A Framework for Dynamic Oligopoly in Concentrated Industries[*]

Bar Ifrach[†] Gabriel Y. Weintraub[‡]

August, 2012

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, in which 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, 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. We illustrate our methods with computational experiments and show that they work well in empirically relevant models, and significantly extend the class of dynamic oligopoly models that can be studied computationally. In addition, our methods can also be used to improve approximations in other settings such as dynamic industry models with a continuum of firms and an aggregate shock and stochastic growth models.

1 Introduction

Ericson and Pakes (1995)-style dynamic oligopoly models (hereafter, EP) offer a framework for modeling dynamic industries with heterogeneous firms. The main goal of the research agenda put forward by EP was to conduct empirical research and evaluate the effects of policy and environmental changes on market outcomes in different industries. The importance of evaluating policy outcomes in a dynamic setting and the broad flexibility and adaptability of the EP framework has generated many applications in industrial organization and other related fields (see Doraszelski and Pakes (2007) for an excellent survey).

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 their Markov perfect equilibrium (MPE). With recent estimation

[*]We would like to thank Vivek Farias for numerous fruitful conversations, valuable insights, important technical input, and careful reading of the paper. We have had helpful conversations with Lanier Benkard, Jeff Campbell, Allan Collard-Wexler, Dean Corbae, Pablo D'Erasmo, Uli Doraszelski, Przemek Jeziorski, Ramesh Johari, Boyan Jovanovic, Sean Meyn, Ariel Pakes, John Rust, Bernard Salanie, Minjae Song, Ben Van Roy, Daniel Xu, as well as seminar participants at UT Austin, Maryland, Informs, IIOC, Society for Economic Dynamics Conference, Econometric Society Summer Meeting, MSOM Conference, and Behavioral and Quantitative Game Theory Conference.

[†]Stanford University, Management Science and Engineering, (bifrach@stanford.edu)

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

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, in the EP framework solving for 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 (2011)). However, in practice computational concerns have typically limited MPE analysis to industries with just a handful of firms, far less 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 oligopolies. Our framework is suited for 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 with few large national firms and many small local firms. Such industries are intractable in the standard EP framework due to the large number of fringe firms. Although individual fringe firms have negligible market power, they may have significant cumulative market share 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 *industry state* is a vector (or "distribution") encoding the number of firms in 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. 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 grows. This renders commonly used dynamic programming algorithms to compute MPE infeasible in many problems of practical interest.

In this work we introduce a new and behaviorally appealing model that 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, we postulate what we believe is a plausible model of firms' behavior: firms closely monitor dominant firms, but keep track of the remainder of the industry—fringe firms—in a less detailed way. More specifically, we assume that firms' strategies depend on (1) the detailed state of dominant firms; and (2) few aggregate statistics (such as few moments) of the distribution of fringe firms (we refer to the fringe firms' distribution or state interchangeably). We call these strategies, *moment-based strategies*, where we use the term 'moments' generically as firms could keep track of other statistics of the fringe firm distribution such as un-normalized moments or quantiles. Based on these strategies, we introduce an equilibrium concept that we call *Moment-Based Markov Equilibrium (MME)*. In MME firms' beliefs on the evolution of the industry satisfy a consistency condition, and firms play their optimal moment based strategies given these beliefs. A MME in which firms keep track of few moments of the fringe firm state is both computationally tractable and, in principle, behaviorally appealing.

A natural question that arises, however, is whether MME strategies remain optimal in a larger set of strategies that may depend on the full fringe firm state, i.e., not solely on few moments. Unfortunately, it is simple to observe that MME strategies are not necessarily optimal; a single firm may increase its profits by unilaterally deviating to such a strategy while competitors continue to play MME strategies. 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 distributions consistent with it, from which the future evolution of the industry is very different. More formally, moments may not induce a sufficient partition of histories and may not summarize all payoff relevant history in the sense of Maskin and Tirole (2001); observing the distribution 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.

Hence, even if competitors use moment-based strategies, a moment-based strategy may not be close to a best response, and, therefore, MME strategies may not be close to a subgame perfect equilibrium. On one hand, we have a computationally tractable model that, in principle, is also behaviorally appealing. On the other, the resulting strategies may not be optimal in a 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, a 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 for fringe firms and may be too restrictive for many empirical applications.

In the second approach we impose less severe restrictions on the model, but instead restrict the strategies fringe firms can play. 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 show that even under this restriction our model generates interesting strategic interactions between the dominant and fringe firms. We also provide a method to ex-post test how severe is our restriction on fringe firms' strategies. We note that we do not restrict the strategies of dominant firms, hence, they become optimal as the number of fringe firms grow. Because the first two approaches only impose restrictions over fringe firms, they may be particularly relevant for applications in which dominant firms are the key focus of analysis and a detailed model of the fringe is not required.

In the third approach we do not restrict the model nor the strategies of fringe firms. Instead we suppose that firms (perhaps 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. Here, 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 novel computationally tractable error bound that measures the extent of sub-optimality of MME strategies in terms of a unilateral deviation, thus relating it to the notion of ϵ-equilibrium. 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. If the bound is small, it may be reasonable to assume firms use MME strategies, as unilaterally deviating to more complex strategies does not significantly increase profits, and doing so may involve costs associated with gathering and processing more information.

We propose computationally efficient algorithms to compute our equilibrium concepts and show that they work well in important classes of models. Specifically, we show that MME generates interesting strategic interactions between dominant and fringe firms, in which for example, dominant firms make investment decisions to deter growth and entry from fringe firms. We also show the applicability of our robust error bound.

To further illustrate the applicability of our model and methods we show how they can be used to endogeneize the industry market structure in a fully dynamic model. In particular, we perform numerical experiments motivated by the long concentration trend in the beer industry in the US during the years 1960-1990. Over the course of those years, the number of active firms dropped dramatically, 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 of the beer industry 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 an industry with hundreds of firms. As further evidence of the usefulness of our approach, Corbae and D'Erasmo (2012) has already used our framework to study the impact of capital requirements in market structure in a calibrated model of banking industry dynamics with dominant and fringe banks.

In summary, our approach offers a computationally tractable model for industries with a dominant/fringe market structure, capturing important and novel strategic interactions, and opens the door to studying novel issues in industry dynamics. As such, our model greatly increases the applicability of dynamic oligopoly models.

Finally, our moment based strategies are similar and follow the same spirit of the seminal paper by Krusell and Smith (1998) that replaces the distribution of wealth over agents in the economy by its moments when computing stationary stochastic equilibrium in a stochastic growth model. While our main focus has been on dynamic oligopoly models, our methods can also be used to find better approximations in stochastic growth models, as well as in macroeconomic dynamic industry models with an infinite number of heterogeneous firms and an aggregate shock (see, e.g., Khan and Thomas (2008) and Clementi and Palazzo (2010)).

The rest of this paper is organized as follows. We discuss related literature in Section 2. Section 3 describes our dynamic oligopoly model. Section 4 introduces our new equilibrium concept. Then, Sections 5, 6, and 7, describe our three approaches to deal with the optimality of MME strategies. Section 8 develops the error bound for the third approach. Section 9 describes the extensions of our methods to the models mentioned in the previous paragraph, and Section 10 concludes.

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 MPE computation, and how our methods relate to those approaches. Second, we discuss other methods that are being developed to alleviate the burden when computing equilibria in dynamic oligopoly models.

Researchers have used in practice different approaches to 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) and Gallant et al. (2010)). 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'Erasmo (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 the industry state information, is assumed to be Markov (e.g., see Kalouptsidi (2011), 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). This relates to our third approach (Section 7) and 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 counterfactuals. In particular, Xu (2008), Qi (2008), Iacovone et al. (2009), and Thurk (2009) among others use the notion of oblivious equilibrium (OE) 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 to industries that are more concentrated.

Our work is related to recent work that alleviates the burden when computing equilibria in dynamic oligopoly models. Farias et al. (2012) uses approximate dynamic programming with value function approximation to approximate MPE. Santos (2012) introduces a state aggregation technique based on quantiles of the industry state distribution to try to break the curse of dimensionality. Fershtman and Pakes (2010) introduces the notion of experienced based equilibria that weakens the restrictions imposed by perfect Bayesian equilibrium for dynamic games of asymmetric information. As we explain later, the approach by Fershtman and Pakes has connections with the approach we discuss in Section 7.

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 only keep track of the state of dominant firms and assume that at every point in time the fringe firm state is equal to the expected state conditional on the state of dominant firms. Our paper builds on Benkard et al. (2011)'s idea of considering a dominant/fringe market structure; however, we believe our approach offers several important contributions with respect to them. First, in our paper firms not only keep track of the dominant firms' states but also of fringe moments. For this reason, our approach should generally produce best responses that are closer to optimal compared to theirs. Second, different to Benkard et al. (2011), our model allows for firms to switch between the fringe and dominant tier, and hence, to fully endogenize the market structure in equilibrium. Third, we present models for which MME strategies are in fact optimal and, in addition, our error bound is valid for dominant firms. On the other hand, the approach of Benkard et al. (2011) does not impose the restriction that the single-period profit function depend only on few moments of the fringe firm state, as we will require below.

We believe that all the previous approaches are useful complements and have different strengths that may be important for different industries and empirical applications. For example, if OE is the starting point of analysis, then the approach with dominant firms provided by Benkard et al. (2011) is a natural refinement. If researchers want to use a computational approach to approximate MPE, then the method by Farias et al. (2012) may be more appropriate. Finally, we believe that among all these papers our approach offers a natural and appealing behavioral model with the advantages discussed in the previous paragraph.

3 Dynamic Oligopoly 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).

Time Horizon. The industry evolves over discrete time periods and an infinite horizon. We index time periods with nonnegative integers $t \in \mathbb{N}$ ($\mathbb{N} = \{0, 1, 2, \ldots\}$). 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. 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. At each time $t \in \mathbb{N}$, we denote the number of incumbent firms as n_t. We assume $n_t \leq N, \ \forall t \in \mathbb{N}$, where the integer number N represents the maximum number of incumbent firms that the industry can accommodate at every point in time.

State Space. Firm heterogeneity is reflected through firm states that represent the 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 $x_{it} \in \mathcal{X} \subseteq \Re^q, \ q \geq 1$. We define the *industry state* $\bar{s}_t$ to be a vector that encodes the individual states of all incumbent firms at time t: $\bar{s}_t = \{x_{it}\}_{i \in S_t}$.

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.

Entry process. We consider an entry process similar to the one in Doraszelski and Pakes (2007). At time period t, there are $N - n_t$ potential entrants, ensuring that the maximum number of incumbent firms that the industry can accommodate is N (we assume $n_0 < N$). Each potential entrant is assigned a unique 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. We assume the random variables $\{\kappa_{it} | t \geq 0 \, i \geq 1\}$ are i.i.d. and independent of all previously defined random quantities, and have a well-defined density function. If the entry cost is below the expected value of entering the industry then the firm will choose to enter. Potential entrants make entry decisions simultaneously. Entrants appear in the following period at state $x^e \in \mathcal{X}$ and can earn profits thereafter.[1] 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 next period.

Transition dynamics. If an incumbent firm decides to remain in the industry, it can take an action to improve its individual state. Let $\mathcal{I} \subseteq \Re_+^k$ ($k \geq 1$) 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 transition to a state at time $t+1$ is described by the following Markov kernel $\mathbf{Q}$:

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

Uncertainty in state transitions may arise, for example, due to the risk associated with a research and development endeavor or a marketing campaign. The dependence of the kernel $\mathbf{Q}$ on the industry state allows, for example, for the existence of investment spillovers across firms. 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 all previously defined random quantities.

Aggregate shock. There is an aggregate profitability shock z_t that is common to all firms. 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 at time period t are given by $\pi(x_{it}, \bar{s}_t, z_t)$, that depend on its individual state x_{it}, the industry state $\bar{s}_t$, and the value of the aggregate shock z_t.

