incumbent firm i observes a nonnegative real-valued sell-off value ϕ_{it} that is private information to the firm. If the sell-off value exceeds the value of continuing in the industry then the firm may choose to exit, in which case it earns the sell-off value and then ceases operations permanently. We assume the random variables $\{\phi_{it} | t \geq 0, i \geq 1\}$ are i.i.d. and have a well-defined density function.

Transition dynamics. If an incumbent firm decides to remain in the industry, it can take an action to improve its individual state. Let $\mathcal{I} \subseteq \mathbb{R}_+^p$ ($p \geq 1$) be a convex and compact action space; for concreteness, we refer to this action as an investment. Given a firm's investment $\iota \in \mathcal{I}$ and state at time t , the firm's state at time $t + 1$ is given by the following transition kernel $\mathbf{Q}$:

$$\mathbf{Q}[x' | x, \iota, \bar{s}] = \mathcal{P}[x_{i,t+1} = x' | x_{it} = x, \iota_{it} = \iota, \bar{s}_t = \bar{s}]. \quad (1)$$

Uncertainty in state transitions may arise, for example, due to the risk associated with a research and development endeavor or a marketing campaign. The cost of investment is given by a nonnegative function $c(\iota_{it}, x_{it})$ that depends on the firm individual state x_{it} and investment level ι_{it} . Even though our approach can accommodate aggregate transition shocks common to all firms, for simplicity we assume that transitions are independent across firms conditional on the industry state and investment levels. We also assume these transitions are independent from the realizations of the sell-off values.

Aggregate shock. We do allow, however, for aggregate profitability shocks that are common to all firms. We denote the aggregate profit shock by z_t . These shocks may represent common demand shocks, a common shock to input prices, or a common technology shock. We assume that $\{z_t \in \mathcal{Z} : t \geq 0\}$ is an independent, finite, and ergodic Markov chain.

Single-Period Profit Function. Each incumbent firm earns profits on a spot market. For firm i , its single period expected profits $\pi(x_{it}, \bar{s}_t, z_t)$ depend on its individual state $x_{it} \in \mathcal{X}$, the industry state $\bar{s}_t \in \bar{\mathcal{S}}$, and the aggregate shock common to all firms $z_t \in \mathcal{Z}$. In addition, it is assumed that profits are bounded, i.e., there exists $\bar{\pi} < \infty$, such that $|\pi(x, \bar{s}, z)| \leq \bar{\pi}$, for all $x \in \mathcal{X}$, $\bar{s} \in \bar{\mathcal{S}}$, $z \in \mathcal{Z}$.

Entry process. Our approach can accommodate different entry processes. For concreteness, we assume that at each period there is a set of $M < \infty$ potential entrants that are assigned a unique positive integer-valued index. In each time period each potential entrant i observes a positive real-valued entry cost κ_{it} that is private information to the firm. If the entry cost is below the expected value of entering the industry then the firm will choose to enter. We assume the random variables $\{\kappa_{it} | t \geq 0, i \geq 1\}$ are i.i.d. and independent of all the previously defined random quantities.

Potential entrants make entry decisions simultaneously. Entrants do not earn profits in the period they decide to enter. They appear in the following period at state $x^e \in \mathcal{X}$ and can earn profits thereafter.² As is common in this literature and to simplify the analysis, we assume potential entrants are short-lived and do not consider the option value of delaying entry. Potential entrants that do not enter the industry disappear and a new generation of potential entrants is created in the next period.

Timing of Events. In each period, events occur in the following order: (1) Each incumbent firm observes its sell-off value and then makes exit and investment decisions; (2) Each potential entrant observes its entry cost and makes entry decisions; (3) Incumbent firms compete in the spot market and receive profits; (4) Exiting firms exit and receive their sell-off values; (4) Investment shock outcomes are determined, new entrants enter, and the industry takes on a new state $\bar{s}_{t+1}$.

Firms' objective. Firms aim to maximize expected net present value. The interest rate is assumed to be positive and constant over time, resulting in a constant discount factor of $\beta \in (0, 1)$ per time period.

Infinite models. An infinite model represents an asymptotic regime with an infinite number of firms. An industry state is represented by a measure over the Borel sets of $\mathcal{X}$. The state space is the space of all such measures. We assume the number of potential entrants is $M = \infty$. The rest of the model primitives are the same as in the finite model. Because of averaging effects across firms, conditional on the current value of the aggregate shock and the industry state, the next period's industry state evolves deterministically. Therefore, the only source of uncertainty in the infinite model is the aggregate shock.

Equilibrium. The most commonly used equilibrium concept in *finite* models is that of symmetric pure strategy Markov perfect equilibrium (MPE) in the sense of Maskin and Tirole (1988). Here, an incumbent firm uses a Markov strategy that depends on its own state x_{it} , the industry state $\bar{s}_t$, and the aggregate shock z_t to maximize expected discounted profits given the strategy of its competitors. Moreover, in equilibrium there is also an entry cut-off strategy that depends on the industry state $\bar{s}_t$ and the aggregate shock z_t . A limitation of MPE is that the set of relevant industry states grows quickly with the number of firms in the industry, making its computation intractable when there is more than few firms, even if one assumes anonymous equilibrium strategies (Doraszelski and Pakes, 2007).

Infinite models that consider a continuum of firms often consider *stationary equilibrium (SE)* (Hopenhayn, 1992), in which in the absence of aggregate shocks, the industry state is assumed to be constant over time due to averaging effects. This alleviates the curse of dimensionality MPE suffers from. In addition, Weintraub et al. (2008) introduce oblivious equilibrium (OE) in finite models, where firms optimize assuming that the industry state is constant over time and equal to its long-run expected value. As a result, OE shares the computational advantages of SE in infinite models. However, in the presence of dominant firms or aggregate shocks, whose actions or realizations might influence the evolution of other firms, responding to such a 'steady state' may be far from optimal. Computing MPE in such models when the number of firms is larger than, say, ten is not feasible. These limitations motivate our alternate approach.

²It is straightforward to generalize the model by assuming that entrants can also invest to improve their initial state.

4 Moment-Based Markov Equilibrium

In this section we introduce a new equilibrium concept that overcomes the curse of dimensionality mentioned above and that we think provides an appealing model of firms' behavior. Throughout this section it will be assumed that both aggregate shocks and dominant firms exist, although the model can accommodate only one of them as well. In fact, the distinction between aggregate shocks and dominant firms is of little importance at this stage, as an aggregate shock can be viewed as a dominant firm with an exogenous evolution.

4.1 Dominant and Fringe Firms

The prototypical industry that we have in mind exhibits the following market structure: there are few *dominant* firms and many *fringe* firms. Let $D_t \subset S_t$ and $F_t \subset S_t$ be the set of indices of dominant and fringe firms at time period t , respectively. The sets D_t and F_t are common knowledge among firms at every period of time. A simple version of our model assumes $D_t = D_{t'}$ for all t, t' , that is, the set of dominant firms is predetermined and does not change over time. A more general version incorporates a mechanism that endogenizes the process through which firms become dominant over time. We discuss specifications of such a process in Subsection 7.2.

The specific division between dominant and fringe firms will depend on the application at hand. We provide a natural example here. Suppose that a firm's individual state is a number that represents the quality of the product it produces, like in a quality ladder model (Pakes and McGuire, 1994). For many commonly used profit functions, such as those derived from random utility models, firms in higher states have larger market shares. It then may be natural to separate dominant firms from fringe firms by an exogenous threshold state $\bar{x}$, such that $i \in D_t$ if and only if $x_{it} \geq \bar{x}$. We let $\mathcal{X}_f \subseteq \mathcal{X} \times \{f\}$ and $\mathcal{X}_d \subseteq \mathcal{X} \times \{d\}$ be the set of feasible individual states for fringe and dominant firms, respectively. Note that the last component of the state encodes whether a firm is fringe or dominant. This will allow for example for fringe and dominant firms to have different model primitives and strategies. To simplify notation we re-define the state space $\mathcal{X}$ as $\mathcal{X} \times \{f, d\}$.

The previous discussion suggests that in the applications we have in mind, dominant firms are few and have significant market power. In contrast, fringe firms are many and individually hold little market power, although their aggregate market share may be significant. This market structure suggests that firms' decisions should be more sensitive to the dominant firms' states, compared with fringe firms' states. Moreover, the fringe firms' state is a highly dimensional object and gathering information on the state of each individual small firm is likely to be more expensive than on larger firms that not only are few, but also usually more visible and often publicly traded. Consequently, as the number of fringe firms grows large it is implausible that firms keep track of the individual state of each one. It seems more plausible that firms only keep track of the state of dominant firms and of some summary statistics of the fringe firms' state distribution. Not only do we think this provides a more appealing model of firms' behavior, but it will also make the equilibrium

computation feasible.

We introduce some useful notation. For convenience and because we will focus on equilibrium strategies that are anonymous with respect to the identity of fringe firms, we define the state of fringe firms f_t as a measure over $\mathcal{X}_f$. In a model with a finite number of firms this measure effectively encodes the states of each of the fringe firms. In an infinite model, this measure yields the ‘fraction’ of fringe firms at a given state $x \in \mathcal{X}_f$. We define $\mathcal{S}_f$ to be the set of all possible states of fringe firms, i.e., all sigma-finite Borel measures over $\mathcal{X}_f$. We call f_t the state of fringe firms and the distribution of fringe firms interchangeably.

We define d_t to be the state of dominant firms that specifies the individual state of each dominant firm at time period t . The set of all possible dominant firms’ states is defined by $\mathcal{S}_d = \mathcal{X}_d^{\overline{D}}$, where $\overline{D}$ is the maximum number of dominant firms the industry can accommodate. Generally, $\overline{D}$ will be determined as part of the equilibrium computation as we explain in the sequel.

With that in mind, we define the state space $\mathcal{S} = \mathcal{S}_f \times \mathcal{S}_d \times \mathcal{Z}$. Hence an industry state $s \in \mathcal{S}$ is now described as a distribution of fringe firms, a state for dominant firms, and the aggregate shock.

4.2 Assumptions

Our approach will require that firms compute best responses in strategies that depend only on a few summary statistics of the fringe firm state. To make this feasible, we will require that the firms’ single period profit function and transition function only depend on a few statistics of the fringe firm state. A set of such summary statistics is a multi-variate function $\theta : \mathcal{S}_f \rightarrow \mathbb{R}^n$. For example, $\theta(f) = \int_{\mathcal{X}_f} y^\alpha df(y)$ is the α moment with respect to the distribution f . For brevity and concreteness, we call such summary statistics *moments* with the understanding that they could include quantities other than just moments, such as fractiles of the distribution of states of fringe firms. We introduce the following assumption that we keep throughout the paper.

Assumption 4.1. *The single period expected profits of firm i $\pi(x_{it}, \tilde{\theta}_t, d_t, z_t)$ depend on its individual state $x_{it} \in \mathcal{X}$, a vector $\tilde{\theta}_t \in \mathbb{R}^l$ of fringe firm moments, the state of dominant firms $d_t \in \mathcal{S}_d$, and the aggregate shock $z_t \in \mathcal{Z}$. The evolution of firm i is Markov condition on its own state and investment, the state of the dominant firms, and an m -dimensional vector of moments. In particular, for an appropriate kernel (and with some abuse of notation) $\mathbf{Q}[\cdot|\cdot]$,*

$$\mathcal{P}[x_{i,t+1} = x | \mathcal{F}_t] = \mathbf{Q}[x | x_{it}, \iota_{it}, \hat{\theta}_t, d_t],$$

where $x_{it} \in \mathcal{X}$ and $\iota_{it} \in \mathcal{I}$ are the state and investment of firm i at time t , $\hat{\theta}_t \in \mathbb{R}^m$ is a vector of moments, and $d_t \in \mathcal{S}_d$ represents the state of the dominant firms.

In our approach, firms keep track of the moments that pin down the profit function and transition function $(\tilde{\theta}_t, \hat{\theta}_t)$. The state space spanned by these moments is much smaller than the original one if both l and m are low dimensional and significantly smaller than $|\mathcal{X}_f|$. In many applications of interest the transition function is independent of the industry state $(\hat{\theta}_t, d_t)$, e.g., if there are no spillovers in investments. Moreover, many

single profit functions of interest depend on few functions of the distribution of firms' states. For example, commonly used profit functions that arise from monopolistic competition models depend on a particular moment of that distribution (Dixit and Stiglitz (1977), Besanko et al. (1990)). A model of price taking fringe with convex costs, and strategic dominant firms with constants marginal costs that compete à la Cournot also gives rise to a profit function that depends on fringe firms only via few moments (see Appendix). We describe another important example below.

Example 4.1. Similarly to Pakes and McGuire (1994), we consider an industry with differentiated products, where each firm's state variable is a number that represents the quality of its product. Hence, a firm's state is given by the tuple (x_{it}, k) , where $x_{it} \in \mathbb{N}$ and $k = \{f, d\}$. We assume no spillover, i.e., $\mathbf{Q}[x'|x, \iota, \hat{\theta}, d] = \mathbf{Q}[x'|x, \iota]$.

There are m consumers in the market. In period t , consumer j receives utility u_{ijt} from consuming the good produced by firm i given by:

$$u_{ijt} = \alpha_1 \ln(x_{it}) + \alpha_2 \ln(Y - p_{it}) + \nu_{ijt}, \quad i \in S_t, \quad j = 1, \dots, m,$$

where Y is the consumer's income, and p_{it} is the price of the good produced by firm i . ν_{ijt} are i.i.d. random variables distributed Gumbel that represent unobserved characteristics for each consumer-good pair. There is also an outside good that provides consumers zero expected utility. We assume consumers buy at most one product each period and that they choose the product that maximizes utility. Under these assumptions our demand system is a classical logit model.

Let $N(x_{it}, p_{it}) = \exp(\alpha_1 \ln(x_{it}) + \alpha_2 \ln(Y - p_{it}))$. Then, the expected market share of each firm is given by:

$$\sigma(x_{it}, s_t, p_t) = \frac{N(x_{it}, p_{it})}{1 + \sum_{j \in S_t} N(x_{jt}, p_{jt})}, \quad \forall i \in S_t,$$

where aggregate shocks are not considered in this example for clarity. We define the "welfare index" $B(x_{it}, s_t, p_t) = 1 + \sum_{j \in S_t} N(x_{jt}, p_{jt})$.³ All firms share the same constant marginal cost c . We assume that dominant firms set prices in the spot market assuming their price change affects B . The first-order conditions for dominant firms are,

$$Y - p_{it} + \alpha_2(p_{it} - c)(\sigma(x_{it}, s_t, p_t) - 1) = 0, \quad \forall i \in D_t. \quad (2)$$

On the other hand, we assume that fringe firms set prices in the spot market assuming their price change has no effect on W , because their size is small relative to the aggregate.⁴ Hence, the first order conditions for fringe firms yield:

$$p^* = \frac{Y + c\alpha_2}{1 + \alpha_2}, \quad \forall i \in F_t. \quad (3)$$

³Note that in this model consumer surplus is equal to $m \log(B)$.