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; (5) 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 discounted profits. 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.

Equilibrium. The most commonly used equilibrium concept in such dynamic oligopoly models is that of

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

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). This motivates our alternative approach.

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.

4.1 Dominant and Fringe Firms

We focus on industries that exhibit 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 incumbent 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; this is discussed in Subsection 7.2.

The specific division between dominant and fringe firms will depend on the specific application at hand. Typically, however, dominant firms will usually be market share leaders or, more generally, firms that most affect competitors' profits, as illustrated in the next example. 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 may then be natural to separate dominant firms from fringe firms by an exogenous threshold state $\overline{x}$, such that $i \in D_t$ if and only if $x_{it} \geq \overline{x}$.

We let $\mathcal{X}_f \subseteq \mathcal{X}$ and $\mathcal{X}_d \subseteq \mathcal{X}$ be the set of feasible individual states for fringe and dominant firms, respectively. We assume that $\mathcal{X}_f \cap \mathcal{X}_d = \emptyset$. This is done with out loss of generality, because we can always append one dimension to the individual state indicating whether the firm is fringe or dominant. This construction is useful because the individual state now encodes whether the firm is fringe or dominant allowing, for example, for fringe and dominant firms to have different model primitives and strategies.

Because we will focus on equilibrium strategies that are anonymous with respect to the identity of firms, we define the state of fringe firms f_t to be a vector over individual states that specifies, for each fringe firm state $x \in \mathcal{X}_f$, the number of incumbent fringe firms at x in period t. We define $\mathcal{S}_f = \left\{ f \in \mathbb{N}^{|\mathcal{X}_f|} \middle| \sum_{x \in \mathcal{X}_f} f(x) \leq N \right\}$ to be the set of all possible states of fringe firms. We call f_t the state or the distribution of fringe firms. 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

$\mathcal{S}_d = \{d \in \mathcal{X}_d^k \Big| k \leq \overline{D}\}$, where $\overline{D}$ is the maximum number of dominant firms the industry can accommodate. We define the state space $\mathcal{S} = \mathcal{S}_f \times \mathcal{S}_d \times \mathcal{Z}$, where an industry state $s \in \mathcal{S}$ is given by a distribution of fringe firms, a state for dominant firms, and the aggregate shock.

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 state of dominant firms, compared to the state of fringe firms. 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. Instead, we postulate that firms only keep track of the state of dominant firms and of few summary statistics of the fringe firms' state distribution. Not only do we think this provides an appealing model of firms' behavior, but it will also make the equilibrium computation feasible.

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. A set of such summary statistics is a multi-variate function $\theta : \mathcal{S}_f \to \Re^n$. For example, when the fringe firm state is discrete and one-dimensional, $\theta(f) = \sum_{y \in \mathcal{X}_f} y^\alpha f(y)$ is the $\alpha-$th un-normalized moment with respect to the distribution f. For brevity and concreteness, we call such summary statistics *fringe firm moments* with the understanding that they could include quantities such as normalized or un-normalized moments, but also quantiles or other functions of the distribution of fringe firms. We introduce the following simplifying assumption that we keep throughout the paper. We briefly discuss how to relax this assumption in Section 10.

Assumption 4.1. *The single period expected profits of firm i at time t, $\pi(x_{it}, \tilde{\theta}_t, d_t, z_t)$, depend on its individual state x_{it}, a vector $\tilde{\theta}_t \in \Re^l$ of fringe firm moments, the state of dominant firms d_t, and the value of the aggregate shock z_t. The transition kernel of firm i at time t, $\mathbf{Q}[\cdot|x_{it}, \iota_{it}, \tilde{\theta}_t, d_t]$ depends on its individual state x_{it}, its investment ι_{it}, a vector $\tilde{\theta}_t$ of fringe firm moments, and the state of dominant firms d_t.*

Note that we slightly abused the notation to re-define the single period profit function and the transition kernel so that their dependence on the state of fringe firms is only through their moments.[2] In our approach, firms will keep track of the moments that determine the profit function and transition function, $\tilde{\theta}_t$. The state space spanned by these moments is much smaller than the original one if l is low dimensional and significantly smaller than $|\mathcal{X}_f|$. In fact, in many applications of interest the transition function is independent of the industry state $(\tilde{\theta}_t, d_t)$, 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)). We describe another important example below.

[2] Also, note that assuming that both functions depend on the same set of moment is done without loss of generality.

Example 4.1 (**Quality-Ladder).** 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 a natural number x_{it} (and another component, that we ignore for simplicity, indicating wether the firm is fringe or dominant). For clarity, we do not consider aggregate shocks in this example.

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:

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

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 standard logit model.

All firms share the same constant marginal cost c. We assume that dominant firms compete Nash in prices with resulting equilibrium prices denoted by p^*_{it}. Similarly to the logit model of monopolistic competition of Besanko et al. (1990), we assume fringe firms set prices assuming they have no market power, so they all set the price $p^* = (Y + c\alpha_2)/(1 + \alpha_2)$. Let $K(x_{it}, p_{it}) = \exp(\alpha_1 \ln(x_{it}) + \alpha_2 \ln(Y - p_{it}))$. It is simple to see that expected profits are then given by:

$$\pi(x_{it}, \bar{s}_t) \quad = \quad m(p^*_{it} - c)\frac{K(x_{it}, p^*_{it})}{1 + (Y - p^*)^{\alpha_2} \sum_{y \in \mathcal{X}_f} y^{\alpha_1} f_t(y) + \sum_{j \in D_t} K(x_{jt}, p^*_{jt})} \ , \ \forall i \in S_t.$$

Therefore, $\pi(x_{it}, \bar{s}_t)$ can be written as $\pi(x_{it}, \tilde{\theta}_t, d_t)$, where $\tilde{\theta}_t$ is 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 or moments of the fringe firm state. 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

$$\theta_t = \theta(f_t) = (\tilde{\theta}_t, \overline{\theta}_t), \tag{2}$$

where $\tilde{\theta}_t$ satisfies Assumption 4.1, and $\overline{\theta}_t$ are additional moments included in θ_t. We define $\mathcal{S}_\theta$ as the set of admissible moments defined by (2). 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 $\rho : \mathcal{X} \times \hat{\mathcal{S}} \rightarrow \Re$ 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 functions by Λ, where an element of Λ is a function $\lambda : \hat{\mathcal{S}} \rightarrow \Re$. It is assumed that all entrants are fringe, that is $x^e \in \mathcal{X}_f$. Note that strategies and the state space are defined with respect to a specific function of moments (2).

With Markov strategies (μ, λ) the *underlying industry state*, $\{s_t = (f_t, d_t, z_t) : t \geq 0\}$, is a Markov process. We denote its transition kernel by $\mathbf{P}_{\mu,\lambda}$. In addition, we denote by $\mathbf{P}_{\mu',\mu,\lambda}$ the transition kernel of (x_{it}, s_t) when firm i uses strategy μ', and its competitors use strategy (μ, λ). Note that given strategies, both these kernels can be derived from the primitives of the model, namely, the distributions of ϕ and κ, the kernel of the aggregate shock, and the kernel $\mathbf{Q}$. We emphasize that the underlying industry state s_t should be distinguished from the moment-based industry state $\hat{s}_t$.

4.4 Moment-Based Markov Equilibrium

A *moment-based Markov equilibrium* (MME) is an equilibrium in moment-based strategies, as will be defined next.

Defining our notion of equilibrium in moment based strategies will require the construction of what can be viewed as a 'Markov' approximation to the dynamics of the moment-based industry state process $\{(x_{it}, \hat{s}_t) : t \geq 0\}$, where i is some generic firm. Note that this process is, in general, *not* Markov even if the dynamics of the underlying industry state $\{(x_{it}, s_t) : t \geq 0\}$ are. To see this, consider Example 4.1 where firms keep track of a single moment of the fringe firm state and for simplicity, assume that $\alpha_1 = 1$. Then, $\theta_t = \theta(f_t) = \sum_{y \in \mathcal{X}_f} y f_t(y)$, so firms only keep track of the first un-normalized moment of the fringe firm state. Suppose the current value of that moment is $\theta_t = 10$; this value 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 probabilistic distribution for the 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. In the process of aggregating information via moments, information is lost, and the resulting process is no longer Markov.

We note that for this reason the introduction of additional *contemporaneous* moments $\overline{\theta}_t$ beyond the moments $\tilde{\theta}_t$ in equation (2) can improve the predictions of the future evolution of the industry. Alternatively, it is possible to define an equilibrium concept where in equation (2) firms keep track of past values of the moment $(\tilde{\theta}_{t-1}, \tilde{\theta}_{t-2}, ...)$ and our methods can be extended accordingly. However, to provide a more direct connection with the commonly used concept of MPE, in which firms keep track only of the current industry state, and to simplify exposition, in this paper we assume that MME strategies only depend on current moments of the fringe firm state.

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 (non-Markov) process $\{(x_{it}, \hat{s}_t) : t \geq 0\}$. To this end, let us suppose that a kernel $\hat{\mathbf{P}}_{\mu,\lambda}$ describes the evolution of an hypothetical Markov process on $\mathcal{S}_\theta$. Having supposed this kernel, the kernel $\hat{\mathbf{P}}_{\mu',\mu,\lambda}[\cdot|\cdot]$ is now define according to:

$$\hat{\mathbf{P}}_{\mu',\mu,\lambda}[x', \hat{s}'|x, \hat{s}] = \mathbf{P}_{\mu',\mu,\lambda}[x', d', z'|x, \hat{s}]\hat{\mathbf{P}}_{\mu,\lambda}[\theta'|\hat{s}], \tag{3}$$

where with some abuse of notation $\mathbf{P}_{\mu',\mu,\lambda}[x', d', z'|x, \hat{s}]$ 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 $\hat{\mathbf{P}}_{\mu',\mu,\lambda}$ 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, the firm correctly assesses the distribution of its next state, the next state of dominant firms, and the next aggregate shock. Note that because firms use moment based strategies, the moment based state $(x, \hat{s})$ is enough to determine the transition probabilities of (x, d, z) according to the transition kernel of the underlying industry state $\mathbf{P}_{\mu',\mu,\lambda}$.

However, 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 potentially depends on the distribution of the fringe firms beyond simply its moments.

In the spirit of approximating the actual moment process, we ask that the perceived transitions described by $\hat{\mathbf{P}}_{\mu,\lambda}$ agree in some manner with the actual transitions observed in equilibrium. In particular, recall that $\mathbf{P}_{\mu,\lambda}$ denotes the transition kernel of the underlying (Markovian) industry state $\{s_t : t \geq 0\}$ when all firms use the strategy (μ, λ). We specify the perceived transition kernel $\hat{\mathbf{P}}_{\mu,\lambda}$ as some transformation of the actual transition 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 (Empirical transitions)**.** *A natural definition for $\hat{\mathbf{P}}_{\mu,\lambda}$ is the kernel that coincides with the long-run average* empirical transitions *from the moment in the current time period to the moment the next time period under strategies (μ, λ). More specifically, we let the industry evolve for a long time under strategies (μ, λ). For each moment based state $\hat{s}$ that is visited, we observe a transition to a new moment θ'. We count the empirical frequency of these transitions and set the kernel $\hat{\mathbf{P}}_{\mu,\lambda}[\theta'|\hat{s}]$ to be the empirical distribution corresponding to these frequencies. A similar construction is used by Fershtman and Pakes (2010) in a setting with asymmetric information.*

We now formalize this definition. We assume a finite state space to simplify the exposition. Recall that the evolution of the underlying industry state is described by the kernel $\mathbf{P}_{\mu,\lambda}$. *We let* $\mathcal{R}$ *be the recurrent class of moment-based states the industry will eventually reach. For all* $\theta' \in \mathcal{S}_\theta$ *and* $\hat{s} \in \mathcal{R}$, *we define the operator:*

$$(\Phi \mathbf{P}_{\mu,\lambda})(\theta'|\hat{s}) = \lim_{T\to\infty} \frac{\sum_{t=1}^{T} \mathbf{1}\{\hat{s}_t = \hat{s},\ \theta_{t+1} = \theta'\}}{\sum_{t=1}^{T} \mathbf{1}\{\hat{s}_t = \hat{s}\}}.$$

We require that

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

for all $\hat{s} \in \mathcal{R}$, *and* $\hat{\mathbf{P}}_{\mu,\lambda}$ *is defined arbitrarily outside this set.*

Having thus defined a Markov process approximating the process $\{(x_{it}, \hat{s}_t) : t \geq 0\}$, 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

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

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 $\hat{\mathbf{P}}_{\mu',\mu,\lambda}$.[3] We will use the shorthand notation $V(x, \hat{s}|\mu, \lambda) \equiv V(x, \hat{s}|\mu, \mu, \lambda)$ 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 (2) and, for every strategy (μ, λ), a transition kernel $\hat{\mathbf{P}}_{\mu,\lambda}$ defined via an operator Φ.