⁴Besanko et al. (1990) introduce a similar assumption in a logit model of monopolistic competition; the assumption is equivalent to assuming that fringe firms solve the dominant firms' first order conditions, but considering that their own market shares are zero when doing so.

Note that all fringe firms set the same price, which is independent of their own state and the industry state. Given this price, dominant firms set their own prices by solving (2). An application of Caplin and Nalebuff (1991) shows that there is a unique Nash equilibrium in pure strategies in this game. Expected profits are given by:

$$\begin{aligned}\pi(x_{it}, s_t) &= m(p_{it}^* - c)\sigma(x_{it}, s_{-i,t}, p_t^*), \forall i \in S_t \\ &= m(p_{it}^* - c) \frac{N(x_{it}, p_{it}^*)}{1 + \int_{y \in \mathcal{X}_f} y^{\alpha_1} df_t(y)(Y - p^*)^{\alpha_2} + \sum_{j \in D_t} N(x_{jt}, p_{jt})}\end{aligned}$$

Therefore, $\pi(x_{it}, s_t)$ can be written as $\pi(x_{it}, \tilde{\theta}_t, d_t)$, where $\tilde{\theta}_t$ is essentially the α_1 -th (un-normalized) moment of f_t .

4.3 Moment-Based Strategies

In this section we introduce firm strategies that depend on the individual states of dominant firms and on few summary statistics of the fringe firms' states. For example, in the setting of Example 4.1, it seems reasonable that firms keep track of their own individual state x_{it} , the state of dominant firms d_t , and the moment $\tilde{\theta}_t$ defined above (and z_t if there is an aggregate shock). We call such strategies *moment-based strategies*. These strategies depend on the distribution of fringe firms via a set of moments

$$\theta_t = \theta(f_t) = (\tilde{\theta}_t, \hat{\theta}_t, \bar{\theta}_t), \quad (4)$$

where $\tilde{\theta}_t$ and $\hat{\theta}_t$ are defined in Assumption 4.1, and $\bar{\theta}_t$ are additional moments included in θ_t . We define $\mathcal{S}_\theta$ as the set of admissible moments defined by (4). That is, $\mathcal{S}_\theta = \{\theta | \exists f \in \mathcal{S}_f \text{ s.t. } \theta = \theta(f)\}$. In light of this, we define the moment-based industry state by $\hat{s} = (\theta, d, z) \in \hat{\mathcal{S}} = \mathcal{S}_\theta \times \mathcal{S}_d \times \mathcal{Z}$. An investment strategy is a function ι such that at each time t , each incumbent firm $i \in S_t$ invests an amount $\iota_{it} = \iota(x_{it}, \hat{s}_t)$. Similarly, each firm follows an exit strategy that takes the form of a cutoff rule: there is a real-valued function ρ such that an incumbent firm $i \in S_t$ exits at time t if and only if $\phi_{it} \geq \rho(x_{it}, \hat{s}_t)$. Let $\mathcal{M}$ denote the set of exit/investment strategies such that an element $\mu \in \mathcal{M}$ is a pair of functions $\mu = (\iota, \rho)$, where $\iota : \mathcal{X} \times \hat{\mathcal{S}} \rightarrow \mathcal{I}$ is an investment strategy and $\rho : \mathcal{X} \times \hat{\mathcal{S}} \rightarrow \mathcal{R}$ is an exit strategy.

Each potential entrant follows an entry strategy that takes the form of a cutoff rule: there is a real-valued function λ such that a potential entrant i enters at time t if and only if $\kappa_{it} \leq \lambda(\hat{s}_t)$. We denote the set of entry rate functions by Λ , where an element of Λ is a function $\lambda : \hat{\mathcal{S}} \rightarrow \mathcal{R}$. It is assumed that all entrants are fringe, i.e., $x^e \in \mathcal{X}_f$. Note that strategies and the state space are defined with respect to a specific function of moments (4).

With Markov strategies (μ, λ) the *underlying industry state*, $s_t = (f_t, d_t, z_t)$, is a Markov process. We denote its transition kernel by $\mathbf{P}_{\mu, \lambda}$. We also define the kernel of the underlying industry state (x_{it}, s_t) when firm i uses strategy μ' , and its competitors use strategy (μ, λ) by $\mathbf{P}_{\mu', \mu, \lambda}$.

4.4 Moment-Based Markov Equilibrium

A *moment-based Markov equilibrium* (MME) is an equilibrium in moment-based strategies, as will be defined next. Before doing so, we notice, in agreement with our discussion above, that the notion of MPE will, in general, call for strategies that depend on the industry state s_t . This, however, is too complex from the computational perspective given the large number of fringe firms, and perhaps also from an informational perspective as discussed above.

Defining our notion of an equilibrium concept in moment based strategies will require the construction of what can be viewed as a ‘Markov’ approximation to the dynamics of the process $\{x_{i,t}, \hat{s}_t | t \geq 0\}$, where i is some generic firm. Notice that this process is, in general, *not* Markov even if the underlying dynamics are. To see this consider Example 4.1 when firms keep track of a single moment of the fringe firm state. For simplicity, assume that $\alpha_1 = 1$ and $(Y - p^*)^{\alpha_2} = 1$. Then, $\theta_t = \theta(f_t) = \int_{y \in \mathcal{X}_f} y df_t(y)$, so firms only keep track of the first (un-normalized) moment of the fringe firm state. Suppose $\theta_1 = 10$. This value of the first moment is consistent with one fringe firm in individual state 10, but also with 10 fringe firms in individual state 1. It is unclear that starting from these two different states will yield the same first moment next period. Therefore, while θ_t is sufficient to compute static profits, it may not be a sufficient statistic to predict the future evolution of the industry, because there are many fringe firm distributions that are consistent with the same value of θ_t . It is simple to observe that previous values of the moment may provide useful information to predict future values of the moment. In the process of aggregating information via moments, information is lost, and the resulting process is no longer Markov.

Assuming that firm i follows the moment based strategy μ' , and that all other firms use strategy (μ, λ) , we will describe a kernel, $\hat{\mathbf{P}}_{\mu', \mu, \lambda}[\cdot | \cdot]$ with the hope that the Markov process described by this kernel is a good approximation to the process $\{x_{i,t}, \hat{s}_t | t \geq 0\}$. To this end, let us suppose a kernel $\hat{\mathbf{P}}_{\mu, \lambda}$, describing the evolution of a hypothetical Markov process on $\mathcal{S}_\theta$ that we view as an approximation of the process $\{\hat{\theta}_t | t \geq 0\}$. Having supposed this kernel, the kernel $\hat{\mathbf{P}}_{\mu', \mu, \lambda}[\cdot | \cdot]$ is now define according to:

$$\hat{\mathbf{P}}_{\mu', \mu, \lambda}[x', \hat{s}' | x, \hat{s}] = \bar{\mathbf{P}}_{\mu', \mu, \lambda}[x', d', z' | x, \hat{s}] \hat{\mathbf{P}}_{\mu, \lambda}[\theta' | \hat{s}], \quad (5)$$

where $\bar{\mathbf{P}}_{\mu', \mu, \lambda}[x', d', z' | x, \hat{s}]$ denotes the marginal distribution of the next state of firm i , the next state of dominant firms, and the next value for the aggregate shock, conditional on the current moment-based state, according to the kernel of the underlying industry state $\mathbf{P}_{\mu', \mu, \lambda}$.

One may view the Markov process described by the above kernel as firm i 's *perception* of the evolution of its own state in tandem with that of the industry. As such, the definition above makes the following facts about this perceived process transparent:

1. Were firm i a fringe firm, the above definition asserts that this fringe firm ignores its own impact on the evolution of industry moments. This is evident in that x' is distributed independently of θ' given x and $\hat{s}$.
2. Given information about the current fringe moments, dominant firms' states, and the state of the

aggregate shock process, the firm correctly assesses the distribution of its next state as well as those of all other dominant firms. It also correctly assesses the distribution over values of the aggregate shock at the next period. Note that because firms use moment-based strategies, the moment-based state $(x, \hat{s})$ is enough to pin-down the transition probabilities of (x, d, z) according to the transition kernel of the underlying state $\mathbf{P}_{\mu', \mu, \lambda}$.

It should be clear that the Markov process given by the above definition remains an approximation since it posits that the evolution of the moments θ are Markov with respect to $\hat{s}$, whereas, in fact, the distribution of moments at the next point in time is potentially influenced by the distribution of the fringe beyond simply its moments. In the spirit of ‘approximating’ the actual moment process, however, we will eventually ask that the transitions described by $\hat{\mathbf{P}}$ agree in some manner with transitions observed in equilibrium. In particular, recall that $\mathbf{P}_{\mu, \lambda}$ denotes the kernel describing the evolution of the Markov process $\{s_t : t \geq 0\}$ when all firms use the strategy (μ, λ) . We will then specify $\hat{\mathbf{P}}_{\mu, \lambda}$ as some transformation of the kernel $\mathbf{P}_{\mu, \lambda}$. In particular,

$$\hat{\mathbf{P}}_{\mu, \lambda} = \Phi \mathbf{P}_{\mu, \lambda}$$

for some operator Φ . We next present a concrete example of such an operator:

Example 4.2 (An Example of Φ). *Let us denote by $q_{\mu, \lambda}(\cdot)$ the stationary distribution induced over $\mathcal{S}$ when all firms employ the strategy (μ, λ) . Moreover, with some abuse of notation we denote by $\mathbf{P}_{\mu, \lambda}(\theta_1 | \hat{s}_0)$ the conditional probability distribution induced assuming that the industry state at time 0 is distributed according to this stationary distribution and that all firms use strategy (μ, λ) . In particular:*

$$\mathbf{P}_{\mu, \lambda}(\theta_1 | \hat{s}_0) = \int_{\mathcal{A}(\hat{s}_0)} \mathbf{P}_{\mu, \lambda}(\theta_1 | s_0) dq_{\mu, \lambda}(s_0)$$

where $\mathcal{A}(\hat{s}_0) = \{s'_0 \in \mathcal{S} | (\theta(f'_0), d'_0, z'_0) = (\theta_0, d_0, z_0)\}$ is the set of industry states ‘compatible’ with the moment-based industry state $\hat{s}_0$. We will require that

$$\hat{\mathbf{P}}_{\mu, \lambda}(\theta_1 | \hat{s}_0) = (\Phi \mathbf{P}_{\mu, \lambda})(\theta_1 | \hat{s}_0) = \mathbf{P}_{\mu, \lambda}(\theta_1 | \hat{s}_0)$$

for all $\hat{s}_0 \in \mathcal{B}$, where $\mathcal{B} \subset \hat{\mathcal{S}}$ is some appropriate subset of moment-based industry states. $\hat{\mathbf{P}}_{\mu, \lambda}$ is defined arbitrarily outside this set.

Having thus defined a Markov process approximating the process $\{x_{i,t}, \hat{s}_t\}$, we next define the *perceived value function* by a deviating firm i when it uses the strategy μ' in response to an incumbent strategy (μ, λ) . Importantly, this value function is consistent with firm i ’s perception of the evolution of its own state and the moment-based industry state as described by the kernel $\hat{\mathbf{P}}_{\mu', \mu, \lambda}$ defined above. In particular, this value is given by

$$V(x, \hat{s} | \mu', \mu, \lambda) = E_{\mu', \mu, \lambda} \left[\sum_{k=t}^{\tau_i} \beta^{k-t} [\pi(x_{ik}, \hat{s}_k) - c(\iota_{ik}, x_{ik})] + \beta^{\tau_i-t} \phi_{i, \tau_i} \middle| x_{it} = x, \hat{s}_t = \hat{s} \right],$$

where τ_i is a random variable representing the time at which firm i exits the industry, and the subscripts of the expectation indicate the strategy followed by firm i , the strategy followed by its competitors, and the entry rate function. The expectation is taken with respect to the perceived transition kernel $\hat{\mathbf{P}}_{\mu', \mu, \lambda}$.⁵ In an abuse of notation, we will use the shorthand, $V(x, \hat{s}|\mu, \lambda) \equiv V(x, \hat{s}|\mu, \mu, \lambda)$, to refer to the expected discounted value of profits when firm i follows the same strategy μ as its competitors.

A moment-based Markov equilibrium (MME) is defined with respect to a function of moments θ in (4) and, for every strategy (μ, λ) , a transition kernel $\hat{\mathbf{P}}_{\mu, \lambda}$ defined via an operator Φ .

Definition 4.1. *An MME of our model comprises of an investment/exit strategy $\mu = (\iota, \rho) \in \mathcal{M}$ and an entry rate function $\lambda \in \Lambda$ that satisfy the following conditions:*

C1: Incumbent firm strategies optimization:

$$\sup_{\mu' \in \mathcal{M}} V(x, \hat{s}|\mu', \mu, \lambda) = V(x, \hat{s}|\mu, \lambda) \quad \forall x \in \mathcal{X}, \forall \hat{s} \in \hat{\mathcal{S}}. \quad (6)$$

C2: At each state, the cut-off entry value is equal to the expected discounted value of profits of entering the industry:

$$\lambda(\hat{s}) = \beta E_{\mu, \lambda} \left[V(x^e, \hat{s}_{t+1}|\mu, \lambda) \middle| \hat{s}_t = \hat{s} \right], \forall \hat{s} \in \hat{\mathcal{S}}.$$

C3: The perceived transition kernel is given by

$$\hat{\mathbf{P}}_{\mu, \lambda} = \Phi \mathbf{P}_{\mu, \lambda} \quad (7)$$

Existence of equilibrium will be discussed in each particular case we study below. With respect to uniqueness, in general we presume that our model may have multiple equilibria. Note that if the function θ is the identity, i.e. $\theta(f) = f$, and Φ is such that $\hat{\mathbf{P}}_{\mu', \mu, \lambda}[x', \hat{s}'|x, \theta(f), d, z] = \mathbf{P}_{\mu', \mu, \lambda}[x', \hat{s}'|x, f, d, z]$ then MME coincides with Markov perfect equilibrium.

Computationally, MME is appealing if agents keep track of few moments of the fringe firm state and there are few dominant firms. In this case, in a MME agents optimize over low dimensional strategies so it is a computationally tractable equilibrium concept. MME also provides an appealing behavioral model. For concreteness, in Example 4.1, as previously suggested, one could define $\theta(f_t) = (Y - p^*)^{\alpha_2} \int_{y \in \mathcal{X}_f} y^{\alpha_1} df_t(y)$, so firms would keep track of a single moment of the fringe firm state.

Theoretically, an MME is appealing if the perceived process of moments is close to the *actual* process of moments. This is related to the performance of moment-based strategies. For instance, is it enough to keep track of a single moment in the example above to derive close to optimal dynamic strategies? We study these questions in the next section.

⁵In the value function above, we have abused notation to denote $\pi(x_{ik}, \theta_k) = \pi(x_{ik}, \theta_k, d_k, z_k)$ instead of $\pi(x_{ik}, \tilde{\theta}_k, d_k, z_k)$. However, recall that $\tilde{\theta}_k$ is included in θ_k .