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

$\mathcal{C}\mathbf{1}$: *Incumbent firm strategies optimization:*

$$\sup_{\mu' \in \mathcal{M}} V(x, \hat{s}|\mu', \mu, \lambda) = V(x, \hat{s}|\mu, \lambda) \qquad \forall x \in \mathcal{X},\ \forall \hat{s} \in \hat{\mathcal{S}}. \tag{4}$$

$\mathcal{C}\mathbf{2}$: *At each state, the cut-off entry value is equal to the expected discounted value of entering the industry:*

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

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

$\mathcal{C}3$: *The perceived transition kernel is given by*

$$\hat{\mathbf{P}}_{\mu,\lambda} = \Phi \mathbf{P}_{\mu,\lambda} \tag{5}$$

Note that because the individual state encodes whether a firm is fringe or dominant, fringe and dominant firms may have different value functions and MME strategies. Also, note that if the function θ is the identity, so that $\theta(f) = f$, 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. Proving existence of MME is model dependent. For example, if the state space is finite one can use a similar argument to Doraszelski and Satterthwaite (2010). We note that in the numerical experiments that we present later in the paper we were always able to computationally find an MME over a large set of primitives that satisfy the standard assumptions required for existence of MPE in EP-style models. With respect to uniqueness, in general we presume that our model may have multiple equilibria.

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 MME agents optimize over low dimensional strategies so it is a computationally tractable equilibrium concept. Moreover, MME also provides what we think is an appealing behavioral model. Theoretically, an MME is appealing if the perceived process of moments is close to the *actual* process of moments. This is related to the optimality of moment-based strategies that we study in the next section.

4.5 Optimality of Moment-Based Strategies

Suppose firms play moment based strategies with moments $\theta(\cdot)$ 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 underlying fringe firm state f_t when dynamics are governed by the *primitive* transition kernel $\mathbf{P}_{\mu,\lambda}$. As previously noted, generally $\{\hat{s}_t : t \geq 0\}$ is not Markov, so it may not be a sufficient statistic to predict the future evolution of the industry. Hence, $\theta(\cdot)$ does not summarize all payoff relevant history in the sense of Maskin and Tirole (2001). As such, observing the underlying fringe firm state may provide valuable information for decision making. It is important to note that while some Φ operators may provide better approximations than others, it may be that few 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 competitors use moment-based strategies, a moment-based strategy may not be close to a best response, and therefore, MME strategies may not be 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 a 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 a finite 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). See Section 5 for a discussion of this approach.

While simple and elegant, these models impose restrictions on the model primitives for fringe firms and may be too restrictive for many applications; this motivates the next approach.

2. In the second approach we consider less restrictive assumptions on the model primitives, but we restrict the 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. Note that we do not restrict the strategies of dominant firms, and therefore, their strategies become optimal as the number of fringe firms increase. 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 assume (perhaps wrongly) that moments form a Markov process and summarize all payoff relevant information. The usefulness of this approach relies on a good choice of the set of moments and on the construction of Φ. Importantly, we introduce a computationally tractable error bound that measures the extent of the sub-optimality of MME 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 asses whether the moments summarize well enough the underlying fringe firm state. See Sections 7 and 8 for a discussion of this approach.

We conclude the section by formalizing the sense in which MME strategies become optimal. 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 functions, respectively, defined with respect to θ^*. Note that $\mathcal{M}^*$ and Λ^* are the sets 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:

$$\Delta_{\mu,\lambda}(x, s) = \sup_{\mu' \in \mathcal{M}^*} V^*(x, s|\mu', \mu, \lambda) - V^*(x, s|\mu, \lambda). \tag{6}$$

We will use the value of the full information deviation to measure the extent of sub-optimality of MME strategies.[4] If this value is small, the MME strategy achieves essentially the same profits compared to the best possible unilateral deviation Markov strategy that keep track of all available information. In this sense, the value of the full information deviation is similar to the notion of $\epsilon-$equilibrium. Note that we expect this value to be small when moments in MME are close to being *sufficient statistics* of the future evolution of the industry. That is, when $\hat{\mathbf{P}}_{\mu,\lambda}[\theta'|\hat{s}] \approx \mathbf{P}_{\mu,\lambda}[\theta'|s]$, so the perceived transition kernel is close to the primitive transition kernel, and the industry state s does not provide additional information to predict the future industry evolution beyond the moment based industry state $\hat{s}$.

[4]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.

5 First Approach: Moments Become Sufficient Statistics

In this section we present a class of industry dynamic models for which a succinct set of moments essentially summarize all payoff relevant information and is close to being sufficient statistics. In these models, the value of the full information deviation is zero, or becomes zero asymptotically as the number of fringe firms grows large.

A simple class of models for which this holds is when fringe firms are homogeneous, in the sense that $\mathcal{X}_f$ is a singleton. In this case, a single moment, namely the number of incumbent fringe firms, is a sufficient statistic of the future evolution of the industry. Next, we describe a model with firm fringe heterogeneity that does not allow for entry and exit, in which moments also become sufficient statistics as the number of fringe firms grows large.

Constant Returns to Scale Model for Fringe Firms. We take $\mathcal{X}_f = \Re^+$ (with another component that we suppress for clarity indicating that the firm is fringe). We assume that there is no entry and exit (therefore $\mu = \iota$), and that there are $N - \overline{D}$ fringe firms. 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 un-normalized moment of the fringe firm state, that is, $\theta_t = \sum_{x \in \mathcal{X}_f} x f_t(x)$. We make the following assumptions on the primitives of fringe firms *only*; importantly, no restrictions are placed on the primitives of dominant firms.

1. The single period profit is linear in the fringe firm's own state, $\pi(x, \hat{s}) = x\pi_1(\hat{s}) + \pi_0(\hat{s})$ with $\sup_{\hat{s} \in \hat{\mathcal{S}}} \{\pi_1(\hat{s}), \pi_0(\hat{s})\} < \infty$. 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 the state. Formally, $c(x, \iota) = (cx)\iota$ with $\iota \in \mathcal{I} = \Re_+$.

3. The dynamics of a fringe firm's evolution are linear in its own state: $x_{i,t+1} = x_{it}\zeta_1(\iota, \hat{s}_t, w_{1it}) + \zeta_0(\hat{s}_t, w_{0it})$, where ι is the amount invested. Each of the sequences of random variables $\{w_{0it} | t \geq 0,\ i \geq 1\}$ and $\{w_{1it} | t \geq 0,\ i \geq 1\}$ is i.i.d. and independent of all previously defined random quantity and of each other. In addition, we assume the functions ζ_0 and ζ_1 are uniformly bounded over all realizations, investment levels, and industry states. The assumption about linear transitions is similar to assuming Gibrat's law in firm's transitions (Sutton, 1997).

In addition, we assume that, under any strategy played by competitors, there exists a unique solution to each firm's investment optimization problem (see Doraszelski and Satterthwaite (2010) for a sufficient condition).

We begin by showing that for any perceived kernel $\hat{\mathbf{P}}_\mu$ the corresponding best response investment strategy for a fringe firm does not depend on its own individual state.[5] All proofs are relegated to Appendix A.

[5] This result is similar to Lucas and Prescott (1971) that studies a dynamic competitive industry model with similar assumptions to ours, but with deterministic transitions and stochastic demand.

Lemma 5.1. *Consider the constant returns to scale model described above. For any MME strategy μ and for every $x, x' \in \mathcal{X}_f$ and $\hat{s} \in \hat{\mathcal{S}}$ we have $\mu(x, \hat{s}) = \mu(x', \hat{s}) = \mu(\hat{s})$.*

We use this result to characterize the evolution of the moment under MME strategies as the number of fringe firms grows. To obtain a meaningful asymptotic regime, we scale the market size together with the number of fringe firms. Formally, we consider a sequence of industries with growing market size m. As specified above, our model does not explicitly depend on market size. However, market size would typically enter the profit function, through the underlying demand system like in Example 4.1. Therefore, we consider a sequence of markets indexed by market sizes $m \in \mathbb{N}$ with profit functions denoted by π^m. We assume the number of firms increases proportionally to the market size, that is $N^m = Nm - \overline{D}$, for some constant $N > 0$. We assume that all other model primitives, including the maximum number of dominant firms, are independent of m. Quantities associated to market size m are indexed with the superscript m. We state the following assumption.

Assumption 5.1. *Let $\{\mu^m : m \geq 1\}$ be a sequence of MME strategies followed by all firms in market m. Then, there exists a compact set $\overline{\mathcal{X}}_f \subset \mathcal{X}_f$ such that for all $m \geq 0$, $t \geq 0$, and $i \in F_t^m$, $\mathcal{P}[x_{it}^m \in \overline{\mathcal{X}}_f] = 1$.*

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 the following result.

Proposition 5.1. *Consider the constant returns to scale model under Assumption 5.1, and suppose that for every $m \geq 1$ all firms use MME strategy μ^m. For a given t, conditional on the realizations of $\{x_{it} \in \overline{\mathcal{X}}_f | i \in F_t^m\}$ and $\hat{s}_t^m$, we have*

$$(1/N^m)\theta_{t+1}^m - \left[\tilde{\zeta}_1^m(\hat{s}_t^m)\,(1/N^m)\,\theta_t^m + \tilde{\zeta}_0(\hat{s}_t^m)\right] \to 0, \ a.s.,$$

as $m \to \infty$, where $\theta_t^m = \sum_{i \in F_t^m} x_{it} = \sum_{x \in \mathcal{X}_f} x f_t^m(x)$, $\tilde{\zeta}_1^m(\hat{s}_t^m) = \mathrm{E}_{\mu^m}[\zeta_{1it}(\mu^m(\hat{s}_t^m), \hat{s}_t^m, w_{1it})]$, and $\tilde{\zeta}_0(\hat{s}_t^m) = \mathrm{E}_{\mu^m}[\zeta_{0it}(\hat{s}_t^m, w_{0it})]$.

The result shows that the first moment becomes a sufficient statistic for the evolution of the next moment. In particular, the next moment becomes a linear function of the current moment. Heuristically, this suggests defining the following perceived transition kernel:[6]

$$\hat{\mathbf{P}}_{\mu^m}[\tilde{\zeta}_1^m(\hat{s}_t^m)\theta_t^m + N^m\tilde{\zeta}_0(\hat{s}_t^m)|\hat{s}_t^m] = 1, \tag{7}$$

which will become a good approximation as m grows large. In fact, with this perceived transition kernel and under suitable continuity conditions, one can show that the value of the full information deviation (appropriately normalized) converges to zero as the market size m approaches infinity.[7]

[6] Note that a derivation similar to that in Proposition 5.1 will show that any k-th moment of fringe firms' states for an integer k would depend on moments $k, k-1, \ldots, 1$ only, as m grows large. Therefore, if higher integer moments are payoff relevant they could be accounted for as well.

[7] This result requires appropriate continuity conditions on the model primitives, and also that the equilibrium investment strategy is continuous in the moment-based fringe state. The result is omitted for brevity and can be obtained from the authors upon request.

The previous result provides the ideal setting for using MME; its equilibrium strategies yield moments that form a Markov process and hence summarize all payoff relevant information. While we did not restrict the model primitives for dominant firms, we impose several restrictions on the model primitives for fringe firms. These restrictions limit the dependence of fringe firms' equilibrium strategies on their own state and, in this way, the aggregation of fringe firms' transitions gives rise to Markov moments. Unfortunately, this type of aggregation is quite limited. To illustrate this, our next result shows that when firms keep track of one moment, moments are sufficient statistics only if fringe firms' *equilibrium transitions* are, in an appropriate sense, linear in their own state. To our knowledge, such equilibrium transitions arise only in the constant returns to scale model presented here or in close variations of it. There, as we saw, equilibrium transitions are linear, because the primitive transitions are linear and the equilibrium investment strategy does not depend on the fringe firm individual state.

Proposition 5.2. *Assume $\mathcal{X}_f$ is a closed interval in $\Re$, $N - \bar{D} \geq 3$, there is no entry and exit in the industry, and the set of moments contains the α un-normalized moment only, that is, $\theta(f) = \sum_{x \in \mathcal{X}_f} x^\alpha f(x)$. Suppose that under MME strategy μ, the moment is a sufficient statistic for the evolution of the industry, that is:*

$$\hat{\mathbf{P}}_{\mu,\lambda}[\theta'|\hat{s}] = \mathbf{P}_{\mu,\lambda}[\theta'|s],$$

for all $\theta' \in \mathcal{S}_\theta$, $s = (f, d, z) \in \mathcal{S}$, and $\hat{s} = (\theta, d, z) \in \hat{\mathcal{S}}$, with $\theta(f) = \theta$. Then, fringe firms' transitions in MME must be linear, in the sense that $\mathrm{E}_\mu[x^\alpha_{i,t+1}|x_{it} = x, \hat{s}_t] = x^\alpha \tilde{\zeta}_1(\hat{s}_t|\mu) + \tilde{\zeta}_0(\hat{s}_t)$.

Finally, we note that this first approach is closely related to the literature in macroeconomics that deals with the aggregation of macroeconomic quantities from the decisions of a single 'representative agent' (Blundell and Stoker, 2005). With heterogeneous agents, similarly to our model, this type of result is only obtained under strong assumptions on consumers' preferences that are also akin to linearity (see, for example, Altug and Labadie (1994)).

6 Second Approach: Restricting Fringe Firms' Strategies

In this section we impose less severe restrictions on the model primitives for fringe firms. Instead, we restrict and loosen the optimality requirement of fringe firms' strategies. This restriction gives rise to a similar result to the first approach: moments become sufficient statistics for the future evolution of the industry as the market size grows. We emphasize that no restrictions will be placed on the strategies or primitives of dominant firms. As such, this approach and the previous one are mostly useful when dominant firms are the key focus of analysis while the detailed model of the fringe is of lesser importance. Moreover, we note that it is possible to check ex-post how sub-optimal the restricted fringe firms' strategies are. Subsection 6.1 provides a numerical experiment that shows that this approach captures interesting strategic interactions between the dominant and the fringe firms.

Let $\tilde{\mathcal{M}} \subset \mathcal{M}$ be a set of *restricted* strategies with typical element $\tilde{\mu}$. This restriction is binding only for fringe firms, so that dominant firms play unrestricted strategies. In addition, let $\hat{\mathbf{P}}_{\tilde{\mu},\lambda}$ be a perceived transition kernel corresponding to the selection of a set of moments and the restricted fringe strategies. We define equilibrium in this approach to be MME, with one exception: fringe firms play a restricted strategy that may be sub-optimal. We require, however, that the restricted strategy played in equilibrium is derived from the optimal unrestricted strategy. Concretely, let $\mu' \in \mathcal{M}$ be the optimal unrestricted strategy (in the sense of $\mathcal{C}1$ in the definition of MME), when the remaining firms play $(\tilde{\mu}, \lambda)$. Then we say that $(\tilde{\mu}, \lambda)$ constitute an equilibrium in this second approach if: (a) for dominant firms $\tilde{\mu} = \mu'$, (b) for fringe firms $\tilde{\mu}$ is derived from μ', (c) λ and $\hat{\mathbf{P}}$ satisfy $\mathcal{C}2$ and $\mathcal{C}3$, respectively, with Φ being appropriately defined given the restricted fringe firm strategies. With some abuse of terminology we also call this restricted equilibrium concept MME.

There are many ways in which one could derive the restricted strategy to be played by the fringe from the optimal unrestricted strategies (step (b) above). For example, one could 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 next.

A Model with Restricted Fringe Firms' Strategies. In the reminder of this section, we illustrate these ideas in a model with entry and exit. The individual states of fringe firm are in $\mathcal{X}_f = [0, \bar{x}]$. Fringe firms' transition dynamics are assumed to take the form,

$$x_{it+1} = (x_{it})^{(1-p)} \zeta(\iota_{it}, w_{it}), \tag{8}$$

where $0 < p < 1$ is assumed to be small. The function ζ is uniformly bounded above and below by $\underline{\zeta}$ and $\bar{\zeta}$, respectively, over all investment levels and realizations. We assume the random variables $\{w_{it} | t \geq 0,\ i \geq 1\}$ are i.i.d. and independent of all previously defined quantities. This functional form has an appealing property: a fringe firm's state at time $t+1$ is, ceteris paribus, increasing at a diminishing rate in the state at time t. Our assumptions imply that a fringe firm cannot grow larger than $\bar{x} = (\bar{\zeta})^{1/p}$.

First, we restrict fringe firms' investment to be $\iota(x, \hat{s}) = \iota(\hat{s})$, where the investment cost is $(cx)\iota$. With a common investment strategy, the transitions in (8) implies that for a small p the growth rate of fringe firms is close to $\zeta_{it}(\iota_{it}, w_{it})$, 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. Although Gibrat's law is disputed, previous work suggests that it is a good approximation for small firms.[8]

Second, we restrict fringe firms' exit strategy so that a fringe firm in state x stays in the industry with probability $(x/\bar{x})^p$. 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. Note that 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. We can generalize the previous specification by allowing the staying probability to depend on the industry state in the following way: $\eta(\hat{s})(x/\bar{x})^p$. In this case, the industry state-dependent constants $\eta(\cdot)$ need to be determined as part of the equilibrium computation, for example, by minimizing the distance between restricted and unrestricted exit strategies.[9] To simplify the exposition

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

[9]In the numerical experiments we use the $1-$ norm for this minimization.

we ignore the η factors in the remaining of this section.

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

$$\mu'(x, \hat{s}) = \underset{\mu \in \mathcal{M}}{\operatorname{argmax}} \; V(x, \hat{s}|\mu, \tilde{\mu}, \lambda) \tag{9}$$

for all $x \in \mathcal{X}_f$ and $\hat{s} \in \hat{\mathcal{S}}$. Note that the optimal unrestricted strategy can be found by solving a single agent dynamic programming problem over the moment-based state space. We take the restricted strategy to be the average strategy:

$$\tilde{\mu}'(x, \hat{s}) = \tilde{\mu}(\hat{s}) = \frac{1}{\bar{x}} \int_0^{\bar{x}} \mu'(y, \hat{s}) \mathrm{d}y. \tag{10}$$

In addition, entry takes a simpler form than that described in Section 3: in state $\hat{s}$ exactly $\lambda(\hat{s})$ firms enter, where λ is determined in equilibrium and there is a fixed entry cost.

Now, assume that the single-period profit function depends on the first un-normalized moment. Under the previous restrictions, we show below that this moment becomes a sufficient statistic as the market size grows. We can also generalize this results in the following way: suppose that the profit function depends on the α-th moment and that firms stay in the industry with probability $(x/\bar{x})^q$ with $q = p/\alpha$. Then, under the same restrictions on strategies, the α-th moment becomes a sufficient statistic as the market size grows.

To formalize the result, we consider a sequence of industries indexed by the market size m and an associated sequence of MME strategies $(\tilde{\mu}^m, \lambda^m)$. We make the following assumption:

Assumption 6.1. *There exists a constant $C > 0$, such that for all sequences of MME strategies $\{(\tilde{\mu}^m, \lambda^m) : m \geq 0\}$ and for all t, we have $|F_t^m| \leq Cm$ almost surely, where F_t^m is the set of incumbent fringe firms for market size m.*

While the previous assumption imposes conditions on equilibrium outcomes, it is quite natural in this context; it states that the equilibrium number of fringe firms cannot grow larger than the market size. We have the following result.

Proposition 6.1. *Consider the model with restricted fringe firms' strategies described above. Suppose that Assumption 6.1 holds, and suppose that for every $m \geq 1$, all firms use MME strategies $(\tilde{\mu}^m, \lambda^m)$. For a given t, conditional on the realizations of $\{x_{it}|i \in F_t^m\}$ and $\hat{s}_t^m$, we have*

$$(1/m)\theta_{t+1}^m - (1/m)\left[\tilde{\zeta}^m(\hat{s}_t^m)\theta_t^m/\bar{x}^p \, + \lambda^m(\hat{s}_t^m)x^e\right] \to 0 \quad a.s.,$$

as $m \to \infty$, where $\theta_t^m = \sum_{i \in F_t^m} x_{it}$ and $\tilde{\zeta}^m(\hat{s}_t) = \mathrm{E}_{\tilde{\mu}^m}[\zeta_{it}(\tilde{\mu}^m(\hat{s}_t), w_{it})]$.

The result shows that the first moment becomes a sufficient statistic and suggests defining the following perceived transition kernel:

$$\hat{\mathbf{P}}_{\tilde{\mu}^m, \lambda^m}[\tilde{\zeta}^m(\hat{s}_t^m)\theta_t^m/\bar{x}^p + \lambda^m(\hat{s}_t^m)x^e \big| \hat{s}_t^m] = 1, \tag{11}$$

which will become a good approximation as m grows large. Similarly to the first approach, with this

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

perceived transition kernel and under suitable continuity conditions, one can show that the value of the full information deviation (appropriately normalized) converges to zero for dominant firms as the market size m approaches infinity. In this second approach, we cannot show this result for fringe firms, because they use restricted strategies in equilibrium. It is possible, however, to measure ex-post the degree of sub-optimality of the restricted fringe strategy. More formally, let $(\tilde{\mu}, \lambda)$ be the MME strategies. Let μ' be the unrestricted best response to $(\tilde{\mu}, \lambda)$ (see equation (9)). Then, using forward simulation, we can compare the actual expected discounted profits achieved by μ' relative to those achieved by $\tilde{\mu}$, that is, the difference $V^*(x, s|\mu', \tilde{\mu}, \lambda) - V^*(x, s|\tilde{\mu}, \lambda)$ over some set of selected industry states (e.g., the most visited states in the long-run under strategies $(\tilde{\mu}, \lambda)$).

6.1 Numerical Experiments

In this section, we illustrate that even the previous model, where fringe firms' strategies are restricted, already generates interesting strategic dynamic interactions between the dominant and fringe firms. We first solve for the MME 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 13.5 with individual states in $[0, 10]$ and 2 dominant firms with individual states in $\{6, 7, \ldots, 11\}$. Solving for MPE with this size of the state space is computationally intractable.