4.5 Moments and the Performance of Moment-Based Strategies

Suppose firms use moment-based strategies with moments θ_t and perceived transition kernel $\hat{\mathbf{P}}_{\mu,\lambda}$. We evaluate the performance of a MME strategy relative to a strategy that keeps track of the full fringe firm state f_t when dynamics are governed by the *primitive* transition kernel $\mathbf{P}_{\mu,\lambda}$. This comparison depends solely on the relationship between $\hat{\mathbf{P}}$ and $\mathbf{P}$, as given by the operator Φ . As previously noted, generally θ_t is not Markov, so it may not be a sufficient statistic to predict the future evolution of the industry. Hence, θ_t does not summarize all payoff relevant history in the sense of Maskin and Tirole (2001). As such, observing the full histogram of fringe firms f_t provides valuable information for decision making. It is important to note that while some Φ operators are better than others, except for special cases, moments are not sufficient statistics for *any* choice of Φ .

The previous arguments raise a concern regarding the performance of moment-based strategies. Even if my competitors use moment-based strategies, a moment-based strategy may not be close to a best response. Hence, MME strategies are not necessarily close to a subgame perfect equilibrium. In the rest of the paper we will deal with this tension: on one hand we have a behaviorally appealing and computationally tractable model, while on the other, the resulting strategies may not be optimal in any meaningful sense.

To deal with this tension we consider three approaches:

1. First, we consider a class of models for which equilibrium strategies yield moments that form a Markov process and hence summarize all payoff relevant information (in an infinite model or as the number of fringe firms becomes large). In this case, MME strategies are subgame perfect (or become subgame perfect as the number of fringe firms grows). Under general assumptions, we show that this class of models is the only one that possesses this nice property. See Section 5 for a discussion of this approach. While simple and elegant, the models discussed here may be too restrictive for many applications; this motivates the next approach.
2. In the second approach we explore the boundaries of the first approach by considering a restricted set of strategies for fringe firms. In this way we are able to obtain similar results to the first approach for a larger class of models. We also provide a method to ex-post test how severe is our restriction on fringe firms' strategies. Note that we do not restrict the strategies of dominant firms. See Section 6 for a discussion of this approach.
3. In the third approach we do not restrict the model. Instead we assume that firms wrongly assume that moments form a Markov process and summarize all payoff relevant information. The usefulness of this approach relies on a good choice of the observable set of moments and on the construction of Φ . We introduce a computationally tractable error bound that measures the extent of the sub-optimality of moment-based equilibrium strategies in terms of a unilateral deviation to a strategy that keeps track of all available information. The error bound is useful because it allows to assess how many and which moments need to be included in the strategies to achieve close to optimal strategies. See Section 7 for a discussion of this approach.

We conclude the section by formalizing the ideas discussed in this subsection. For simplicity, we define the concept of moments being sufficient statistics in the context of an infinite model.

Definition 4.2. *Moments are sufficient statistics in MME if*

$$\hat{\mathbf{P}}_{\mu,\lambda}[\theta'|\hat{s}] = (\Phi\mathbf{P}_{\mu,\lambda})[\theta'|\hat{s}] = \mathbf{P}_{\mu,\lambda}[\theta'|s] \quad (8)$$

for all $s \in \mathcal{S}$, where (μ, λ) are MME strategies and $\hat{s} = (\theta(f), d, z)$.

In the model with finitely many firms we say that moments become sufficient statistics if (8) holds in the limit as the number of fringe firms grows large.

Define the function of moments $\theta^*(f) = f$. Hence, a moment-based strategy with respect to θ^* is a Markov strategy that keeps track of the full fringe firm state. We denote $\mathcal{M}^*$ and Λ^* as the set of exit/investment strategies and entry rate functions, respectively, defined with respect to θ^* . Note that $\mathcal{M}^*$ and Λ^* are the set of standard Markov strategies. Similarly to the value function V defined above, we define a value function $V^*(x, s|\mu', \mu, \lambda)$ where transitions are assumed to be consistent with the primitive transition kernel $\mathbf{P}_{\mu',\mu,\lambda}$. Hence, $V^*(x, s|\mu', \mu, \lambda)$ is the expected net present value for a firm at state x when the industry state is s , given that its competitors each follows a common strategy μ , the entry rate function is λ , the firm itself follows strategy μ' , and transitions are governed by the kernel of the underlying industry state $\mathbf{P}_{\mu',\mu,\lambda}$. In words, V^* provides the *actual* expected discounted profits a firm would get in the industry.

For MME strategies (μ, λ) , define the *value of the full information deviation* by

$$\Delta_{\mu,\lambda}(x, s) = \sup_{\mu' \in \mathcal{M}^*} V^*(x, s|\mu', \mu, \lambda) - V^*(x, s|\mu, \lambda). \quad (9)$$

Note that we are not comparing the value of the optimal deviation to $V(x, \hat{s}|\mu, \lambda)$, since it is generally *not* the actual value of following strategy μ , as $\hat{\mathbf{P}}_{\mu',\mu,\lambda}$ is not the actual transition kernel. We will use the value of the full information deviation to measure the extent of sub-optimality of MME strategies. If this value is small, the MME strategy achieves essentially the same profits compared to the best possible Markov strategy. The three approaches below will address this notion of optimality in different ways.

5 First Approach: The Ideal Model

The first approach considers models in which moments summarize all payoff relevant information and the value the full information deviation is expected to be small. In order for this to hold, the Markov transitions of fringe firms, which are governed by their strategies, have to aggregate in a way that the current fringe firms' moments are almost sufficient statistics of the future evolution of the moments. This means that there exists a perceived transition kernel that approximates the actual transitions of moments well. In this section we present a subclass of industry dynamic model for which moments are close to sufficient statistics and the value of the full information deviation is zero, or asymptotically zero. We then show that under general

conditions this subclass is essentially the only one that possesses this property.

5.1 Constant Returns to Scale with Linear Dynamics

We take $\mathcal{X}_f = \mathbb{R}_+ \times \{f\}$ with a typical element $x = (x^1, f)$ and a fixed number of fringe firms $|F|$, where $F = F_t$ for all t . Entry and exit are treated in the sequel. The analysis below can be applied to single-period profit functions and transition kernels that depend on any integer moments of the fringe firm state. However, to simplify the exposition we assume that they both depend only on the first moment. A model like Example 4.1 with $\alpha_1 = 1$ would give rise to this type of profit function. Accordingly, we assume that firms keep track only of the first moment of the fringe firm histogram, i.e., $\theta_t = \int_{\mathcal{X}_f} x^1 df_t(x)$. We make the following assumptions on the primitives of fringe firms:

1. The single period profit is linear in the fringe firm's own state, $\pi(x, \hat{s}) = x^1 \pi_1(\hat{s}) + \pi_0(\hat{s})$. The assumption imposes constant returns to scale.
2. For a fixed state x , the cost function increases linearly with the investment level ι . In addition, the marginal investment cost increases linearly with state. Formally, $c(x, \iota) = (cx)\iota$ with $\iota \in \mathcal{I} = \mathbb{R}_+$.
3. The dynamics of a fringe firm's evolution are linear in its own state: $x_{it+1}^1 = x_{it}^1 \zeta_{1it}(\iota, \hat{s}_t) + \zeta_{0it}(\hat{s}_t)$, where ι is the amount invested. In addition, we assume the random functions ζ_{0it} and ζ_{1it} are independent of each other, and that $0 \leq \zeta_{0it} \leq \bar{\zeta}_0 < \infty$ and $0 \leq \zeta_{1it} \leq \bar{\zeta}_1 < \infty$ hold for all investment values and industry states. The assumption about linear transitions is similar to assuming Gibrat's law in firm's transitions (Sutton, 1997).

Importantly, note that the assumptions above do not place any restrictions on the primitives of dominant firms.

5.2 Moments Transitions

We begin with showing that for any perceived kernel $\hat{\mathbf{P}}_\mu$ the corresponding best response investment strategy for a fringe firm is linear in its state. In particular, we establish the following result:

Lemma 5.1. *For any perceived kernel, $\hat{\mathbf{P}}_\mu$, we have for a fringe firm in state x , that*

$$\sup_{\mu' \in \mathcal{M}} V(x, \hat{s} | \mu', \mu) = x^1 V_1(\hat{s}) + V_0(\hat{s})$$

for appropriate functions $V_1(\cdot)$ and $V_0(\cdot)$. Moreover, the best response strategy μ^* is independent of the fringe firms' state:

$$\mu^*(x, \hat{s}) = \mu_1(\hat{s})$$

for some appropriate function $\mu_1(\cdot)$.

Proof. Under the assumptions of model N of Chapter 9 in Bertsekas and Shreve (1978) (which are satisfied here assuming that π_1 and π_0 are non-negative) we have from Proposition 9.8 in the reference, that the optimal value function satisfies:

$$(TV)(x, \hat{s}|\mu) \triangleq \max_{\iota \in \mathcal{I}} \left\{ x^1 \pi_1(\hat{s}_0) + \pi_0(\hat{s}_0) - cx^1 \iota + \beta E_\mu \left[V(x_1, \hat{s}_1) \middle| \iota, \hat{s}_0 = \hat{s}, x_0 = x \right] \right\} = V(x, \hat{s}|\mu)$$

Moreover, we have that $T^n \hat{V} \rightarrow V$ if $\hat{V} = 0$ by Proposition 9.14.

It suffices to show that the posited form of the perceived value function is stable under an application of the Bellman operator. We have:

$$\begin{aligned} (T\hat{V})(x, \hat{s}|\mu) &= \max_{\iota \in \mathcal{I}} \left\{ x^1 \pi_1(\hat{s}) + \pi_0(\hat{s}) - cx^1 \iota + \beta E_\mu \left[(x^1 \zeta_1(\iota, \hat{s}_t) + \zeta_0(\hat{s}_t)) \hat{V}_1(\hat{s}_{t+1}) \right. \right. \\ &\quad \left. \left. + \hat{V}_0(\hat{s}_{t+1}) \middle| \iota, \hat{s}_t = \hat{s} \right] \right\} \\ &= x^1 \max_{\iota \in \mathcal{I}} \left\{ -c\iota + \beta E_\mu \left[\zeta_1(\iota, \hat{s}) \hat{V}_1(\hat{s}_{t+1}) \middle| \iota, \hat{s}_t = \hat{s} \right] \right\} + x^1 \pi_1(\hat{s}) + \tilde{V}_0(\hat{s}) \\ &= x^1 \tilde{V}_1(\hat{s}) + \tilde{V}_0(\hat{s}), \end{aligned}$$

where we define $\tilde{V}_0(\hat{s}) = \pi_0(\hat{s}) + \beta E_\mu \left[\zeta_0(\hat{s}_t) \hat{V}_1(\hat{s}_{t+1}) + \hat{V}_0(\hat{s}_{t+1}) \middle| \hat{s}_t = \hat{s} \right]$. Now, let us denote by $\hat{V}^n$ the iterates obtained by applying the Bellman operator T . Then, we have concluded that

$$x^1 \hat{V}_1^n(\hat{s}) + \hat{V}_0^n(\hat{s}) \rightarrow V(x, \hat{s})$$

But, since the above holds for at least two distinct values of x^1 for any given $\hat{s}$, this suffices to conclude that $\hat{V}_1^n(\hat{s}) \rightarrow \hat{V}_1^\infty(\hat{s})$ and $\hat{V}_0^n(\hat{s}) \rightarrow \hat{V}_0^\infty(\hat{s})$. This completes the proof of the first part of the Theorem.

Now, under the additional Assumption C of Chapter 4 in Bertsekas and Shreve (1978), and further assuming that the supremum implicit in the dynamic programming operator applied to V is attained for every $(x, \hat{s})$, the second claim follows immediately from the penultimate equality above and Proposition 4.3 of the reference. $\square$

We can now consider the evolution of the moment under a MME strategy μ . Note that by the previous lemma, in any MME fringe firm strategies must be independent of their own state. We begin with an assumption:

Assumption 5.1. *Suppose all firms use the MME strategy μ . Let $\bar{\mathcal{X}}_f \subset \mathcal{X}_f$ be compact. Then, for all $i \in F$ and for all t , $\mathcal{P}[x_{it} \in \bar{\mathcal{X}}_f] = 1$.*

While the previous assumption imposes conditions on equilibrium outcomes, it is quite natural in this context; MME is a sensible equilibrium concept only if fringe firms do not grow unboundedly large. We

have under this Assumption:

Proposition 5.1. *Suppose all firms play MME strategy μ and that Assumption 5.1 holds. For a given t , conditional on the realizations of $\{x_{it} \in \bar{\mathcal{X}}_f | i \in F\}$ and $\hat{s}_t$, we have*

$$1/|F| \sum_{i \in F} x_{it+1}^1 - \left[\tilde{\zeta}_1(\hat{s}_t) (1/|F|) \sum_{i \in F} x_{it}^1 + \tilde{\zeta}_0(\hat{s}_t) \right] \rightarrow 0, \text{ a.s.}$$

as $|F| \rightarrow \infty$, where $\tilde{\zeta}_1(\hat{s}_t) = \mathbb{E}[\zeta_{1,i}(\mu_1(\hat{s}_t), \hat{s}_t)]$ and $\tilde{\zeta}_0(\hat{s}_t) = \mathbb{E}[\zeta_{0,i}(\hat{s}_t)]$.

Proof. Fix an industry state $\hat{s}_t = (\theta_t, d_t, z_t)$. The evolution of moments is governed by,

$$\begin{aligned} \frac{1}{F} \sum_{i \in F} x_{it+1}^1 &= \frac{1}{F} \sum_{i \in F} [x_{it}^1 \zeta_{1,i}(\mu_1(\hat{s}_t), \hat{s}_t) + \zeta_{0,i}(\hat{s}_t)] \\ &\rightarrow \mathbb{E}[\zeta_{1,i}(\mu_1(\hat{s}_t), \hat{s}_t)] \sum_{i \in F} \frac{1}{|F|} x_{it}^1 + \mathbb{E}[\zeta_{0,i}(\hat{s}_t)] \\ &= \tilde{\zeta}_1(\hat{s}_t) \sum_{i \in F} \frac{1}{|F|} x_{it}^1 + \tilde{\zeta}_0(\hat{s}_t), \end{aligned} \tag{10}$$

where (10) follows, for instance, via the application of a Chernoff bound (employing Assumption 5.1) and the first Borel Cantelli Lemma. $\square$

The result shows that for large $|F|$, we have the following relation for the evolution of the normalized first moment:

$$\theta_{t+1} \approx \tilde{\zeta}_1(\hat{s}_t) \theta_t + \tilde{\zeta}_0(\hat{s}_t).$$

Thus motivated we define the perceived transition kernel⁶:

$$\hat{\mathbf{P}}_\mu[\tilde{\zeta}_1(\hat{s}_t) \theta_t + \tilde{\zeta}_0(\hat{s}_t) | \hat{s}_t] = 1 \tag{11}$$

Because in an infinite model a law of large numbers is assumed to hold exactly, the perceived transition kernel defined above will coincide with the actual transitions of the underlying industry state. In this case, the first moment is a sufficient statistic in MME and the value of the full information deviation is exactly zero.