We discretize the state space for moments and the set of fringe firm states and solve for the MME. In short, in our algorithm, given strategies $(\tilde{\mu}, \lambda)$ the perceived transition kernel is derived. Given these perceived transitions, we compute firms' best responses ($\mathcal{C}1$) using value iteration and linear interpolation around grid points. We then use an inner loop to update the entry strategy until the zero profit condition ($\mathcal{C}2$) is satisfied. This loop alternates between modifying λ in the appropriate direction and recomputing the perceived value functions and transition kernel. To speed up the computation, we do not recompute incumbent firms' strategies during this subroutine. We iterate these steps until convergence.

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

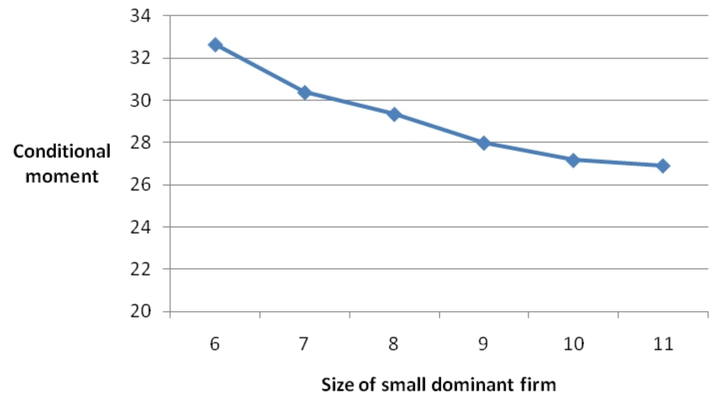


Figure 1: Conditional Moments (Big dominant firm at maximal state, 11)

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 $x^e = 1$. 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' spot 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. We compute the MPE of the 'normalized' EP model and simulate the industry. The results show that the average dominant firm state decreases to 6.25 from the MME value of 9.15. This suggests that deterring entry and pushing down the investment in the fringe tier are key determinants in dominant firms' investment incentives. Moreover, ignoring fringe firms may bias downwards the predicted 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 for fringe firms. In this section, we do not impose such restrictions. Instead we study MME assuming that firms, perhaps wrongly, suppose that moments form a Markov process that summarizes all payoff relevant information related to fringe firms. In this setting, we discuss how to extend our dynamic industry model to allow fringe firms to become dominant and vice versa, thus fully endogenizing the market structure in MME.

There are two main challenges that arise in this approach. First, we need to construct natural and meaningful candidates for the perceived transition kernel, $\hat{\mathbf{P}}$, that is, we need to choose a Φ operator. Such 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 in general; 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 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. In fact, if the value of the full information deviation is small, it is plausible that firms would use the relatively simpler MME strategies as oppose to more complex Markov strategies that do not yield significant additional benefits to them.

The reminder of this section is organized as follows: Subsection 7.1 describes two candidates 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 numerical experiments on a calibrated model of the beer industry in the United States over the second half of the previous century. Subsection 7.5 numerically compares MME with MPE strategies. 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 of Example 4.2. Note that this kernel does not specify transitions outside the recurrent class induced by firms' strategies. Transitions outside the recurrent class, however, may affect the equilibrium play in that firms' perception about these transitions may affect their equilibrium actions, and under these actions those states are indeed never reached. One 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. Alternatively, one can consider the limit of models with diminishing noise in transitions.

In addition, we briefly describe another construction of the perceived transition kernel that has been successfully used in stochastic 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 functions ξ_0 and ξ_1 that approximate the actual transitions best, for instance by employing linear regressions. In comparison to the empirical transitions kernel, this perceived transition kernel has the disadvantage of imposing strong parametric restrictions and assuming deterministic transitions. On the other hand, these same restrictions 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 firms 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 (one could also assume that this event occurs with some probability). In every period where $|D_t| < \bar{D}$, so that 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 firms 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.

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 tiers. Recall that $\mathbf{P}_{\mu',\mu,\lambda}[x', d', z'|x, \hat{s}]$ (see (3)) is the kernel that describes the actual evolution of (x_{it}, d_t, z_t), which firms can determine exactly from the primitive transition kernel when transitions between the fringe and dominant tiers are not possible. However, with such transitions moments may not contain sufficiently detailed information to determine exactly the transition probabilities between the fringe and dominant tiers. Therefore, in this case, we need to modify the kernel that describes the evolution of (x_{it}, d_t, z_t) to incorporate 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 (3) we implicitly assume that the evolution of a single fringe firm does not affect the evolution of the moments. Although in this more general setting a large fringe firm that 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 best response-type algorithm. The solver starts with a strategy profile, checks for the equilibrium conditions, and updates the strategies by best responding to the incumbent strategy until an equilibrium is found. In each iteration we construct the perceived transition kernel from firms' strategies according to the operator Φ. Depending on the operator used this could be done analytically or by simulation. We illustrate the general algorithm below for the empirical transitions kernel and comment on the generalization when appropriate. We use simulation to compute the empirical transition kernel $\hat{\mathbf{P}}_{\mu,\lambda}$ given strategies (μ, λ). A more detailed description is provided in Algorithm 1.[10] The following remarks

[10]For 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 that a fringe firm becomes dominant. These probabilities should coincide with the empirical transition probabilities observed under the equilibrium strategies. We also note that in the algorithm, $0 < \sigma < 1$ is chosen to speed-up convergence.

are important:

1. The algorithm terminates when the norm of the distance between a strategy and its best responseis 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, hence states are weighted according to their relevance. Moreover, this is useful because simulation errors in the estimated transition kernel are higher for states that are visited infrequently.

2. Even if in theory all states are recurrent, in a finite length simulation it is possible that some states will not be visited, and for them the perceived transition kernel cannot be computed. For those states we set the transitions in $\hat{\mathbf{P}}$ to be some predetermined value, for example, transition with probability one to the closest state (under some suitable norm) that has been actually visited. We tried different specifications for these states and found that they did not make much difference on MME outcomes.

Algorithm 1 Equilibrium solver

1: Initialize with (μ, λ) and industry state $s_0 = (f_0, d_0, z_0)$ with corresponding $\hat{s}_0$;
2: $n := 1, \delta := \epsilon + 1$;
3: **while** $\delta > \epsilon$ **do**
4: Simulate a T period sample path $\{(f_t, d_t, z_t)\}_{t=1}^T$ with corresponding $\{\hat{s}_t\}_{t=1}^T$ for large 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 $h(\hat{s}) > 0$

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

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

If the algorithm terminates with $\epsilon = 0$ (and if all states are recurrent) we have found an MME. A positive value of ϵ allows for numerical error. For different perceived transition kernels, lines 4-6 in Algorithm 1 will be different. For example in the parameterized kernel of Subsection 7.1, line 6 may include a regression for each (d, z) to find the parameters $\xi(d, z)$ that best fit the simulated transitions.

We found that for the empirical transitions kernel a real time stochastic algorithm similar to Pakes and McGuire (2001) and Fershtman and Pakes (2010) is much faster than Algorithm 1. In this variant the kernel is not simulated in every iteration, but instead continuation value functions that allow to solve for firms' optimal strategies are kept in memory and updated through simulation draws. In this sense, this modified real-time algorithm performs the simulation and optimization steps simultaneously. We use one step of

Algorithm 1 to check whether $\mathcal{C}1$ and $\mathcal{C}2$ hold and convergence has been achieved. The details of this variant of the algorithm are presented in Appendix D.1.

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 of the beer industry to examine the role advertising may have on market structure. The model is similar to the dynamic advertising model by Doraszelski and Markovich (2007). We emphasize that it is not our goal to develop a complete and exhaustive empirical model of the beer industry, but rather to illustrate our approach in a setting of potential empirical relevance.

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

$$\sigma(x_{it}, s_t) \propto (x_{it})^{\alpha_1}(Y - p_{it})^{\alpha_2}.$$

Firms invest in advertising to increase their goodwill stock over time and compete in prices every period. The number of dominant firms is determined endogenously in the equilibrium 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 as captured by the parameter α_1. Firms keep track of the α_1-th unnormalized moment and MME is computed for each parameter specification.

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 $(\alpha_D, \alpha_C, \alpha_I) = (.85,\ 1,\ 1.1)$. 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). Firms become dominant when their individual goodwill level becomes larger than a predetermined value.

We calibrate the model parameters from a variety of empirical research that studies the beer industry

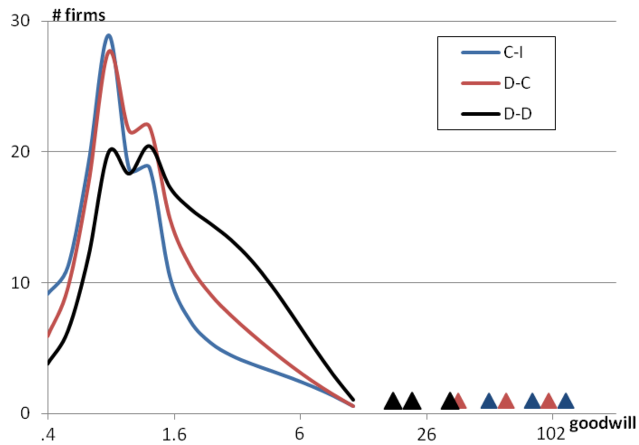


Figure 2: Average size distribution (un-normalized) of firms in log-scale. The solid line represents fringe firms and the triangles represent dominant firms.

or related advertising settings.[11] 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, and Table 2 reports some average industry statistics. The experiments suggest that higher returns to advertising indeed give rise to more right-skewed size distributions, as it is expected. Indeed, it is clear that in the DRA-DRA case there is on average 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. Also, dominant firms remain dominant for a much larger period of time. In addition, as the industry becomes more concentrated, there are fewer incumbent fringe firms, they are smaller, and spend less time in the industry on average. In that sense, large dominant firms deter the growth of fringe 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 these numerical experiments highlight the usefulness of our approach. In particular, we envision that it may be specially useful to perform counterfactuals when the market structure is endogenously determined.

7.5 Comparison with Markov Perfect Equilibrium

In the previous section, we presented numerical experiments in which MPE is infeasible to compute. In this subsection we present a set of experiments to compare our equilibrium concept with MPE in small industries in which we can compute MPE. In particular, we consider industries with a total of 6 firms (dominant and

[11]We thank Victor Tremblay and Carol Termblay for providing supplementary data.

Industry averages	CRA-IRA	DRA-CRA	DRA-DRA
Concentration ratio C_2	0.42	0.36	0.13
Concentration ratio C_3	0.53	0.44	0.17
First moment (normalized by # fringe)	1.085	1.22	1.74
Active fringe firms (#)	147	167	180
Active dominant firms (#)	3	3	2.3
Size incumbent 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

Table 2: Average industry statistics

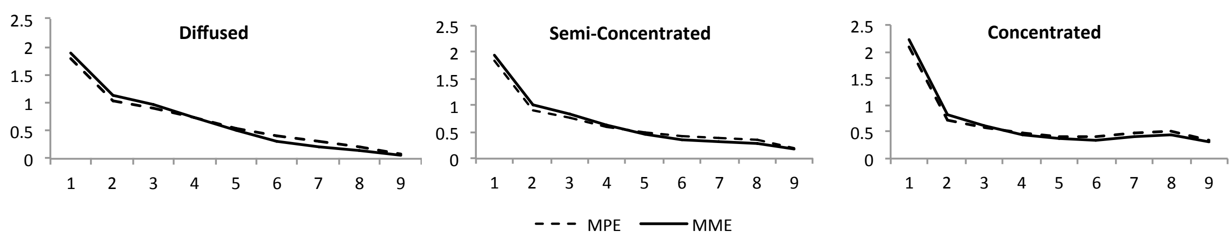


Figure 3: Average size distribution of firms under MME and MPE.

fringe) and 9 individual states. In spite of the small number of firms, we find that the two equilibria produce very similar market outcomes as will be shown below.

We consider the same model specification as the beer industry experiments, expect for the changes outlined next. We scale down the market size due to smaller number of firms and rule out entry and exit. We again assume that bigger firms are able to achieve higher returns for advertising, as captured by the parameter α_1. Here the returns to advertising of firms in states 1 to 5 and of firms in states 6 to 9 are captured by parameters α_1^{low} and α_1^{high}, respectively. We take $\alpha_1^{\text{low}} = 1$ and consider three cases with $\alpha_1^{\text{high}} \in \{1.5, 2, 2.5\}$ that we call diffused, semi-concentrated, and concentrated, respectively, as we expect higher returns to advertising to result in more concentration.

We first solve for MPE for each of the three cases. Then we take individual state 6 to be the one separating between fringe and dominant firms, and solve for MME with a single moment. That is, firms with states 5 or lower are fringe, and the rest are dominant. The number of active dominant firms is endogenous (see Subsection 7.2) with a maximum of 4 dominant firms.

Figure 3 shows the long-run average size distribution of firms in both equilibria for the three cases. Notably, the average market structure is almost identical under both equilibria concepts for the three cases. A closer inspection shows that, on average, firms spend slightly more time in higher states in MPE than in MME. This is expected, since small (fringe) firms ignore their market power in MME, but not in MPE. Therefore, small firms have stronger incentives to grow in MPE, resulting in a more right skewed distribu-

	Diffused	Semi-concentrated	Concentrated
$\mathrm{CS}_{\mathrm{MPE}}/\mathrm{CS}_{\mathrm{MME}}$	6.9%	7.4%	6.5%
$\mathrm{PS}_{\mathrm{MPE}}/\mathrm{PS}_{\mathrm{MME}}$	5.5%	5.7%	4.8%

Table 3: Comparison of consumer and producer surplus under MME and MPE.

tion. This incentive should decrease with the number of firms in the market. Thus it is reasonable to expect that the gap between the size distributions will be even smaller for industries with many firms. In addition, Table 3 compares both the consumer and producer surpluses (CS and PS, respectively). The difference between those quantities under both equilibria is moderate and ranges from 4.8% to 7.4%.

These experiments suggest that MME and MPE can give rise to similar market outcomes, even in industries with few firms and when moments are not neccesarily sufficient statistics. Finally, we would like to comment that even with 6 firms, the MME computation was about 10 times faster than MPE. Moreover, with a fixed number of dominant firms, as the number of fringe firms scales, the running time for the MPE solver would scale exponentially, whereas the running time for the MME solver would remain constant.

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, that is, a unilateral deviation to a strategy that keeps track of the underlying state of the industry s. In theory, one could compute the value of the unilateral deviation exactly, however this is almost as computationally challenging as solving for the MPE in the underlying state space. As such, we suggest a novel computationally tractable error bound that upper bounds the value of the full information deviation. This error bound is derived by observing a connection between the unilateral deviation problem of a firm in MME and problems in robust dynamic programming (RDP); see Iyengar (2005) for a reference on this literature.[12] We note that to the best of our knowledge ideas from RDP have not been previously used to derive sub-optimality error bounds for dynamic programs with partial information as we do here.

Let (μ, λ) be a fixed MME strategy for the reminder of this section, and assume that firms' profit function $\pi(x, \hat{s})$ and cost function $c(x, \iota)$ are uniformly bounded for all states $(x, \hat{s})$ and investment levels $\iota \in \mathcal{I}$. Recall that $V^*(x, s|\mu, \lambda)$ is the actual value of playing MME strategies starting from (x, s). Denote by $\overline{V}^*(x, s|\mu, \lambda) = \sup_{\mu' \in \mathcal{M}^*} V^*(x, s|\mu', \mu, \lambda)$. The value of the full informational deviation is $\Delta_{\mu,\lambda}(x, s) = \overline{V}^*(x, s|\mu, \lambda) - V^*(x, s|\mu, \lambda)$ (see (6)). Note that given MME strategies (μ, λ), $V^*(x, s|\mu, \lambda)$ can be easily computed using forward simulation. However, the problem of finding the optimal strategy that achieves $\overline{V}^*$ is subject to the curse of dimensionality. Instead we find an upper bound to $\overline{V}^*$ with which we can upper bound $\Delta_{\mu,\lambda}$. To do so we construct a *robust Bellman operator* as follows. For every $\hat{s} = (\theta, d, z) \in \hat{\mathcal{S}}$

[12]RDP considers Markov decision processes with unknown transition kernels in which the decision maker chooses a 'robust' strategy to mitigate this ambiguity. In our case, observing the moment, but not the underlying industry state, is equivalent to having ambiguity about the underlying transition probabilities. Therefore, the unilateral deviation problem of an agent in MME can be reformulated as a RDP problem.

define the *consistency set* $\mathcal{S}_f(\hat{s}) = \{f \in \mathcal{S}_f \big| \theta(f) = \theta\}$, that is the set of all fringe firm distributions that are consistent with the value of the moments in state $\hat{s}$. Note that 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 distributions in the consistency set may have different future evolutions. Now, define

$$(T_R\hat{V})(x,\hat{s}) = \sup_{\substack{\iota\in\mathcal{I}\\ \rho\geq 0}} \sup_{f\in\mathcal{S}_f(\hat{s})} \Big\{\pi(x,\hat{s}) + \mathrm{E}\Big[\phi\mathbf{1}\{\phi\geq\rho\} + \mathbf{1}\{\phi<\rho\}\big[-c(x,\iota) + \beta\,\mathrm{E}_{\mu,\lambda}[\hat{V}(x_{i,t+1},\hat{s}_{t+1})\big|x_{it}=x, s_t=(f,d,z),\iota]\big]\Big]\Big\}, \quad (12)$$

where $\hat{V} \in \hat{\mathcal{V}}$ is a bounded vector $\hat{V} : \mathcal{X}\times\hat{\mathcal{S}} \to \Re$ and $\mathbf{1}\{\cdot\}$ is the indicator function. 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 with the best response in $\mathcal{C}1$ in the definition of MME, except that the firm can also choose any underlying fringe firm state consistent with the moment θ.

In Lemma A.1 in the Appendix, we show that the robust Bellman operator T_R has a unique fixed point $\hat{V}^*$ that we call the *robust value function*. The next result relates it to the optimal value function $\overline{V}^*$.

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

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

where $\hat{s}$ is the moment based industry state that is consistent with s, that is, $\theta = \theta(f)$

In essence, the robust value function resolves the indeterminacy of moments transitions by choosing the best fringe firm state from the associated consistency set. Intuitively, this should provide an upper bound for $\overline{V}^*$. 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 moment-based industry state of s. The advantage of computing $\hat{V}^*$ over $\overline{V}^*$ is that T_R operates on the moment based state space $\hat{\mathcal{S}}$ which is much smaller than $\mathcal{S}$. Nevertheless, the computation of $\hat{V}^*$ is still demanding as we explain next.

Finding the fixed point of the operator T_R is generally a hard computational problem, in fact it is NP-complete (Iyengar, 2005, §3). This is not surprising, since the optimization over consistency sets may be very complex. However, iterating the operator T_R becomes much simpler if there is a large number of fringe firms. In this case, the one-step transition of the fringe firm state is close to being deterministic, because, conditional on the current state, the transitions of individual fringe firms average out at the aggregate level by a law of large numbers. When one-step transitions are deterministic, finding the optimal consistent f in (12) is equivalent to choosing the next moment from a set of moments that are accessible from the current industry state. This considerably simplifies the computation of the inner maximization in (12), since the moment accessibility sets are low dimensional. Moreover, characterizating these accessibility sets can be done efficiently. We provide details of this procedure in Appendix D.2. With this modification in place, loosely speaking, computing the bound is tractable for problems for which computing MME is tractable.[13]

[13] In applications, the number of fringe firms N is finite, hence, the robust bound based on deterministic transitions for the fringe

It is simple to observe that if moments are sufficient statistics to predict the future evolution of the industry, then the bound is tight. If this is not the case, the gap between the actual value of the full information and the quantities computed with the robust bound, $\hat{\Delta}_{\mu,\lambda}(x,s)$ depend on the distance $\hat{V}^*(x,\hat{s}) - \overline{V}^*(x,s)$. To lower this gap and make the bound tighter we can compute the robust bound over a refinement of the moment-based state space used in the equilibrium computation by including moments in the robust operator not included in MME. For example, we could add another contemporaneous moment like a quantile, or add a lagged moment. This will reduce the size of the consistency sets making the bound tighter, albeit increasing the computational cost. We illustrate this with numerical experiments in the next subsection.

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 ($|D_t| = \overline{D}$), the bound does not change. In other industry states, the deviating firm can pick a fringe firm state that does not allow any fringe firm to transition to a state from which it can become dominant, that is, a distribution f with $f(x) = 0$ for all $x \in \mathcal{K}_f$. In this case the optimization over consistency sets yields the deviating firm more influence on the evolution of the industry that results in a looser bound. This can be mitigated, however, by augmenting the state space, for example by keeping track of the number of firms in $\mathcal{K}_f$.

8.1 Numerical Experiments

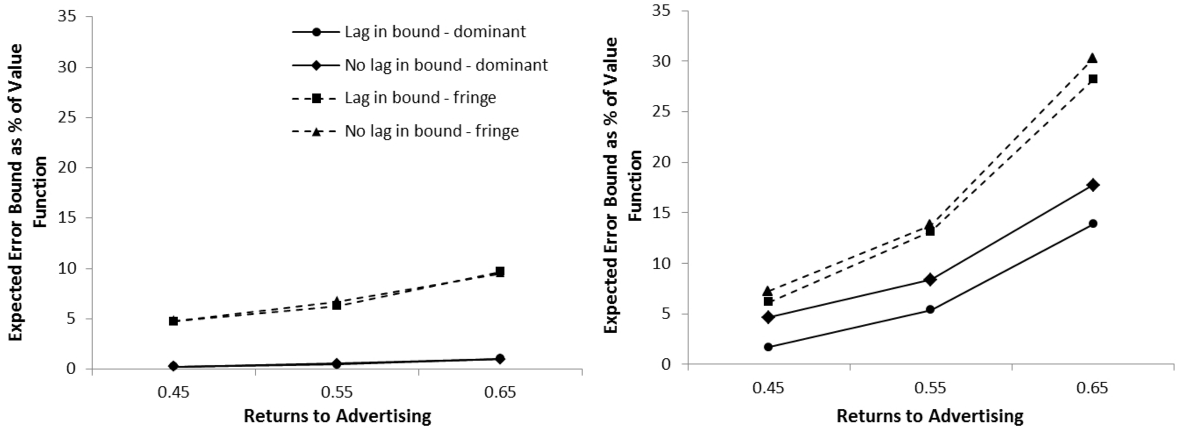
We have done extensive numerical experiments using the robust bound. First, in one set of experiments we computed the robust bound for several instances of models in which fringe firms are constrained in their strategies as in Section 6. Here, when the number of fringe firms is large, the value of the 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.

We also experimented in models of dynamic oligopoly competition when moments are not sufficient statistics corresponding to the third approach. In particular, we consider two simplified models similar to the beer industry experiments of Subsection 7.4 with a single dominant firm (D is a singleton) that operates under constant returns to advertising. In the first model $N = 200$ and fringe firms can enter and exit the market (model E), in the second $N = 100$ and firms cannot enter or exit the industry (model NE). For each model we vary α_1^f, the parameter controlling fringe firms' returns to advertising, from .45 to .65. The lower the returns to advertising, the smaller the reward from investment and the more homogenous the fringe firms are in equilibrium. Since moments are sufficient statistics when fringe firms are homogenous, we expect the error bound to be lower for low values of α_1^f.