Now, we consider finite models. Often in these models, the profit function depends on the un-normalized moment of the fringe firm state (see Example 4.1), so we assume firms keep track of the first un-normalized

⁶Note that a derivation similar to (10) will show that any k -th moment of fringe firms' states for an integer k would depend on moments $k, k-1, \dots, 1$ only as $|F|$ grows large. Therefore, if higher integer moments are payoff relevant they could be accounted for as well.

moment. Motivated by the previous discussion, we define the following transition kernel:

$$\hat{\mathbf{P}}_\mu[\tilde{\zeta}_1(\hat{s}_t)\theta_t + |F|\tilde{\zeta}_0(\hat{s}_t)|\hat{s}_t] = 1, \quad (12)$$

where here $\theta_t = \sum_{i \in F} x_{it}^1$. Proposition 5.1 suggests that when $|F|$ is large the first moment should become a sufficient statistic in MME. To formalize this notion we consider a sequence of industries index by $|F|$. To obtain meaningful models, we scale the market size proportionally to the number of fringe firms along this sequence. The proof is not presented in this draft.

5.3 Extensions and Special Cases

Entry and Exit So far we did not consider entry and exit in the constant returns to scale model. It is easy to add an entry process and still obtain the same results. However, with exit of fringe firms, moments are no longer sufficient statistics. The reason for this is that smaller fringe firms will tend to exit the industry more frequently since their continuation value from staying in the industry is lower than that of bigger fringe firms. Thus, two fringe histograms with the same moment value, one with many small firms and the other with mostly big firms will entail different evolution for the moments. If the exit process is such that the probability of exiting the market is independent of the fringe firm's own state, the previous result holds.

Homogenous Fringe A simple particular case of the linear model above assumes that all fringe firms are homogeneous. This can be done by setting $\zeta_1(\iota, \hat{s}) = 0$ and $\zeta_0(\hat{s}) = 1$ for all investment level and industry states. In this case the first moment of the fringe firm distribution would equal the number of fringe firms. This simple structure can accommodate entry and exit. Here firms only keep track of the number of fringe firms in the industry.

I.I.D. Fringe Suppose that the distribution of a fringe firm's state at time $t + 1$ is independent from its own state at time t . If that is the case, all incumbent fringe firms will invest the same amount. Consequently, if the number of fringe firms is large, the randomness in the transition of fringe firms will wash out due to a law of large numbers, and moments will be sufficient statistics. The linear model can accommodate this extension by setting $\zeta_1(\iota, \hat{s}) = 0$ for all investment levels and industry states, and allowing ζ_0 to depend on investment. Entry and exit could be incorporated to this model.

5.4 Necessary Conditions for Markov Moments

It is interesting to decompose the conditions that allow for this attractive result. The first necessary condition is that each fringe firm will have diminishing impact on its competitors as we increase the market size (or the number of fringe firms). The second necessary condition is that the dependence of fringe firms' equilibrium strategies on their own state will have some structure. Arbitrary dependence would mean that different histograms with the same moment value may have very different investment patterns. This hinders

aggregation of the individual fringe firms' transitions to moment transition, even if the number of fringe firms is large. To see this more clearly note that in order for the moment to be a sufficient statistic for the future evolution of moments, then at least the expected value of the next moment given industry state $\hat{s} = (\theta_t, d_t, z_t)$ should be the same for all histograms that are consistent with moment θ_t . This will be the case only if there is some pattern that relates investment with fringe firm's own state. In our previous exposition we had a linear relation.

If the set of moments contains only one moment, then under equilibrium strategies, fringe firms' transitions must be linear in order for moments to be Markov, as the next result shows.

Proposition 5.2. *Assume $\mathcal{X}_f = \mathbb{R}_+ \times \{f\}$ and that the set of moments contains the α moment only, i.e., $\theta(f) = \int_{x \in \mathcal{X}_f} x^\alpha df(x)$. If the moment is a sufficient statistic for the evolution of the industry, then fringe firms' transitions in MME are linear, $E_\mu[x_{i,t+1}^\alpha | x_{it} = x, \hat{s}_t] = x^\alpha \tilde{\zeta}_1(\hat{s}_t | \mu) + \zeta_0(\hat{s}_t | \mu)$.*

Proof. If $\alpha \neq 1$ redefine the state of fringe firms to be $y = x^\alpha$ (assume x_{it} is the state of a fringe firm throughout this proof). For moments to be sufficient statistics in the sense of Definition 4.2. It must be the case that the expected next moment $E[\theta(f_{t+1}) | f_t, \hat{s}_t]$ is independent of f_t for all $f_t \in \hat{\mathcal{S}}(\theta_t) = \{f \in \mathcal{S}_f | \int_{\mathcal{X}_f} x df(x) = \theta_t\}$. Therefore, $E[\theta(f_{t+1}) | f_t, \hat{s}_t] = \int E[x_{i,t+1} | x_{it} = x, \hat{s}_t] df_t(x) = \theta_{t+1}$ for all distribution f_t such that $\int_{\mathcal{X}_f} x df(x) = \theta_t$. In particular, treating $E[x_{i,t+1} | x_{it} = x, \hat{s}_t] \triangleq g(x)$ as a function of x , we require that under the measure f

$$E[g(X)] = \theta_{t+1}.$$

But considering the measure that places unit mass on $x = \theta_t$,

$$E[x_{i,t+1} | x_{it} = \theta_t, \hat{s}_t] = \theta_{t+1}$$

So that we have $g(E[X]) = E[g(X)]$ for all distributions f with mean θ_t . Since our choice of θ_t was arbitrary in $\mathcal{X}_f$, this suffices to establish the linearity of $g(\cdot)$. $\square$

Linear equilibrium transitions naturally emerge in the constant returns to scale. We are not aware of nonlinear models for which equilibrium strategies result in linear transitions.

6 Second Approach: Restricting Strategies

Motivated by the analytical results of the first approach, we explore its boundaries in order to relax the assumptions of the constant returns to scale model. To achieve this, we loosen the optimality requirement of fringe firms' strategies. However, no restrictions will be placed on the strategies or primitives of dominant firms. As such, this approach is mostly useful when dominant firms are the key focus of analysis. We note, however, that it is possible to check ex-post how far the restricted fringe strategies are from the optimal strategies. Subsection 6.1 provides a numerical experiment that shows that this approach captures interesting strategic interactions between the dominant and the fringe firms.

The idea of restricting agents' strategies is reminiscent of the idea of a representative agent in macroeconomics. In such models the evolution of the aggregate quantities in the economy (moments in our case) is derived from the decision of a single representative agent in a particular individual state. Typically, however, the optimal strategy would depend on the agent's individual state in a nontrivial way, and the transition laws of individual agents will not 'aggregate', i.e., it will not be the case that the evolution of some low dimensional statistic of the distribution of fringe firms will constitute a Markov process when augmented with the state of dominant firms and aggregate shocks. In fact, as far as we can tell, the only known case wherein such an aggregation is possible is when fringe firms have identical homothetic preferences, which in turn correspond to the linear dynamics of our first approach (see Hartley, 1997, for reference on representative agent in macroeconomics). We view our second approach as a means of obtaining an 'aggregate' transition law for the fringe at the cost of, in some sense, limiting the heterogeneity in fringe firms' strategies.

Suppose that the assumptions of constant returns to scale are suitable for a certain application, except that the single period profit function is not linear, for example $\pi(x, \hat{s}) = (x^1)^\alpha \pi_1(\hat{s}) + \pi_0(\hat{s})$ with $0 < \alpha < 1$. Then, the value function will not be linear in the fringe firm's state. As a result, the optimal investment strategy would not be as described in Lemma 5.1 and Proposition 5.1 will not hold. However, if we restrict fringe firms' strategies to be as in the result of Lemma 5.1, then we can 'aggregate' fringe firms in that it will suffice to track the appropriate moment of the fringe firms' state as in the first approach.

For some restriction on fringe firms' strategies that allows for the 'aggregation' alluded to above, denote by $\tilde{\mathcal{M}}$ the set of such restricted strategies with typical element $\tilde{\mu}$. For a dominant firm, $\tilde{\mu}$ is not restricted. Let $\hat{\mathbf{P}}_{\tilde{\mu}, \lambda}$ be the perceived transition kernel corresponding to the selection of some set of moments as a succinct description of the fringe firm state. There are many ways in which one could choose the restricted strategy to be played by the fringe. For example, one could follow the representative agent literature and take the strategy of the average agent, or take the restricted strategy to be the one that is closest, by some measure, to the optimal unrestricted strategies, as will be shown in an example.

We define equilibrium in this approach to be MME, except that the strategy that fringe firms play is a restricted one, and this restricted strategy is derived from their optimal unrestricted strategies. Concretely, let μ' be the optimal unrestricted strategy (in the sense of C1 in the definition of MME), when the remaining firms play $(\tilde{\mu}, \lambda)$. Then we say that $(\tilde{\mu}, \lambda)$ constitute an MME in this second approach if: (a) for dominant firms $\tilde{\mu} = \mu'$, (b) for fringe firms $\tilde{\mu}$ is derived from μ' (by a projection for example), (c) λ and $\hat{\mathbf{P}}$ satisfy C2 and C3, respectively, with Φ being appropriately defined given the restricted fringe firm strategies.

In the reminder of this section, we illustrate these ideas in a model with entry and exit. The state of a fringe firm is $(x^1, f) \in \mathcal{X}_f = [0, \bar{x}] \times \{f\}$. Fringe firms' transition dynamics are assumed to take the form,

$$x_{it+1}^1 = (x_{it}^1)^{(1-p)} \zeta_{it}(\iota_{it}), \quad (13)$$

where $0 < p < 1$ is assumed to be small, and $\zeta_{it}(\iota_{it})$ are random variables defined for every i, t and independent of all other random quantities, given ι_{it} . We omit the time and identity indexes of ζ when referring to its general properties. This functional form has a desirable property: a fringe firm's state at time

$t + 1$ is, ceteris paribus, increasing at a diminishing rate in the state at time t . We assume that $0 < \underline{\zeta} \leq \zeta \leq \bar{\zeta} < \infty$ so that a fringe firm cannot grow larger than $\bar{x} = (\bar{\zeta})^{1/p}$.

We restrict fringe firms' investment to be $\iota((x^1, f), \hat{s}) = \iota(\hat{s})$ and the investment cost $(cx)\iota$. Moreover, a fringe firm in state x stays in the industry with probability $(x^1/\bar{x})^p$. This is equivalent to the restriction $\rho(x, \hat{s}) = F_\phi^{-1}(x^p/\bar{x}^p)$, where F_ϕ is the cumulative distribution function of the sell-off values. With a common investment strategy, the transitions in (13) implies that for a small p the growth rate of fringe firms is close to $\zeta(\iota)$, that is the growth of fringe firms is close to proportional to their current state. Proportional and semi-proportional growth of firms is often referred to in the literature as Gibrat's law.⁷ Although Gibrat's law is disputed, previous work suggests that it is a good approximation for small firms. In addition, the probabilities of staying in the industry are increasing with the firm's own state, as is reasonable to assume, since the continuation value will generally be increasing with the fringe firm's own state.

The restricted investment strategies are chosen to be the best compromise among all fringe firms in different individual states. Specifically, given strategies $(\tilde{\mu}, \lambda)$, consider the optimal unrestricted strategy

$$\mu'(x, \hat{s}) = \operatorname{argmax}_{\mu \in \mathcal{M}} V(x, \hat{s} | \mu, \tilde{\mu}, \lambda)$$

for all $x \in \mathcal{X}_f$ and $\hat{s} \in \hat{\mathcal{S}}$. We take the best compromise strategy to be

$$\tilde{\mu}'(x, \hat{s}) \triangleq \tilde{\mu}(\hat{s}) = \int_0^{\bar{x}} \mu'((y, f), \hat{s}) dy.$$

Firms keep track of the first fringe moment only⁸. Entry takes a simpler form than that described in Section 3: in state $\hat{s}$ exactly $\lambda(\hat{s})$ firms enter (this is the limiting entry rate of the previously assumed entry strategy as the number of fringe firms grows large). Under some technical conditions, when the restricted model aggregates, we can obtain similar results to that of Proposition 5.1, as we now show.

We consider a sequence of industries indexed by the market size m (see Example 4.1) and an associated sequence of strategies $(\tilde{\mu}^m, \lambda^m)$. Similarly, we index the underlying industry state that corresponds to firms by m , (f_t^m, d_t^m) . We make the following assumption:

Assumption 6.1. *For all sequences of equilibrium strategies $(\tilde{\mu}^m, \lambda^m)_{m \geq 1}$ and for all t we have $|F_t^m| = \Theta(m)$ a.s., where F_t^m is the set of active fringe firms.*

This assumption is reasonable as it states that the profit collected by each active fringe firm remains stable as the market size grows large. Under this assumption we have:

Proposition 6.1. *Suppose all firms play equilibrium strategies $(\tilde{\mu}, \lambda)$. For a given t conditional on the*

⁷See Sutton (1997) for an excellent survey and Evans (1987) for an estimation of p .

⁸The following generalization is available: suppose that the profit function depends on the b -th moment and that firms stay in the industry with probability $(x^1/\bar{x})^q$ where $q = p/b$. Then the same restrictions on strategies will yield a Markov b -th moment as the number of fringe firms grows large.

realizations of $\{x_{it}|i \in F\}$ and $\hat{s}_t$, we have

$$(1/m) \sum_{i \in F_{t+1}^m} x_{i,t+1}^1 - (1/m) \left[\tilde{\zeta}^m(\hat{s}_t) \sum_{i \in F_t^m} x_{it}^1 / \bar{x}^p + \lambda(\hat{s}_t) x^e \right] \rightarrow 0 \text{ a.s.},$$

as $m \rightarrow \infty$, where $\tilde{\zeta}^m(\hat{s}_t) = \mathbb{E}[\zeta(\tilde{\mu}^m(\hat{s}_t))]$.

Proof. Fix $\hat{s}_t$. Recall that $0 \leq x_{it} \leq \bar{x} < \infty$. For $i \in F_t^m$ note that

$$\mathbb{E}_{\tilde{\mu}}[x_{i,t+1}^{1,m} | x_{it}^1, \hat{s}_t] = \tilde{\zeta}^m(\hat{s}_t) (x_{it}^1)^{(1-p)} \mathcal{P}[i \in F_{t+1}^m | x_{it}^1] = \tilde{\zeta}^m(\hat{s}_t) x_{it}^1 / (\bar{x})^p.$$

Now,

$$\sum_{i \in F_{t+1}^m} x_{i,t+1}^1 = \sum_{i \in F_t^m} (x_{i,t}^1)^{(1-p)} \zeta_{it}(\tilde{\mu}^m(\hat{s}_t)) \mathbf{1}_{i \in F_{t+1}^m} + \lambda(\hat{s}_t) x^e,$$

and so the result follows via the proof of Proposition 5.1 and using the fact that $|F_t^m| = \Theta(m)$ by assumption. $\square$

This result gives rise to the update equation for moments:

$$\theta_{t+1} \approx \mathbb{E}[\zeta(\tilde{\mu}(\hat{s}), \hat{s})] \theta_t / \bar{x}^p + \lambda(\hat{s}) x^e,$$

and we take

$$\hat{\mathbf{P}} = [\mathbb{E}[\zeta(\tilde{\mu}(\hat{s}), \hat{s})] \theta_t / \bar{x}^p + \lambda(\hat{s}) x^e | \hat{s}] = 1 \quad (14)$$

as in the first approach.