In both models E and NE we use the parametric approach of Subsection 7.1 to construct the perceived transition kernel. In particular, a linear relationship

$$\theta_{t+1} = \xi_0(d_t) + \xi_1(d_t)\theta_t$$

firm state is only an approximation to an upper bound. It is possible to formally derive a probabilistic version of the robust bound that corrects for this using standard probability bounds. Loosely speaking, by a central limit theorem the correction term is of the order $\sqrt{N}$. However, given our numerical experience, we believe that in many settings of interest in which N is large, the square root N correction will not have a significant impact for practical purposes.



(a) No Entry/Exit

(b) Entry/Exit

Figure 4: Error Bound

is assumed, and the parameters $\{(\xi_0(d), \xi_1(d))\}_{d \in \mathcal{S}_d}$ are taken to fit the simulated transitions best using ordinary least squares. To simplify, we assume there is no aggregate shock. We report in Figures 4a and 4b the expected error bound for both dominant and fringe firms as a percentage of the actual value function. Namely, we plot

$$\mathrm{E}_{\mu,\lambda}\Big[\frac{\hat{\Delta}_{\mu,\lambda}(x,s)}{V^*(x,s|\mu,\lambda)}\Big],$$

where the expectation is taken with respect to the invariant distribution of the underlying industry state s.[14] In addition to the standard robust bound, we consider a variation where a lagged fringe firm state enters the computation of the consistency sets. This reduces the size of these sets and so it decreases the gap between the actual value of the full information and the error bound. Indeed we see that the lagged state decreases the error bound, in particular in model E. Adding further contemporaneous or lagged moments should generally make the error bound even tighter.

The experiments agree with our conjecture that a more homogenous fringe tier (low α_1^f) will result in a lower error bound. In addition, we see that the error bounds are generally lower for model NE. This is explained by the better fit of the linear moment transitions without entry and exit. Entry and exit decisions are inherently nonlinear. As such, the assumed linear moment transitions approximate the actual moment transitions in NE better than in E resulting in a lower value of the full information deviation.

These experiments suggest that the robust bound can be useful to test the extent of sub-optimality of MME strategies in terms of a unilateral deviation. Based on our numerical results, we found that depending on the model, the bound can be sometimes tight indicating that the extent of a unilateral deviation is small, but it can also be loose. In the latter cases, refining the state space by adding additional moments when computing the robust bound can be helpful to make the bound tighter and more useful. Also, adding additional moments to firms' MME strategies should generally improve the accuracy of the perceived transition kernel

[14]For dominant firms, the expectation is also taken with respect to the invariant distribution of their individual state evolution. For fringe firms, x is taken to be the most visited fringe state.

and it is plausible to expect that this would decrease the bound as well. There may still be settings in which even after adding several moments, the bound will be loose. This is not entirely surprising as the error bound derived in this paper is very generic and does not use any problem specific information. We leave potential refinements of our bound that use problem specific information as matter for further research.

9 Extensions

Our approach is, of course, related to the important literature in macroeconomics that studies models with heterogeneous agents in the presence of aggregate shocks. Notably, Krusell and Smith (1998) studies a stochastic growth model with heterogeneity in consumer income and wealth. Other papers focus on firm level heterogeneity in capital and productivity (see, for example, Khan and Thomas (2008) and Clementi and Palazzo (2010)). All of these models assume a continuum of agents and therefore agents' dynamic programming problems are infinite dimensional; strategies depend on the distribution of individual states. Motivated by the seminal idea in Krusell and Smith (1998), to overcome the curse of dimensionality, in these papers agents are assumed to keep track of only few statistics of this distribution. Note that while these macroeconomic models do not have dominant agents, our framework can easily accommodate them, as an aggregate shock can be understood as an exogenous dominant firm.

While our approach is inspired by this previous literature, we believe that in turn our results can also be useful in these macroeconomic models as we now explain. We start describing the connections to the literature that focuses on firm level heterogeneity, because it is more directly related to our model, and then discuss Krusell and Smith-style models.

9.1 Dynamic Industry Models

Dynamic industry models with a continuum of firms have been applied to several settings in macroeconomics such as business cycles and international trade policy. These models often extend the pioneering model of Hopenhayn (1992) to include an aggregate shock. Because these models assume an infinite number of firms we refer to them as *infinite* models.

In partial equilibrium, infinite models are often similar to the model we introduced in Section 3, with minor modifications that we now discuss.[15] An infinite model represents an asymptotic regime where the number of firms and the market size become infinite. 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. There is an infinite mass of potential entrants. The rest of the model primitives are the same as in the (finite) model of Section 3. 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.[16]

[15] Weintraub et al. (2011) provide more details regarding the connection between (finite) dynamic oligopoly models and infinite models.

[16] Note that in the presence of an aggregate shock, the commonly used concept of *stationary equilibrium (SE)*, in which the industry state is assumed to be constant over time due to averaging effects, is no longer a reasonable equilibrium concept, because the industry state fluctuates with the aggregate shock.

Our results are connected to infinite models in several levels. Because in an infinite model a law of large numbers is assumed to hold exactly to predict the one period evolution of the industry state, the perceived transition kernel defined in equations (7) and (11) for the first and second approaches, respectively, will coincide exactly 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* (recall that the latter is only true for dominant firms in the second approach). Moreover, the robust bound derived in the third approach based on a law of large numbers will be a valid upper bound without further corrections. We explore in more detail the usefulness of the bound in the context of stochastic growth models next.

9.2 Stochastic Growth Models

The seminal paper of Krusell and Smith (1998) studies an infinite-horizon, stochastic general equilibrium model with aggregate and individual level shocks, and incomplete insurance (see Krusell and Smith (2006) for a recent survey of this literature). In the model, forward looking agents seek to maximize the stream of expected discounted utility by making consumption/saving decisions. To make these decisions, consumers need to forecast the prices in the market, such as interest rates and wages. In equilibrium, those in turn depend on both aggregate shocks and the distribution of wealth across agents.

Therefore, in principle, to make decisions agents need to keep track of the distribution of wealth, which is highly dimensional. However, Krusell and Smith (1998) show that in this model the first moment of this distribution is essentially a sufficient statistic for the evolution of the economy. The reason is that agents' equilibrium decisions turn out to be almost linear in their state. They call this property "approximate aggregation". In fact, if decisions were exactly linear, the first moment would be an exact sufficient statistic, as in our first approach.

While the approximate aggregation property has shown to be quite robust in this class of models, Krusell and Smith (2006) and Algan et al. (2010) acknowledge that it does not hold for all models and may not hold for important models considered in the future. For this reason, it is important to understand the boundaries of approximate aggregation by testing its accuracy. Den Haan (2010) describes the limitations of commonly used accuracy tests, such as the so called "R^2 test" and provide other alternatives. These tests often try to measure the accuracy of the perceived transition kernel proposed. We believe the bound from our third approach provides a more direct test of the validity of a state aggregation technique, because it directly measures improvements in payoffs. Note that the accuracy of the perceived transition kernel is embedded in our bound, as poor approximations of the kernel will lead to large bounds.

More specifically, we have introduced a computationally tractable algorithm that provides an upper bound on how much an agent can improve its expected discounted utility (in monetary terms) by unilaterally deviating from the approximate equilibrium strategy to a strategy that keeps track of the full distribution of wealth.[17] In general, the bound provides a direct test on whether the state aggregation is appropriate or whether a finer state aggregation is necessary, for example by adding more moments.[18] In fact, numerical

[17] Specifically, there exist a nonnegative monetary quantity (that depends on the individual and economy states), such that a consumer will be indifferent between receiving this quantity and playing the equilibrium strategy (in wealth moments), or not receiving it and playing the best response that may depend of the wealth distribution. We upper bound this quantity.

[18] Unlike our dynamic oligopoly model, in this stochastic growth model agents' individual transitions (whether they are employed

experiments we have conducted show that our bound is fairly tight and that, as expected, deviating from approximate aggregation to a strategy that keeps track of all available information provides little gains to agents in the original model in Krusell and Smith (1998).[19]

The previous results are encouraging and suggest that our bound and potential extensions can help exploring the boundaries of approximate aggregation. In particular, our bound can be an aid to decide which statistic one should add next to the state space if approximate aggregation fails. Researchers have often tried the second moment of the distribution of wealth, but our preliminary results suggests that in some cases other statistics such as quantiles can provide more valuable information for agents.

10 Conclusion

In this paper we introduced a new framework to study dynamic oligopoly in concentrated industries that opens the door to study new issues in the empirical analysis of industry dynamics. Our first two approaches in Sections 5 and 6 provide models that impose no restrictions over dominant firms primitives and for which MME strategies for dominant firms become optimal. They do impose, however, restrictions over fringe firms. We believe these models may be particularly relevant for applications in which dominant firms are they key focus of analysis and a detailed model of the fringe is not required. In addition, we believe that our approach in Section 7, in which firms approximate the non-Markov process of moments by a Markov process and the error in the approximation is measured with our error bound can prove particularly useful in other empirical applications, because it imposes no restrictions on fringe firms.

We believe that our work suggests several future directions for research. First, our error bound is very general and we envision that tighter bounds can be derived using problem specific information. Second, in our definition of MME firms keep track of current moments of the fringe firm state. Our equilibrium concept and error bound can be modified to allow firms to keep track of past values of the moments. Third, one can relax the assumption that single-period profits are a function of just few fringe moments. In this case, firms would need to have perceptions of the expected profits given the moments they observe, for example, by using the empirical profits received at each state. Finally, our definition of MME and numerical experiments in Section 7 intend to study the long-run equilibrium market structure. We believe that an appropriate modification of our approach would allow to study the short-transitional dynamics and how would the market structure evolve in few years after a policy or an environmental change, such as a merger.

In Section 9 we discussed how our methods can be extended to other important models in economics. In addition, we envision that some of our ideas could potentially also be useful to study dynamic models with forward looking consumers. For example, in a dynamic oligopoly model with durable goods, firms may need to make pricing and investment decisions, keeping track of the distribution of consumers' ownership, which is highly dimensional (e.g., Goettler and Gordon (2011)). Our ideas may be useful in this context as well, where one could replace this distribution by some of its moments. We leave all these extensions for future research.

or unemployed) are correlated through the aggregate shock even conditional on the current market state. Thus, the computation of the bound has to be slightly modified to let moment accessibility sets depend on the transition of the aggregate shock as well.

[19]These results can be obtained from the authors upon request.

References

Algan, Y., O. Allais, W. J. Den Haan, P. Rendahl. 2010. Solving and simulating models with heterogeneous agents and aggregate uncertainty. Working Paper.

Altug, S., P. Labadie. 1994. *Dynamic Choice and Asset Markets*. Academic Press.

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, D. P., S. 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.

Blundell, R., T.M. Stoker. 2005. Heterogeneity and aggregation. *Journal of Economic Literature* **XLIII** 347 – 391.

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'Erasmo. 2011. A quantitative model of banking industry dynamics. Working paper, UT Austin.

Corbae, D., P. D'Erasmo. 2012. Capital requirements in a quantitative model of banking industry dynamics. Working Paper, Maryland University.

Den Haan, W. J. 2010. Assessing the accuracy of the aggregate law of motion in models with heterogeneous agents. *Journal of Economic Dynamics and Control* **34** 79 – 99.

Dixit, A. K., J. E. Stiglitz. 1977. Monopolistic competition and optimum product diversity. *American Economic Review* **67**(3) 297 – 308.

Donoghue, W. F. 1969. *Distributions and Fourier Transforms*. Academic Press.

Doraszelski, U., K. Judd. 2011. Avoiding the curse of dimensionality in dynamic stochastic games. *Quantitative Economics* **3**(1) 53–93.

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.

Doraszelski, U., M. Satterthwaite. 2010. Computable markov-perfect industry dynamics. *RAND Journal of Economics* **41** 215–243.