Similarly to the first approach, this result suggests that the value of full information deviation for dominant firms would converge to zero as the market size grows large. The key condition in proving this result is that the strategies are Lipschitz continuous with respect to the moment and to the market size. The proof of this result is not presented in this draft. For infinite models, just as in the first approach, the value of full information deviation is zero for dominant firms, since moments evolve deterministically according to (14). Since fringe firms play suboptimal strategies, we cannot obtain a similar result for them. It is possible, however, to measure ex-post the degree of sub-optimality of the restricted fringe strategy by solving a single agent DP.

Lastly, we comment here that there are no general theorems that could be applied to prove existence of MME either in this approach or in the first approach. The main challenge to prove such a result is that in these two approaches we consider uncountable state spaces. However, over a large set of primitives that satisfy the standard assumptions required for existence in EP-style models, we were always to computationally find an MME.

Table 1: Industry Averages

State of a dominant firm	9.15
State of a fringe firm	2.22
Number of fringe firms	13.5
First (un-normalized) moment	29.8
Number of fringe firms exiting/entering per period	.73
Size of exiting fringe firms	1.8

6.1 Numerical Experiments

We first solve for equilibrium of the model, then simulate an industry and report the industry statistics. To illustrate that the problem we are analyzing could not have been analyzed in a standard dynamic oligopoly framework, we report upfront that the average number of fringe firms in equilibrium is 21. Solving for MPE with this many fringe firms and additional dominant firms varies from computationally demanding to intractable, depending on the number of states fringe firms can take.

The algorithm we use is a combination of the algorithm in Pakes and McGuire (1994) for dominant firms and the algorithm in Weintraub et al. (2010) for fringe firms. It is useful to distinguish between the two in the algorithm since fringe firms do not internalize their effect on the industry state. The perceived transition kernel is the one following from (14).

We consider 2 dominant firms and a fringe tier. For simplicity we consider a Dixit-Stiglitz profit function of monopolistic competition $\pi(x_i, x_{-i}) = m \frac{x_i^b}{x_i^b + \sum_{j \neq i} x_j^b}$ with $b = 1$ where m is the market size (note that this profit function is very similar to Example 4.1). The transition probabilities for dominant firms are a generalization of that found in Pakes and McGuire (1994) and are given in detail in the Appendix as well as a list of parameters. We assume the identity of dominant firms does not change over time.

We compute MME and simulate 10,500 periods and remove the first 500 periods. Table 1 summarizes some industry averages. The industry statistics go in the direction one would expect. For example, the average fringe state is higher than the entry state. Some of the strategic interaction between the fringe firms and dominant firms is captured in figure 1. The figure shows that on average the higher the state of dominant firms the lower the un-normalized moment of fringe firms, and consequently the lower the cumulative size of fringe firms (specifically, we vary the size of one dominant firm when the other dominant firm state is held constant). Because fringe firms' spots profit are decreasing with the state of dominant firms, entry and investment are less profitable for fringe firms the higher the state of the dominant firm. This suggests that dominant firms invest to deter entry and investment from the fringe tier.

We also compare MME with an EP-style equivalent model with no fringe firms. Ignoring fringe firms is a common practice in the applied literature to simplify computation. In order to make the comparison fair we normalize the profit function in the EP model by fixing the fringe firms' moment to its average state from the MME simulation, which is 30. We compute the MPE of the "normalized" EP model and simulate

Figure 1: Conditional Moments (Big dominant firm at maximal state, 11)

the industry. The results show that the average dominant firm state decreases to 6.25 from the MME value, 9.15. This suggests that deterring entry in the fringe tier and pushing down the fringe firms' investment are key determinants in dominant firms' investment incentives. Moreover, ignoring fringe firms may bias downwards the investment efforts exerted by dominant firms. The collective presence of fringe firms, in spite of their weak individual market power, disciplines dominant firms and forces them to invest more than in the duopoly case. We conclude that explicitly modeling fringe firms may have important effects on conclusions derived in counterfactuals.

7 Third Approach: Unrestricted Model

The previous two approaches restrict the model's primitives and the set of allowable strategies. In this section, we do not impose such restrictions. Instead we assume that firms wrongly suppose that moments form a Markov process that summarizes all payoff relevant information related to fringe firms. The equilibrium concept under consideration is MME. One advantage of this approach is that there is a natural extension of our dynamic industry model to allow fringe firms to become dominant and vice versa. In this way, we will be able to fully endogeneize the dominant/fringe market structure in MME. We discuss this in the sequel.

There are, however, two main challenges that arise in this approach. First, we need to construct natural and meaningful candidates for the perceived transition kernel, $\hat{P}$, i.e., we need to choose a Φ operator. Candidate perceived transition kernels should ideally approximate well the *actual* transitions of moments under equilibrium strategies. Second, MME strategies will generally not be optimal, because moments may not summarize all payoff relevant information. To assess the extent of sub-optimality of MME strategies, one could ideally compute the value of the full information deviation. However this is not possible; computing a Markov best response suffers from the curse of dimensionality. To address this issue, we introduce a

computationally tractable error bound that provides an upper bound to the value of the full information deviation. This error bound is useful because it allows to assess how many and which moments need to be included in the strategies to achieve close to optimal strategies.

The reminder of this section is organized as follows: Subsection 7.1 describes a candidate for the perceived transition kernel. Subsection 7.2 shows how to extend the model to endogenize the set of dominant firms. Subsection 7.3 describes an algorithm to solve for MME; and Subsection 7.4 presents some numerical experiments motivated by the massive concentration of the beer industry in the United States over the second half for the previous century. The error bound is discussed in Section 8.

7.1 Candidate Perceived Transition Kernels

A natural choice for the perceived transition kernel is the empirical transitions of industry states. Recall that the underlying state of the industry s_t forms a Markov process for every moment-based strategy. We assume that for each such pair of strategies (μ, λ) there is a recurrent class $\mathcal{R}_{\mu, \lambda} \subset \mathcal{S}$ of underlying industry states that are visited with positive probability and an associated stationary distribution $q(s)$. An appealing choice of the perceived transition kernel $\hat{\mathbf{P}}_{\mu, \lambda}$ is the empirical transitions, as was presented in Example 4.2. We call this kernel the empirical transitions since it would coincide with the long run transitions from states in $\hat{\mathcal{S}}$. A similar concept was used by Fershtman and Pakes (2010) in a setting with asymmetric information.

Note that for a state $\hat{s}$ outside the recurrent class, the transitions detailed in Example 4.2 are not well defined, since the set $\mathcal{A}(\hat{s}) = \emptyset$. This renders consistency conditions outside $\mathcal{R}$ vacuous. In spite of that, the specification of the perceived transition kernel outside $\mathcal{R}$ may effect the equilibrium play. For example, firms may believe that in a particular state, the moment will shoot up resulting in fierce competition. As a result, firms may invest excessively to avoid reaching that state, and in equilibrium it is indeed never reached. One convenient approach to mitigating the effect of beliefs outside the equilibrium recurrent class is to assume that firms' transitions exhibit a small degree of noise, so that all industry states are in the recurrent class.⁹

To conclude this subsection we briefly describe another construction of the perceived transition kernel that has been successfully used in growth models in macroeconomics (Krusell and Smith, 1998) and subsequent literature. This perceived kernel assumes a parameterized and deterministic evolution for moments. That is, starting from industry state $\hat{s}_t = (\theta_t, d_t, z_t)$, the next moment value is assumed to be

$$\theta_{t+1} = G(\theta_t; \xi(d_t, z_t)),$$

where $\xi(d_t, z_t) \in \Xi$ are parameters. For example, this could represent a linear relationship with one moment, $\theta_{t+1} = \xi_0(d_t, z_t) + \xi_1(d_t, z_t)\theta_t$. In this case the goal would be to choose the functions ξ_0 and ξ_1 that approximate the actual transitions best, for instance by employing linear regressions. In comparison to empirical transitions, this perceived transition kernel has the disadvantage of being parametric and assuming

⁹If this assumption cannot be reconciled with the model of interest, one can consider the limit of models with diminishing noise in transitions.

deterministic transitions. These disadvantages, however, significantly reduce the computational burden of solving for MME.

7.2 Endogenous Market Structure

We now briefly describe a possible way to endogenize the set of dominant firms. Denote by $\mathcal{K}_f \subset \mathcal{X}_f$ and $\mathcal{K}_d \subset \mathcal{X}_d$ the sets of states from which a fringe firm may become dominant and a dominant firm becomes fringe, respectively. In every period, if an incumbent dominant firm enters $\mathcal{K}_d$, it becomes a fringe firm in state $x^{df} \in \mathcal{X}_f/\mathcal{K}_f$ in the next period with certainty. In every period where $|D_t| < \bar{D}$, i.e., the number of incumbent dominant firms is less than its maximum allowable value, one of the fringe firms who enters $\mathcal{K}_f$ in that period becomes a dominant firm in state $x^{fd} \in \mathcal{X}_d/\mathcal{K}_d$ in the next period. If more than one fringe firm enters $\mathcal{K}_f$ in that period, one of them is chosen at random and transitions to x^{fd} , the other firms transition with certainty to some state in $\mathcal{X}_f/\mathcal{K}_f$. Note that under this specification the transitions among the fringe and dominant tiers are naturally embedded in the transitions of firms. A specific example that fits this specification is that a fringe firm becomes dominant when growing above a pre-determined size. In this case, the transitions from fringe to dominant firms will naturally arise by the evolution of fringe firms given by their investment strategies.

While this extension does not require a modification of MME, the perceived transition kernel needs to explicitly account for the transitions between the fringe and dominant tier. Recall that $\bar{\mathbf{P}}_{\mu', \mu, \lambda}[x', d', z'|x, \hat{s}]$ (see (5)) is the kernel that describes the actual evolution of (x_{it}, d_t, z_t) , which firms can pin down exactly when transitions between the fringe and dominant tiers are not allowed. However, with such transitions moments may not contain sufficiently detailed information to pin down the transition probabilities between the fringe and dominant tiers. Therefore, we incorporate into $\bar{\mathbf{P}}$ firms' *perceived probabilities* that such transitions will occur given the moment-based industry state. In MME Φ will specify consistency conditions for these events in addition to moments transitions. We note that in the definition of $\hat{\mathbf{P}}$ in (5) we implicitly assume that the evolution of a single fringe firm does not affect the evolution of the moments. Even though, in this more general setting when a large fringe firm leaves the fringe tier this may affect the value of the moments, for simplicity we keep the independence assumption.

7.3 Computation

To compute MME we employ a real time stochastic algorithm similar to Pakes and McGuire (2001) and Fershtman and Pakes (2010). As is common in equilibrium solvers for dynamic games, the solver starts with a strategy profile, checks for equilibrium conditions, and updates the strategies until an equilibrium is found. Given strategies (μ, λ) and their associated value functions, it is useful to define

$$W(x, \hat{s}|\mu, \lambda) = \mathbb{E}_{\hat{\mathbf{P}}_{\mu, \lambda}}[V(x, \hat{s}_{t+1}|\mu, \lambda)|\hat{s}_t = \hat{s}], \quad (15)$$

where the expectation is taken with respect to the perceived transition kernel. W is the expected continuation value starting from industry state $\hat{s}$ and landing in state x in the next period. Note that we only integrate over the possible transitions of $\hat{s}$. It is worth emphasizing that if $x \in \mathcal{X}_d$ in (15) then $\hat{s}_{t+1}$ depends on x , whereas if $x \in \mathcal{X}_f$ the next industry state, $\hat{s}_{t+1}$, is independent of x . Namely, dominant firm i who transitions to x will integrate over $(\theta_{t+1}, d_{-i,t+1}, z_{t+1})$ with $d_{t+1} = (x, d_{-1,t+1})$. In light of that, we write the Bellman equation associated with C1 as follows (ignoring the subindices (μ, λ) to simplify notation):

$$V(x, \hat{s}; W) = \sup_{\substack{\iota \in \mathcal{I} \\ \rho \geq 0}} \left\{ \pi(x, \hat{s}) + \mathbb{E} \left[\phi \mathbf{1}\{\phi \geq \rho\} + \mathbf{1}\{\phi < \rho\} \left[-c(x, \iota) + \beta \mathbb{E}[W(x_{i,t+1}, \hat{s}) | x_{it} = x, \iota] \right] \right] \right\}$$

where the first expectation is taken with respect to the sell-off random value ϕ , and the second with respect to the firm's transition under investment level ι . Note that when evaluated at the optimal value function, the function W is sufficient to compute a best response strategy.

Suppose we want to compute MME with the empirical transition kernel described in the previous section. An important challenge is that for given strategies (μ, λ) we cannot explicitly write the kernel $\hat{\mathbf{P}}_{\mu, \lambda}$. We propose using simulation to address this issue; we simulate long enough sample paths and approximate $\hat{\mathbf{P}}_{\mu, \lambda}$ from the empirical transitions. One could then envision the following algorithm to compute MME. Start with strategies and compute via simulation the perceived transition kernel. Given the kernel, compute a best response. Then, iterate until the computed best response coincides with the initial strategies used to compute the perceived transition kernel. See Algorithm 1.¹⁰

The following remarks are important¹¹:

1. The algorithm terminates when the norm of the distance between a strategy and the best response to it is small. We consider the following norm, $\|\mu - \mu'\|_h = \max_{x \in \mathcal{X}} \left\{ \sum_{\hat{s} \in \hat{\mathcal{S}}} |\mu(x, \hat{s}) - \mu'(x, \hat{s})| h(\hat{s}) \right\}$ where h is a probability vector. We take h to be the frequency in which each industry state is visited. This is useful since simulation errors to estimate the perceived transition kernel will be higher for states that are visited infrequently.
2. In a finite length simulation it is possible that some states will not be visited, and for those states the perceived transition kernel cannot be computed. In those state we set the transitions in $\hat{\mathbf{P}}$ to be some predetermined ones, for example transition with certainty to the closest state (under some suitable norm) that was visited.

We found that a real-time dynamic programming algorithm that performs the simulation and optimization simultaneously is much faster than one proposed above. In this algorithm we update the function W

¹⁰For simplicity, the algorithm does not specify the formation of the perceived transition kernel when transitions across tiers are allowed. This is done by keeping track in the course of the algorithm, at each state $\hat{s}$, of the probability a fringe firm becomes dominant and the probability that a firm that enters $\mathcal{K}_f$ becomes dominant. In equilibrium these probabilities should coincide with the empirical transition probabilities observed under the equilibrium strategies.

¹¹In the algorithm, $0 < \sigma < 1$ is chosen to speed-up convergence.

Algorithm 1 Equilibrium solver

- 1: Initial with some (μ, λ) and industry state $s_0 = (f_0, d_0, z_0)$ with corresponding $\hat{s}_0$;
- 2: $n = 1, \delta'_{str} = \delta_{str} + 1$;
- 3: **while** $\delta'_{str} > \delta_{str}$ **do**
- 4: Simulate a T period sample path $\{(f_t, d_t, z_t)\}_{t=1}^T$ with corresponding $\{\hat{s}_t\}_{t=1}^T$;
- 5: Calculate the empirical frequencies of industry states $h(\hat{s}) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}\{\hat{s}_t = \hat{s}\}$ for all $\hat{s} \in \hat{\mathcal{S}}$;
- 6: Calculate for every $\hat{s} \in \hat{\mathcal{S}}$ with $h(\hat{s}) > 0$

$$\hat{\mathbf{P}}_{\mu, \lambda}[\theta' | \hat{s}] = \frac{\sum_{t=1}^T \mathbf{1}\{\hat{s}_t = \hat{s}, \theta(f_{t+1}) = \theta'\}}{\sum_{t=1}^T \mathbf{1}\{\hat{s}_t = \hat{s}\}};$$

- 7: Solve $\mu' = \operatorname{argmax}_{\mu' \in \mathcal{M}} V(x, \hat{s} | \mu', \mu, \lambda)$ for all $(x, \hat{s}) \in \mathcal{X} \times \hat{\mathcal{S}}$;
 - 8: Let $\lambda'(\hat{s}) = \mathbb{E}_{\mu, \lambda}[V(x^e, \hat{s}_{t+1} | \mu', \mu, \lambda) | \hat{s}_t = \hat{s}]$ for all $\hat{s} \in \hat{\mathcal{S}}$;
 - 9: $\delta'_{str} = \max(\|\mu - \mu'\|_h, \|\lambda - \lambda'\|_h)$;
 - 10: $\mu = \mu + (\mu' - \mu)/(1 + n^\sigma)$;
 - 11: $\lambda = \lambda + (\lambda' - \lambda)/(1 + n^\sigma)$;
 - 12: $n = n + 1$;
 - 13: **end while**;
-

while simulating sample paths. We use one step of Algorithm 1 to check whether $\mathcal{C}1$ holds and convergence has been achieved. The details of the algorithm are provided in Algorithm 2 in the Appendix.

7.4 Numerical Experiments – The Beer Industry

Some questions that have puzzled economists for decades are: What are the determinants of market structure? Why some industries become dominated by a handful of firms while still holding many small firms? How does the resulting market structure affect market outcomes? We believe that the approach developed in this section can be useful to shed light on these questions. In particular, our model and algorithm can be used to develop counterfactuals in different empirical settings in which the market structure is endogeneized in a fully dynamic model.

To illustrate the applicability of our method, we perform numerical experiments that are motivated by the long concentration trend in the beer industry in the US during the years 1960-1990. In the course of those years, the number of active firms dropped from about 150 to 30, and three industry leaders emerged: Anheuser-Busch (Budweiser brand among others), Miller, and Coors. Two competing explanations for this trend are common in the literature (see Tremblay et al. (2005)): an increase in the minimum efficient scale (MES), and an increase in the importance of advertising that with the emergence of national television has benefited big firms. The role of advertising as an “endogenous sunk cost” in determining market structure is discussed in detail in Sutton (1991) (see Chapter 13 for a discussion on the beer industry). In this section we calibrate a model to examine the role advertising may have on market structure. The model is similar to the dynamic advertising model by Doraszelski and Markovich (2007).

The model follows Example 4.1, where x_{it} is the goodwill of firm i in period t with associated market share

$$\sigma(x_{it}, s_t) = \frac{(x_{it})^{\alpha_1} (Y - p_{it})^{\alpha_2}}{B(s_t, p_t)}.$$

Firms invest in advertising to increase their goodwill stock over time and compete in prices in the spot market. The model parameters are calibrated from industry statistics available from different sources. The number of dominant firms is determined endogenously with a maximum number of three. The transitions between goodwill states are similar to those in Pakes and McGuire (1994), and are specified in Appendix C. The numerical experiments examine the effect of different specifications of the contribution of goodwill on firms' profits. If indeed big firms have an advantage over small firms in the way they convert goodwill to sales, this should be captured by the profit functions¹².

We say that the profit function exhibits: decreasing returns to advertising (DRA) if $\alpha_1 < 1$, constant returns to advertising (CRA) if $\alpha_1 = 1$, and increasing returns to advertising (IRA) if $\alpha_1 > 1$. We consider three specifications of returns to advertising with α_1^f and α_1^d controlling the returns to advertising for fringe and dominant firms, respectively. These take values in $(\alpha_1^f, \alpha_1^d) \in \{\alpha_D, \alpha_C\} \times \{\alpha_D, \alpha_C, \alpha_I\}$, where $\alpha_D = .85$, $\alpha_C = 1$, and $\alpha_I = 1.1$. The three cases under consideration are: (1) DRA-DRA with (α_D, α_D) , (2) DRA-CRA with (α_D, α_C) , and (3) CRA-IRA with (α_C, α_I) . Theoretical analysis in Adlakha et al. (2011) draws a sharp line between diffused and concentrated market structure: the market is concentrated if and only if $\alpha_1^d \geq 1$. This suggest that nondecreasing returns to advertising for dominant firms will be more appropriate in our case.

We calibrate the model parameters from a variety of empirical research that studies the beer industry or related advertising settings¹³. For example, the goodwill level captures a measure of the discounted expenditure on advertising, α_2 and Y are chosen to match the price elasticity in the average price, and the average sell-off value is taken from costs of used manufacturing plants. See Appendix C for a list of the parameters and their sources.

Figure 2 plots (on a log-scale) the average goodwill distribution of firms for the three cases. The un-normalized distribution of fringe firms is the solid line and the triangles are dominant firms (the average state of dominant firm is conditional on being active). Table 2 repots some average industry statistics.

The experiments suggest that higher returns on advertising indeed give rise to more right-skewed size distributions, as it is expected. Indeed, it is clear that the dominant/fringe structure is not present with DRA, as on average there is a vacancy in the dominant tier, and dominant firms are not much bigger than the biggest fringe firms. In both DRA-CRA and CRA-IRA the industry is much more concentrated and dominant firms are much larger than fringe firms. Second, we can see that the more concentrated the industry becomes, the fewer fringe firms are active on average. Yet, there is a reverse relationship between the number of active fringe firms and the average number of entrants, since higher concentration entails a shorter life span for fringe firms.

¹²Alternatively in the cost of accumulating goodwill, but we focus on the first.

¹³We thank Victor Tremblay and Carol Tremblay for providing supplementary data.

Figure 2: Size distribution of Firms (log scale). The solid line represents fringe firms and the triangles represent dominant firms.

We emphasize that each experiment includes 200 firms and 29 different individual states, which makes it much larger than any problem that can be solved if MPE was used as an equilibrium concept. We hope that the previous numerical experiments highlight the usefulness of our approach. In particular, we hope that it will be specially useful to perform counterfactuals when the market structure is endogenously determined as above.

Finally, we examine the effect fringe firms have on dominant firms by computing MPE with dominant firms only. To normalize the profit function and make more fair comparison we fix the moment on its average value. We obtain that in the DRA-CRA the average size of dominant firms drops from 64.9 to 54.7; furthermore the average size of the smallest dominant firm falls from 36.2 to 18.7, almost 50%. In CRA-IRA a similar, but a subtler outcome is observed; the average size of dominant firms lowers from 83.4 to 77.9 and the size of the smallest dominant firm drops by almost 20% from 51 to 42. This suggests that competition from the fringe tier does have an influence on dominant firms. In particular, the strategic interactions between fringe firms and the smaller dominant firms seem to be fairly intense, since they are close competitors. If the fringe firms were omitted from the analysis, as is some times done for computationally tractability, important investment incentives for dominant firms that are driven by fringe firms may wrongly ignored.

8 Bounding the Value of the Full Informational Deviation

In the third approach moments are not sufficient statistics for the future evolution of fringe firms. We evaluate the performance of MME strategies by considering the value of the full information deviation, i.e., a deviation to a strategy that keeps track of the full underlying state of the industry s . In theory, one could compute the value of unilateral deviation exactly, however this is almost as computationally challenging as

Industry average	CRA-IRA	DRA-CRA	DRA-DRA
Market share dominant	60%	42%	10%
First moment	217	244	348
Active fringe firms (#)	147	167	180
Active dominant firms (#)	3	3	2.3
Size fringe (goodwill)	1.1	1.4	2
Size dominant (goodwill)	83.4	64.9	24.1
Entrants per period (#)	16.5	12.3	7.6
Time in industry fringe	12.9	16	26.2
Time as dominant	1536	360	21
Consumer surplus per period	141.4	133.2	130.5

Table 2: Average industry statistics

solving for the equilibrium in the underlying state space $\mathcal{S}$. As such, we suggest a computationally tractable error bound that upper bounds the value of the full information deviation. This error bound is based on ideas from robust dynamic programming (RDP).

For every $\hat{s} = (\theta, d, z) \in \hat{\mathcal{S}}$ define the *consistency set* $\mathcal{S}_f(\hat{s}) = \{f \in \mathcal{S}_f \mid \theta(f) = \theta\}$, i.e., this is the set of all histograms that are consistent with the value of the moments in state $\hat{s}$. Moments are not sufficient statistics for the evolution of the industry, because typically $\mathcal{S}_f(\hat{s})$ is not a singleton and different fringe firm states in the consistency set may have different future evolutions. The error bound let the deviating firm choose for every industry state a histogram from the associated consistency set, and then to choose the optimal investment/exit strategy. Theorem ?? shows that this procedure produces an upper bound on the value of the full information deviation. Generally, computing this error bound will be intractable. However, we can take advantage of the fact that the number of fringe firms is large to considerably simplify this computation.

Let (μ, λ) be some fixed MME strategies for the reminder of this section. Recall that $V^*(x, s \mid \mu, \lambda)$ is the actual value of playing MME strategies starting from (x, s) . Denote by $\bar{V}^*(x, s \mid \mu, \lambda) = \sup_{\mu' \in \mathcal{M}^*} V(x, s \mid \mu', \mu, \lambda)$ the value of unilateral deviation to a strategy that may depend on the underlying industry state. The value of the full informational deviation is $\Delta_{\mu, \lambda}(x, s) = \bar{V}^*(x, s \mid \mu, \lambda) - V^*(x, s \mid \mu, \lambda)$. Note that given MME strategies (μ, λ) , $V^*(x, s \mid \mu, \lambda)$ can be easily computed using forward simulation. However, the problem of finding the optimal strategy that achieves $\bar{V}^*$ is subject to the curse of dimensionality. Instead we find an upper bound to $\bar{V}^*$. With this we can upper bound $\Delta_{\mu, \lambda}$. To do so we construct a “robust Bellman operator” as follows

$$\begin{aligned}
(T_R \hat{J})(x, \hat{s}) = & \sup_{\substack{\iota \in \mathcal{I} \\ \rho \geq 0}} \sup_{f \in \mathcal{S}_f(\hat{s})} \left\{ \pi(x, \hat{s}) + \mathbb{E} \left[\phi \mathbf{1}\{\phi \geq \rho\} \right. \right. \\
& \left. \left. + \mathbf{1}\{\phi < \rho\} \left[-c(x, \iota) + \beta \mathbb{E}[\hat{J}(x_{i,t+1}, \hat{s}_{t+1}) \mid x_{it} = x, s = (f, d, z), \iota] \right] \right\}, \quad (16)
\end{aligned}$$

where $\hat{J} \in \hat{\mathcal{J}}$ is a bounded vector $\hat{J} : \mathcal{X} \times \hat{\mathcal{S}} \rightarrow \mathbb{R}$. Recall that we assume all competitors use MME strategies (μ, λ) . That is, the robust Bellman operator is defined on $\mathcal{X} \times \hat{\mathcal{S}}$ and it is identical to the standard Bellman operator associated to the best response in $\mathcal{C}1$, except that the firm can also choose any underlying fringe firm state consistent with the moment θ . The next lemma follows from the literature on RDP.

Lemma 8.1. *The operator T_R satisfies the following properties:*

1. T_R is a contraction mapping modulo β . That is, for $\hat{J}, \hat{J}' \in \hat{\mathcal{J}}$, $\|T_R \hat{J} - T_R \hat{J}'\|_\infty \leq \beta \|\hat{J} - \hat{J}'\|_\infty$.
2. The equation $T_R \hat{J} = \hat{J}$ has a unique solution $\hat{J}^*$.
3. $\hat{J}^* = \lim_{k \rightarrow \infty} T_R^k \hat{J}$ for all $\hat{J} \in \hat{\mathcal{J}}$.

We call $\hat{J}^*$ the *robust value function*. The next result relates it to the optimal value function $\bar{V}^*$.

Theorem 8.1. *For all $(x, s) \in \mathcal{X} \times \mathcal{S}$*

$$\bar{V}^*(x, s) \leq \hat{V}(x, \hat{s}),$$

where $\hat{s}$ is the moment-based industry state that is consistent with s .

Both proofs are given in Appendix D. In essence, the robust value function resolves the indeterminacy of moments transitions by choosing the most favorable moment transition for the deviating firm. Intuitively, this should provide an upper bound for $\bar{V}^*$. Therefore, we can bound the value of full informational deviation with $\hat{\Delta}(x, s) = \hat{V}(x, \hat{s}) - V^*(x, s)$, where $\hat{s}$ is the moments-based industry state of s . The advantage of computing $\hat{V}$ over $\bar{V}^*$ is that T_R operates on the state space $\hat{\mathcal{S}}$ which is much smaller than $\mathcal{S}$. Nevertheless, the computation of $\hat{V}$ is still demanding as will be shown in Subsection 8.1. We comment here that we can augment the state of moments in the robust operator to include moments that are not considered in MME. This will lower the robust bound since it decreases the size of the consistency sets in (16).

8.1 Computation

Finding the fixed point of the operator T_R is generally NP-complete (Iyengar, 2005, §3). This is not surprising, since the optimization over consistency sets may be very complex. However, as we now show, iterating operator T_R becomes much simpler if there is a large numbers of fringe firms, and, as a consequence, the transitions of the fringe firms are close to being deterministic. This considerably simplifies the computation of the second expectation in (16), as for any f_t the evolution of θ_{t+1} becomes deterministic. Moreover, it is simple to show that in this case the evolution of the next moment is linear in the fringe firm state. Thus, finding the optimal consistent f in (16) is equivalent to choosing the next moment from a set of moments that are accessible from the current industry state. This is a manageable optimization problem because its feasible set is described by linear equations and the decision variables are moments which are described by a low dimensional vector.

In the reminder of this section we assume that the set of individual fringe firm states is discrete and univariate, $\mathcal{X}_f = \{x^1, \dots, x^{\bar{n}}\} \times \{f\}$, where $x^n \in \mathbb{R}$, for all $n = 1, \dots, \bar{n} < \infty$ ¹⁴. In addition it is assumed that the entry process is such that exactly $\lambda(\hat{s})$ firms enter at industry state $\hat{s}$, and that there are no tier transitions. We comment about tier transitions at the end of the section.

Finding the set of accessible moments amounts to solving an integer feasibility problem. To simplify, let us assume that θ consist of only one moment of the form $\theta = \sum_{x \in \tilde{\mathcal{X}}_f} f(x) x^b$ for $b \geq 0$, where $\tilde{\mathcal{X}}_f \subset \{x^1, \dots, x^{\bar{n}}\}$. This representation is general enough to include moments, quantiles and others statistics. The extension to more moments is direct.

Assuming deterministic fringe transitions due to a law of large numbers, moment θ' is accessible from moment θ in industry state $\hat{s}$ if there exists a fringe firm state $f \in \mathcal{S}_f$ that solves the following system of linear equations,

$$\sum_{x \in \tilde{\mathcal{X}}_f} f(x) x^b = \theta \quad (17)$$

$$\sum_{x \in \tilde{\mathcal{X}}_f} f(x) \mathbb{E}_\mu[(x_{i,t+1})^b | (x_{it}, \hat{s}_t) = (x, \hat{s})] + \mathbf{1}\{x^e \in \tilde{\mathcal{X}}_f\} \lambda(\hat{s}) (x^e)^b = \theta' \quad (18)$$

$$f \in \mathbb{N}^{\bar{n}}, \quad (19)$$

where the first equation states that the moment is consistent the fringe firm state, and the second that the expected next moment is θ' . We say that moment θ' is accessible from $\hat{s}$ if this system of linear equations has a solution. This motivates the definition of the *accessibility set* $A(\hat{s})$, where $\theta' \in A(\hat{s})$ if and only if it is accessible from $\hat{s}$, i.e., if there is a fringe firm state consistent with $\hat{s}$ such that the expected next moment is θ' . Due to the integrability constraint $f(x) \in \mathbb{N}^{\bar{n}}$ the computation of $A(\hat{s})$ is demanding, however we can relax this constraint by replacing it with $f \geq 0$. With that, the accessibility problem amounts to solving a feasibility problem of a system of linear equations that can be solved easily. We denote the *relaxed accessibility set* by $\hat{A}(\hat{s})$; this set contains $A(\hat{s})$.

Define the operator

$$(\hat{T}\hat{J})(x, \hat{s}) = \sup_{\substack{\iota \in \mathcal{I} \\ \rho \geq 0}} \sup_{\theta' \in A(\hat{s})} \left\{ \pi(x, \hat{s}) + \mathbb{E} \left[\phi \mathbf{1}\{\phi \geq \rho\} \right. \right. \quad (20)$$

$$\left. \left. + \mathbf{1}\{\phi < \rho\} \left[-c(x, \iota) + \beta \mathbb{E}[\hat{J}(x_{i,t+1}, (\theta', d_{t+1}, z_{t+1})) | x_{it} = x, \hat{s}_t = \hat{s}, \iota] \right] \right] \right\}, \quad (21)$$

where $\hat{J} \in \hat{\mathcal{J}}$. The next lemma shows that we can search over accessibility sets instead of the much larger consistency sets.

Lemma 8.2. *Assume that there are no transitions between tiers and that fringe firms follow deterministic*

¹⁴The extension to multivariate x^n is simple, and the extension to a continuous state space will require discretization.

transitions. Then

$$(T_R \hat{J}) = (\hat{T} \hat{J}). \quad (22)$$

for every $\hat{J} \in \hat{\mathcal{J}}$.

Based on this, we propose the following computationally tractable algorithm to find the robust error bound: (1) construct the relaxed accessibility sets by solving the relaxed feasibility problem for all $\hat{s} \in \hat{\mathcal{S}}$ and $\theta' \in \mathcal{S}_\theta$ and store them; and (2) iterate the operator $\hat{T}$ over the relaxed accessibility sets until a fixed point is found. Since the relaxed accessibility sets contain the accessibility sets, this provides an upper bound to $\bar{V}^*$, assuming deterministic fringe transitions. For any problem for which MME is solvable this procedure is computationally manageable.

Since in practice the number of fringe firms is finite, the transition of moments is not deterministic and the robust bound is only an approximation to an upper bound.¹⁵ With $|F|$ fringe firms, by a central limit theorem, the error in the moment transition would be of order $\sqrt{|F|}$ from the mean moment transition, which corresponds to deterministic transitions. It is possible to formally derive a probabilistic version of the robust bound using standard probability bounds. However, we believe that in many settings of interest, where the number of fringe firms is relatively large, the robust bound derived above will provide a valid upper bound.

Lastly, we comment about the necessary modification to the robust bound when firms can transition between tiers. In industry states where there are no vacancies in the dominant tier ($|D_t| = \bar{D}$), the bound does not change. In the remainder of industry states, the deviating firm can pick a histogram that does not allow any fringe firm to transition to a state from which it can become dominant. As such, the robust bound may not perform well, unless the space of moments is augmented in some way that limits the ability of the deviating firm block rivals from becoming dominant¹⁶.

8.2 Numerical Experiments

We have done extensive numerical experiments using the robust bound:

1. In one set of experiments we computed the robust bound for several instances of models in which fringe firms are constrained in their strategies like in Section 6. Here, when the number of fringe firms is large, the value of the full information deviation for dominant firms should be small as suggested by the analysis in that section. The robust bound, indeed, confirmed this, taking very small values.
2. We also studied a sequence of models in which we allow fringe firms strategies to become increasingly far apart from the restricted ones imposed in Section 6. As we depart from the latter, the moments

¹⁵Note that if the $\hat{V}$ is concave with respect to the moment, then $E_{\theta'}[\hat{V}(x, \theta')|\hat{s}] \leq \hat{V}(x', E_{\theta'}[\theta'|\hat{s}])$ so by considering deterministic transitions we will further increase the robust bound.

¹⁶For example, we could augment the space of moments to include the state of the biggest fringe firms if $\mathcal{K}_f = \{x^{\bar{n}}\}$.

cease to be sufficient statistics. The robust bound again exhibited the expected behavior; it increased monotonically as the fringe firm strategies departed from the ones imposed in Section 6.

3. We also computed the robust bound in another important model where moments are known to be sufficient statistics, namely the stochastic growth model by Krusell and Smith (1998). Again, the error bound was small.

The previous results suggest that the robust bound can be useful to test the extent of sub-optimality of MME strategies and determine whether they provide an accurate approximation.

On a less positive note, we also computed the robust bound for the numerical experiments on the beer industry, and so far the robust bound with few moments seems to be looser than we expected in this setting. Based on the analysis of the results there, we think that adding lagged moments would improve the robust bound significantly compared to the contemporaneous moments we have considered so far. We are modifying the algorithm to handle this case and we hope to report the results shortly.

9 Conclusion and Future Work

Our model opens up the door to study new issues in empirical analysis of dynamic oligopoly models. Our methods allow fringe firms to be introduced to dynamic oligopoly models in the spirit of EP in varying degrees of flexibility and computational burden. We hope our models will prove useful in empirical studies.

Finally, we envision that our Robust bound in the third approach can guide researchers solving stochastic growth models in the spirit of Krusell and Smith (1998) to find appropriate ways of aggregating the state space. We also hope that our methods can be helpful to study dynamic models with forward looking consumers. We leave the application of our approach in these settings for future research.

References

- Adlakha, S., R. Johari, G. Y. Weintraub. 2011. Equilibria of dynamic games with many players: Existence, approximation, and market structure. *Working paper*.
- Bajari, P., C. L. Benkard, J. Levin. 2007. Estimating dynamic models of imperfect competition. *Econometrica* **75**(5) 1331 – 1370.
- Benkard, C. L. 2004. A dynamic analysis of the market for wide-bodied commercial aircraft. *Review of Economic Studies* **71**(3) 581 – 611.
- Benkard, C. L., A. Bodoh-Creed, J. Lazarev. 2010. Simulating the dynamic effects of horizontal mergers: U.S. airlines. Working Paper, Yale.
- Benkard, C. L., P. Jeziorski, G. Y. Weintraub. 2011. Oblivious equilibrium for concentrated industries. Working Paper, Columbia University.
- Bertsekas, Dimitir P., Steven Shreve. 1978. *Stochastic Optimal Control: The Discrete-Time Case*. Academic Press Inc.
- Besanko, D., U. Doraszelski, Y. Kryukov, M. Satterthwaite. 2010. Learning-by-doing, organizational forgetting, and industry dynamics. *Econometrica* **78**(2).
- Besanko, D., M. K. Perry, R. H. Spady. 1990. The logit model of monopolistic competition: Brand diversity. *The Journal of Industrial Economics* **38**(4) 397 – 415.
- Caplin, A., B. Nalebuff. 1991. Aggregation and imperfect competition - on the existence of equilibrium. *Econometrica* **59**(1) 25 – 59.
- Clementi, G.L., D. Palazzo. 2010. Entry, exit, firm dynamics, and aggregate fluctuations. *Working paper*.
- Collard-Wexler, A. 2010a. Demand fluctuations in the ready-mix concrete industry. Working Paper, NYU.
- Collard-Wexler, A. 2010b. Mergers and sunk costs: An application to the ready-mix concrete industry. Working Paper, NYU.
- Collard-Wexler, A. 2011. Productivity dispersion and plant selection. Working Paper, NYU.
- Corbae, D., P. D’Erasmus. 2011. A quantitative model of banking industry dynamics. Working paper, UT Austin.
- Dixit, A. K., J. E. Stiglitz. 1977. Monopolistic competition and optimum product diversity. *American Economic Review* **67**(3) 297 – 308.

- Doraszelski, U., K. Judd. 2006. Avoiding the curse of dimensionality in dynamic stochastic games. Working Paper, Hoover Institution.
- Doraszelski, U., S. Markovich. 2007. Advertising dynamics and competitive advantage. *RAND Journal of Economics* **38**(3) 557–592.
- Doraszelski, U., A. Pakes. 2007. A framework for applied dynamic analysis in IO. *Handbook of Industrial Organization, Volume 3*. North-Holland, Amsterdam.
- Ericson, R., A. Pakes. 1995. Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies* **62**(1) 53 – 82.
- Evans, D. S. 1987. The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *The Journal of Industrial Economics* **35**(4) 567 – 581.
- Farias, V., D. Saure, G.Y. Weintraub. 2010. An approximate dynamic programming approach to solving dynamic oligopoly models. Working Paper.
- Fershtman, C., A. Pakes. 2010. Oligopolistic dynamics with asymmetric information: A framework for empirical work. Working Paper, Harvard University.
- Hartley, James E. 1997. *The Representative Agent in Macroeconomics*. Routledge.
- Hopenhayn, H. A. 1992. Entry, exit and firm dynamics in long run equilibrium. *Econometrica* **60**(5) 1127 – 1150.
- Iacovone, L., B. Javorcik, W. Keller, J. Tybout. 2009. Walmart in Mexico: The impact of FDI on innovation and industry productivity. Working paper, Penn State University.
- Iyengar, G. 2005. Robust dynamic programming. *Mathematics of Operations Research* **30**(2) 257–280.
- Jia, P., P. Pathak. 2011. The cost of free entry: Evidence from real estate brokers in greater Boston. Working Paper, MIT.
- Judd, K. 1998. *Numerical Methods in Economics*. MIT Press.
- Khan, A., J.K Thomas. 2008. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica* **76**(2) 395 – 436.
- Krusell, P., A. A. Smith, Jr. 1998. Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy* **106**(5) 867–896.
- Lee, R. S. 2010. Dynamic demand estimation in platform and two-sided markets. Working Paper, NYU.
- Maskin, E., J. Tirole. 1988. A theory of dynamic oligopoly, I and II. *Econometrica* **56**(3) 549 – 570.

- Maskin, E., J. Tirole. 2001. Markov perfect equilibrium i. observable actions. *Journal of Economic Theory* **100** 191 – 219.
- Pakes, A., P. McGuire. 1994. Computing Markov-perfect Nash equilibria: Numerical implications of a dynamic differentiated product model. *RAND Journal of Economics* **25**(4) 555 – 589.
- Pakes, A., P. McGuire. 2001. Stochastic algorithms, symmetric Markov perfect equilibrium, and the ‘curse’ of dimensionality. *Econometrica* **69**(5) 1261 – 1281.
- Qi, S. 2008. The impact of advertising regulation on industry: The cigarette advertising ban of 1971. Working paper, University of Minnesota.
- Roberts, Mark J., Larry Samuelson. 1988. An empirical analysis of dynamic, nonprice competition in an oligopolistic industry. *The RAND Journal of Economics* **19**(2) pp. 200–220. URL <http://www.jstor.org/stable/2555700>.
- Rojas, Christian. 2008. Price competition in u.s. brewing. *The Journal of Industrial Economics* **56**(1) 1–31. doi:10.1111/j.1467-6451.2008.00330.x. URL <http://dx.doi.org/10.1111/j.1467-6451.2008.00330.x>.
- Ryan, S. 2010. The costs of environmental regulation in a concentrated industry. Working paper, MIT.
- Santos, C. D. 2010. Sunk costs of R&D, trade and productivity: the moulds industry case. Working Paper, U. of Alicante.
- Sutton, J. 1991. *Sunk Costs and Market Structure*. 1st ed. MIT Press.
- Sutton, J. 1997. Gibrat’s legacy. *Journal of Economic Literature* **35**(1) 40 – 59.
- Sweeting, Andrew. 2007. Dynamic product repositioning in differentiated product markets: The case of format switching in the commercial radio industry. Working Paper, Duke.
- Thurk, J. 2009. Market effects of patent reform in the U.S. semiconductor industry. Working paper, University of Texas at Austin.
- Tomlin, B. 2008. Exchange rate volatility, plant turnover and productivity. Working Paper, Boston University.
- Tremblay, Victor J., Natsuko Iwasaki, Carol Horton Tremblay. 2005. The dynamics of industry concentration for u.s. micro and macro brewers. *Review of Industrial Organization* **26** 307–324. 10.1007/s11151-004-8114-9.
- Tremblay, Victor J., Carol Horton Tremblay. 2005. *The US Brewing Industry: Data and Economic Analysis*, MIT Press Books, vol. 1. The MIT Press. URL <http://ideas.repec.org/b/mtp/titles/0262201518.html>.

- Weintraub, Gabriel Y., C. Lanier Benkard, Benjamin Van Roy. 2008. Markov perfect industry dynamics with many firms. *Econometrica* **76**(6) 1375–1411.
- Weintraub, Gabriel Y., C. Lanier Benkard, Benjamin Van Roy. 2010. Computational methods for oblivious equilibrium. *Operations Research (Special Issue in Computational Economics)* **58**(4) 1247–1265.
- Xu, Y. 2008. A structural empirical model of R&D, firm heterogeneity, and industry evolution. Working paper, NYU University.

A Stackelberg-Cournot Profit Function with Moments

Consider k dominant firms with constant marginal costs mc_i for $i = 1, \dots, k$. In addition there are n fringe firms in the market with cost function of the form $c_j(q) = \frac{1}{c_j} q^\alpha$, where c_j is positive, $\alpha > 1$ and q is the quantity produced. We will use index i for dominant firms and index j for fringe firms to avoid confusion. Accordingly, the states of dominant firms are their marginal costs and those of fringe firms are the coefficients in their cost function. Firms compete in quantities. Fringe firms are price takers since they are small, However dominant firms internalize their effect on the price and take into account the response of fringe firms. With one dominant firm this is the well known Stackelberg game. With more than one dominant firm all dominant firms play a Cournot game on the residual demand given the fringe firms' response. The inverse demand is given by a differentiable function $P(Q)$. Fringe firms optimal production problem is concave and the first order conditions are sufficient,

$$q_j = \left(\frac{P c_j}{\alpha} \right)^{\frac{1}{\alpha-1}},$$

where P is the market clearing price. Consequently, the aggregate fringe production is,

$$Q_F = \sum_j \left(\frac{P c_j}{\alpha} \right)^{\frac{1}{\alpha-1}} = \left(\frac{P}{\alpha} \right)^{\frac{1}{\alpha-1}} \sum_j c_j^{\frac{1}{\alpha-1}} := \left(\frac{P}{\alpha} \right) \tilde{c}.$$

Taking into account the inverse demand function Q_F solves,

$$Q_F = \left(\frac{P(Q_F + Q_D)}{\alpha} \right)^{\frac{1}{\alpha-1}} \tilde{c}. \quad (23)$$

This exhibits a simple solution for linear demand function and some constant elasticity demand functions. Dominant firm i 's problem is to maximize the following quantity over non-negative quantities,

$$\pi_i = q_i (P(q_i + Q_F + Q_{D-i}) - mc_i).$$

We assume the problem is concave. First order conditions imply,

$$P(Q) - mc_i + q_i P'(Q) = 0.$$

Aggregating the first order conditions for all dominant (assuming interior solution for all dominant firms) we have,

$$Q_D = \frac{\sum_i mc_i - kP(Q)}{P'(Q)} := \frac{\widetilde{mc} - kP(Q)}{P'(Q)}. \quad (24)$$

In an equilibrium with full participation equations 23 and 24 must hold.

It is clear that the profit of both dominant and fringe firms will depend only on the moments of the fringe

firm's distribution. Some demand function coupled with certain parameter value exhibit an analytic form, like in linear demand functions some constant elasticity demand functions. Nevertheless, the equilibrium can be evaluated numerically fairly easily in other cases. The further analyze the linear demand case.

A.0.1 Linear demand $P(Q) = a - bQ$

We have the following, $Q_D = \frac{\tilde{m}c - ka + kb(Q_D + Q_F)}{-b}$ which simplifies to,

$$Q_D = \frac{ka - \tilde{m}c}{b(1+k)} - \frac{k}{k+1}Q_F := \tilde{d} + \frac{k}{k+1}Q_F$$

and,

$$Q = \tilde{d} + \frac{1}{k}Q_F.$$

From 23 we have,

$$Q_F^{\alpha-1} + \frac{b\tilde{c}^{\alpha-1}}{\alpha(1+k)} + \frac{(b\tilde{d} - 1)\tilde{c}^{\alpha-1}}{\alpha} = 0.$$

It is clear that the profits of all firms depend only on the $1/(\alpha - 1)$ moment of fringe firms and on the some of marginal costs of dominant firms. Moreover, if α equals 2 or 3 an analytic solution to the equilibrium can be found.

We further analyze the case $\alpha = 2$. We have the following quantities,

$$Q_F = \frac{(a - b\tilde{d})(1+k)\tilde{c}}{2(1+k) + b\tilde{c}},$$

$$Q_D = \tilde{d} - \frac{k(a - b\tilde{d})\tilde{c}}{\alpha(1+k) + b\tilde{c}},$$

$$Q = \frac{2\tilde{d}(1+k) + a\tilde{c}}{2(1+k) + b\tilde{c}}.$$

B Second Approach

Fringe firms transitions. This appendix will propose a distribution for ζ'_f and ζ_d . The proposal is based on the dynamics in Weintraub et al. (2010). For dominant firms,

$$\zeta_d(x, \iota) = \begin{cases} x + 1 & \text{w.p. } \frac{\delta a \iota}{1 + a \iota} \\ x & \text{w.p. } \frac{(1 - \delta') + (1 - \delta)a \iota}{1 + a \iota} \\ x - 1 & \text{w.p. } \frac{\delta'}{1 + a \iota}, \end{cases}$$

Table 3: Model parameters for the experiments of Subsection 6.1

Notation	Value	Description
β	.95	Discounting factor
m	30	Market size
p	.05	Power in fringe transition
$\mathcal{X}_d$	$\{6, 7, \dots, 11\}$	Dominant firms' state space
η	1.1	Dominant investment cost rate
δ_d, δ'_d, a_d	.4, .6, 1.5	Dominant transition parameters
$\bar{x}$	6	Upper bound on fringe firms
$(\zeta_1, \dots, \zeta_5)$	(.93, .96, 1, 1.09, 1.12)	Fringe transition parameters
$(\delta'_1, \delta'_2, \delta_4, \delta_5)$	(.2, .6, .6, .2)	Fringe transition parameters
a_f	3	Fringe transition parameters
x^e	1	Entry state
κ	29	Entry cost
$\bar{\phi}$	9.1	Expected sell-off value (exponential)

where $\delta, \delta' \in (0, 1)$ and $a > 0$ are constants. If x is on the upper boundary of X_d we add the probability of going to the probability of staying at the same state place and similarly if x is on the lower boundary of X_d we add the probability of decreasing to the probability of staying at the same state.

The transition of fringe firms is a generalization of the above. Let ζ'_f take values in $\{\zeta_1, \zeta_2, \dots, \zeta_L\}$ with $\zeta_l < \zeta_{l+1}$ and let $L > \bar{l} > 1$ be some interior index. Note that $\zeta_l - 1$ is approximately the growth rate if x_{it} is close to 1. We define the random transition by,

$$\zeta'_f(l) = \begin{cases} \zeta_l & \text{w.p. } \frac{\delta'_l}{1+al} \text{ for } l = 1, \dots, \bar{l} - 1 \\ \zeta_{\bar{l}} & \text{w.p. } \frac{(1-\delta')+(1-\delta)al}{1+al} \\ \zeta_l & \text{w.p. } \frac{\delta_l al}{1+al} \text{ for } l = \bar{l} + 1, \dots, L, \end{cases}$$

with $\sum_{l=1}^{\bar{l}} \delta'_l = \delta' < 1$, $\sum_{l=\bar{l}+1}^L \delta_l = \delta < 1$, δ'_l is positive for all $l = 1, \dots, \bar{l} - 1$, δ_l is positive for all $l = \bar{l} + 1, \dots, L$ and a is a positive parameter. It is easy to see that ζ'_f is stochastically increasing in investment and that the distribution is well defined for non-negative investment.

C Beer Industry Experiments

Denote by x_{it} the goodwill of firm i at time t , with the interpretation that goodwill x is approximately the long run average expenditure on advertising (in millions of dollars). The evolution of goodwill is similar to

Pakes and McGuire (1994)

$$x_{it+1} = \begin{cases} x_{it}(1 + \rho) & \text{w.p. } \frac{\gamma\psi(x)l_{it}}{1+\psi(x)l_{it}} \\ x_{it} & \text{w.p. } \frac{1-\gamma'+(1-\gamma)\psi(x)l_{it}}{1+\psi(x)l_{it}} \\ x_{it}/(1 + \rho) & \text{w.p. } \frac{\gamma'}{1+\psi(x)l_{it}}. \end{cases}$$

This equivalent to a depreciation factor $1/(1 + \rho)$ as is common in the literature on goodwill. With this in mind we define grid of states $\{x^{out}, 0, x^1, x^2, \dots, x^n\}$ for the possible values of goodwill firms can take, where $x^k = x^0(1 + \delta)^{k-1}$ for some $x^0 > 0$, and x^{out} is an inactive state. To maintain the relationship between goodwill and advertising costs, we choose the parameter $\psi(x)$ such that $E[x_{it+1}|x_{it} = x, l_{it} = x] = x$, i.e., a firm with goodwill x has to invest x dollars in advertising to maintain goodwill level x on average. It follows that $\psi(x) = \frac{\gamma'}{\gamma} \frac{1-(1+\delta)^{-1}}{\delta x}$.

The next table lists some of the model parameters with their relevant source.

Description	Value	Source
Number of firms	200	This figure is chosen to be greater than the maximal number of active firms in this period
Maximal number of dominant firms ($\bar{D}$)	3	This is the actual number of dominant firms in the industry
Depreciation of goodwill (ρ)	.25	Roberts and Samuelson (1988) estimate this by .2 for the cigarette market, we increase this figure to reduce the size of the state space
Production cost per barrel	120	Rojas (2008) estimated the markup to be about a third of the price, and the average price is \$165 per barrel
Fixed cost per period fringe (double for dominant)	10^6	This is necessary for negative profits in some states
Average entry cost (exponential dist.)	35×10^6	Based on costs of new plants
Sell-off value (exponential dist.)	7×10^6	Based on costs of used plans
Profit function parameters (Y and α)	200 and 1, resp.	Chosen to match price elasticity (-.5) for the average price, see (Tremblay and Tremblay, 2005, p. 23)

In addition we take $\beta = .925$ and $(\delta', \delta) = (1, .55)$ are the transition parameters. Finally, after some experimentation we choose the market size $m = 30$. Our entry model is different from that of Section 3. Here there is a maximal number of firms (200) out of them some are active and some are inactive (state x^{out}). Every firm in the inactive state will draw an entry cost in every period and may decide to enter, and every exiting fringe firm will become inactive. Inactive firms are shortly lived, i.e., they do not strategize over their entry epoch.

D Third Approach

Algorithm 2 Equilibrium solver with real-time dynamic programming

```

1: Initiate  $W(x, \hat{s}) = W_1(x, \hat{s})$  for all  $(x, \hat{s}) \in \mathcal{X} \times \hat{\mathcal{S}}$ ;
2:  $c(\hat{s}) = 0$  for all  $\hat{s} \in \hat{\mathcal{S}}$ ;
3: Initiate industry state  $(f_1, d_1, z_1)$  and  $\hat{s}_1 = (\theta(f_1), d_1, z_1)$ ;
4:  $t = 1, \delta'_w = \delta_w + 1$ ;
5: while  $\delta'_w > \delta_w$  do
6:    $W'(x, \hat{s}) = W(x, \hat{s})$  for all  $(x, \hat{s}) \in \mathcal{X} \times \hat{\mathcal{S}}$ ;
7:    $t = 1$ ;
8:   while  $t \leq K$  do
9:     for all  $x$  with  $f_t(x) > 0$  or  $x \in d_t$ , do
10:      Compute optimal strategies in  $V(x, \hat{s}_t; W)$  and store them;
11:    end for;
12:    Compute optimal entry cutoff from  $V(x^e, \hat{s}_t; W)$  and store it;
13:    Simulate  $(f_{t+1}, d_{t+1}, z_{t+1})$  and  $\hat{s}_{t+1}$  from these strategies;
14:    Let  $\gamma = \frac{1}{n^\sigma + c(\hat{s}_t)}$ ;
15:    for all  $x' \in \mathcal{X}_f$  do
16:      Compute  $V(x', \hat{s}_{t+1}; W)$ ;
17:      Update  $W(x', \hat{s}_t) = \gamma V(x', \hat{s}_{t+1}) + (1 - \gamma)W(x', \hat{s}_t)$ ;
18:    end for;
19:    for all Dominant firm  $i$  and  $x' \in \mathcal{X}_d$  that is accessible in one step from  $x_{it}$  do
20:      Define  $\hat{s}'_{t+1}$  to be the industry state  $\hat{s}_{t+1}$  when firm  $i$  transitions to state  $x'$ ;
21:      Compute  $V(x', \hat{s}'_{t+1}; W)$ ;
22:      Update  $W(x', \hat{s}_t) = \gamma V(x', \hat{s}'_{t+1}; W) + (1 - \gamma)W(x', \hat{s}_t)$ ;
23:    end for;
24:     $c(\hat{s}_t) = c(\hat{s}_t) + 1, t = t + 1$ ;
25:  end while;
26:   $\delta'_w = \|W' - W\|_\infty$ ;
27:   $c(\hat{s}) = 0$  for all  $\hat{s} \in \hat{\mathcal{S}}$ ;
28:   $(f_1, d_1, z_1) = (f_{K+1}, d_{K+1}, z_{K+1})$ , and  $\hat{s}_1 = (\theta(f_1), d_1, z_1)$ ;
29: end while;
30: Compute  $\mu(x, \hat{s})$  and  $\lambda(\hat{s})$  from  $V(x, \hat{s}; W)$  for all  $(x, \hat{s}) \in \mathcal{X} \times \hat{\mathcal{S}}$ ;
31: Run Algorithm 1 with these strategies;

```

Proof of Lemma 8.1. Denote by $\mathcal{Q}$ the set of distribution functions over $\mathcal{S}_f$. Define $Q(\hat{s}) = \{q \in \mathcal{Q} \mid q(f') = \mathbf{P}[f' | \theta(f), d, z] \forall f' \in \mathcal{S}_f \text{ and } \forall f \in \mathcal{S}_f(\hat{s})\}$, i.e., the set of distributions over f' starting from any fringe firms states that is consistent with $\hat{s}$. Define the operator

$$(T_q \hat{J})(x, \hat{s}) = \sup_{\substack{\iota \in \mathcal{I} \\ \rho \geq 0}} \sup_{q \in Q(\hat{s})} \left\{ \pi(x, \hat{s}) + \mathbf{E} \left[\phi \mathbf{1}\{\phi \geq \rho\} \right. \right. \quad (25)$$

$$\left. \left. + \mathbf{1}\{\phi < \rho\} \left[-c(x, \iota) + \beta \mathbf{E}_q[\hat{J}(x_{i,t+1}, (f_{t+1}, d_{t+1}, z_{t+1})) \mid x_{it} = x, \hat{s}_t = \hat{s}, \iota] \right] \right] \right\}, \quad (26)$$

where in the second expectation the distribution of f_{t+1} is q and $\mathbf{P}$ is the distribution of $(x_{t+1}, d_{t+1}, z_{t+1})$. It follows that $T_R \hat{J} = T_Q \hat{J}$. To see this note that any f consistent with θ that achieves the maximum in one operator can be chosen in the other. This construction falls in the framework of Iyengar (2005) and the lemma follows from Theorem 3.2 there. $\square$

Proof of Theorem 8.1. Take vectors $\hat{J} \in \hat{\mathcal{J}}$ and J , where $J : \mathcal{X} \times \mathcal{S} \rightarrow \mathbb{R}$, such that $J(x, s) = \hat{J}(x, \hat{s})$ for all $\hat{s}$ that is consistent with s . Let T^* be the bellman operator associated with θ^* . It follows that $T^* J(x, s) \leq \hat{T} \hat{J}(x, \hat{s})$ for all x and all $\hat{s}$ and s consistent. By monotonicity of $\hat{T}$ we conclude that $(T^*)^k J(x, s) \leq \hat{T}^k \hat{J}(x, \hat{s})$ for all $k \geq 1$. Taking k to infinity we get $\hat{T}^k \hat{J} \rightarrow \hat{V}$ from Lemma 8.1 and $(T^*)^k J \rightarrow \bar{V}^*$. Therefore, $\bar{V}^*(x, s) \leq \hat{V}(x, \hat{s})$ for s and $\hat{s}$ consistent. $\square$