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 relashionship between frim 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. 2012. An approximate dynamic programming approach to solving dynamic oligopoly models. *The RAND Journal of Economics* **43**(2) 253–282.

Fershtman, C., A. Pakes. 2010. Oligopolistic dynamics with asymmetric information: A framework for empirical work. Working Paper, Harvard University.

Gallant, A. R., H. Hong, A. Khwaja. 2010. Dynamic entry with cross product spillovers: An application to the generic drug industry. Working paper.

Goettler, R. L., B. Gordon. 2011. Does AMD spur Intel to innovate more? *Journal of Political Economy* **119**(6) 1141–1200.

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.

Kalouptsidi, Myrto. 2011. Time to build and fluctuations in bulk shipping. Working Paper.

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.

Krusell, P., A. A. Smith, Jr. 2006. Quantitative macroeconomic models for heterogeneous agents. Working Paper.

Lee, R. S. 2010. Dynamic demand estimation in platform and two-sided markets. Working Paper, NYU.

Lucas, R. E., E. C. Prescott. 1971. Investment under uncertainty. *Econometrica* **39**(5) 659 – 681.

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.

Resnick, S.I. 1998. *A Probability Path*. Birhauser.

Roberts, M. J., L. Samuelson. 1988. An empirical analysis of dynamic, nonprice competition in an oligopolistic industry. *The RAND Journal of Economics* **19**(2) pp. 200–220.

Rojas, C. 2008. Price competition in u.s. brewing. *The Journal of Industrial Economics* **56**(1) 1–31.

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.

Santos, C. D. 2012. An aggregation method to solve dynamic games. *Working paper* .

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, A. 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, V. J., N. Iwasaki, C. H. Tremblay. 2005. The dynamics of industry concentration for u.s. micro and macro brewers. *Review of Industrial Organization* **26** 307–324.

Tremblay, V. J., C. H. Tremblay. 2005. *The US Brewing Industry: Data and Economic Analysis*, *MIT Press Books*, vol. 1. The MIT Press.

Weintraub, G. Y., C. L. Benkard, B. Van Roy. 2008. Markov perfect industry dynamics with many firms. *Econometrica* **76**(6) 1375–1411.

Weintraub, G. Y., C. L. Benkard, B. Van Roy. 2010. Computational methods for oblivious equilibrium. *Operations Research (Special Issue in Computational Economics)* **58**(4) 1247–1265.

Weintraub, G. Y., C. L. Benkard, B. Van Roy. 2011. Industry dynamics: Foundations for models with an infinite number of firms. *Journal of Economic Theory* **146** 1965–1994.

Xu, Y. 2008. A structural empirical model of R&D, firm heterogeneity, and industry evolution. Working paper, NYU University.

Appendices

A Proofs

Proof of Lemma 5.1. Under the assumptions of model (N) of Chapter 9 in Bertsekas and Shreve (1978) we have from Proposition 9.8, that the optimal value function satisfies:

$$(TV)(x,\hat{s}|\mu) := \max_{\iota\in\mathcal{I}}\left\{x\pi_1(\hat{s}_0)+\pi_0(\hat{s}_0)-cx\iota+\beta\,\mathrm{E}_{\mu}\left[V(x_1,\hat{s}_1)\Big|\iota,\hat{s}_0=\hat{s},x_0=x\right]\right\}=V(x,\hat{s}|\mu)$$

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

$$\sup_{\mu'\in\mathcal{M}} V(x,\hat{s}|\mu',\mu)=xV^1(\hat{s})+V^0(\hat{s})$$

for appropriate functions $V^1(\cdot)$ and $V^0(\cdot)$ by demonstrating that the posited form of the perceived value function is stable under an application of the Bellman operator. We have:

$$\begin{aligned}(T\hat{V})(x,\hat{s}|\mu) &= \max_{\iota\in\mathcal{I}}\Bigg\{x\pi_1(\hat{s})+\pi_0(\hat{s})-cx\iota+\beta\,\mathrm{E}_{\mu}\Big[(x\zeta_1(\iota,\hat{s},w_1)+\zeta_0(\hat{s},w_0))\hat{V}_1(\hat{s}_{t+1})\\ &\quad+\hat{V}_0(\hat{s}_{t+1})\Big|\iota,\hat{s}_t=\hat{s}\Big]\Bigg\}\\ &= x\max_{\iota\in\mathcal{I}}\left\{-c\iota+\beta\,\mathrm{E}_{\mu}\left[\zeta_1(\iota,\hat{s},w_1)\hat{V}_1(\hat{s}_{t+1})\Big|\iota,\hat{s}_t=\hat{s}\right]\right\}+x\pi_1(\hat{s})+\tilde{V}_0(\hat{s}) \qquad (13)\\ &= x\tilde{V}_1(\hat{s})+\tilde{V}_0(\hat{s}),\end{aligned}$$

where we define $\tilde{V}_0(\hat{s})=\pi_0(\hat{s})+\beta\,\mathrm{E}_{\mu}\left[\zeta_0(\hat{s},w_0)\hat{V}_1(\hat{s}_{t+1})+\hat{V}_0(\hat{s}_{t+1})\Big|\hat{s}_t=\hat{s}\right]$. Now, let us denote by $\hat{V}^n$ the iterates obtained by applying the Bellman operator T. Then, we have concluded that

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

as $n\to\infty$, where V is the optimal value function. But since the above holds for at least two distinct values of x for any given $\hat{s}$, this suffices to conclude that $\hat{V}_1^n(\hat{s})\to\hat{V}_1^\infty(\hat{s})$ and $\hat{V}_0^n(\hat{s})\to\hat{V}_0^\infty(\hat{s})$.

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 equation (13) and Proposition 4.3 of the reference. □

Proof of Proposition 5.1. Note that

$$\theta_{t+1}^m=\sum_{i=1}^{N^m}x_{i,t+1}=\sum_{i=1}^{N^m}[x_{it}\zeta_1(\mu^m(\hat{s}_t^m),\hat{s}_t^m,w_{1it})+\zeta_0(\hat{s}_t^m,w_{0it})],$$

where (w_{0it}, w_{1it}) are i.i.d. We need to show that conditional on s_t^m,

$$(1/N^m)\left|\theta_{t+1}^m - \mathrm{E}_{\mu^m}[\theta_{t+1}^m|s_t^m]\right| \to 0,\ a.s.,$$

where $\mathrm{E}_{\mu^m}[\theta_{t+1}|s_t^m] = \tilde{\zeta}_1^m(\hat{s}_t^m)\theta_t^m + N^m\tilde{\zeta}_0(\hat{s}_t^m)$. The result follows by Assumption 5.1 and a law of large numbers (see, for example, Corollary 7.4.1 in Resnick (1998)). □

Proof of Proposition 5.2. If $\alpha \neq 1$ redefine the state of fringe firms to be $y = x^\alpha$. Without loss of generality assume $\mathcal{X}_f = [0, \bar{x}]$. Let s_t, s_t' be two consistent underlying industry states, that is, $d_t = d_t'$, $z_t = z_t'$, and $\theta(f_t) = \theta(f_t') = \theta_t$. For moments to be sufficient statistics it must be that for any such consistent underlying states the expected next moment are the same $\mathrm{E}_\mu[\theta(f_{t+1})|s_t] = \mathrm{E}_\mu[\theta(f_{t+1})|s_t']$.

Let us fix a moment-based industry state $\hat{s}$ and let us define the function $g(x;\hat{s}) = \mathrm{E}_\mu[x_{i,t+1}|x_{it} = x, \hat{s}_t = s]$. A function h is midpoint convex if for all $x_1, x_2 \in (\underline{x}, \bar{x}) \subset \Re$ the following holds $h((x_1 + x_2)/2) \leq (h(x_1) + h(x_2))/2$. Midpoint concavity is defined by reversing the inequality. A function that is midpoint convex (concave) and bounded (and so Lebesgue measurable) is convex (concave) (see Donoghue (1969)). We will show that $g(\cdot;\hat{s})$ is both convex and concave and so linear by showing that midpoint convexity and midpoint concavity hold. Note that we need to show this only for $x \leq \theta$ as a single fringe firm cannot be larger than the moment.

For any x_0 in the interior of $\mathcal{X}_f$ with $2x_0 \leq \theta_t$ we can find (we assumed $N - \overline{D} \geq 3$) a fringe state f with $f(x_0) \geq 2$ and $\delta > 0$ such that $x_0 \pm \delta \in \mathcal{X}_f$. Construct f' with $f'(x_0) = f(x_0) - 2$, $f'(x_0 - \delta) = f(x_0 - \delta) + 1$, $f'(x_0 + \delta) = f(x_0 + \delta) + 1$ and $f'(x) = f(x)$ for all $x \notin \{x_0, x_0 \pm \delta\}$. Let $s = (f, d, z)$ and $s' = (f', d, z)$. Clearly $\hat{s} = \hat{s}'$, because $\theta(f) = \theta(f')$. By assumption $\mathrm{E}_\mu[\theta(f_{t+1})|s_t] - \mathrm{E}_\mu[\theta(f_{t+1})|s_t'] = 2g(x_0;\hat{s}) - g(x_0 - \delta;\hat{s}) - g(x_0 + \delta;\hat{s}) = 0$ or $g(x_0;\hat{s}) = (g(x_0 - \delta;\hat{s}) + g(x_0 + \delta;\hat{s}))/2$. That is $g(x;\hat{s})$ is midpoint convex and midpoint concave in the variable x, and so linear.

For $x_0 \geq \theta/2$ we use the following construction. From the result above we have that $g(\theta_t/2 - \delta;\hat{s})$ is linear for $0 \leq \delta \leq \theta_t/2$. By assumption, $g(\theta_t/2 + \delta;\hat{s}_t) + g(\theta_t/2 - \delta;\hat{s}_t) = 2g(\theta_t/2;\hat{s}_t)$, or $g(\theta_t/2 + \delta;\hat{s}_t) = 2g(\theta_t/2;\hat{s}_t) - g(\theta_t/2 - \delta;\hat{s}_t)$. Substituting for the right hand side we get $g(\theta_t/2 + \delta;\hat{s}_t) = (\theta_t/2 + \delta)\tilde{\zeta}_1(\hat{s}_t|\mu) + \tilde{\zeta}_0(\hat{s}_t|\mu)$ for appropriately defined $\tilde{\zeta}_0, \tilde{\zeta}_1$ and for all $0 \leq \delta \leq \min(\bar{x} - \theta_t/2, \theta_t/2)$. This completes the proof.

□

Proof of Proposition 6.1. Recall that $0 \leq x_{it}^m \leq \bar{x} < \infty$. For $i \in F_t^m$ note that

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

Now,

$$\mathrm{E}_{\mu^m,\lambda^m}[\theta_{t+1}^m|s_t^m] = \sum_{i\in F_t^m}\tilde{\zeta}^m(\hat{s}_t^m)x_{it}^m/(\bar{x})^p + \lambda^m(\hat{s}_t^m)x^e,$$

and so the result follows via the proof of Proposition 5.1. □

Lemma A.1. *The operator T_R satisfies the following properties:*

1. *T_R is a contraction mapping modulo β. That is, for $\hat{V}, \hat{V}' \in \hat{\mathcal{V}}$, $\|T_R\hat{V} - T_R\hat{V}'\|_\infty \leq \beta\|\hat{V} - \hat{V}'\|_\infty$.*

2. *The equation* $T_R\hat{V} = \hat{V}$ *has a unique solution* $\hat{V}^*$.

3. $\hat{V}^* = \lim_{k\to\infty} T_R^k \hat{V}$ *for all* $\hat{V} \in \hat{\mathcal{V}}$.

Proof. Our result is based on a special case of Iyengar (2005) (see Theorem 3.2). However, to allow for greater generality in the state space and action space we use a different proof based on classic dynamic programming results. Define

$$(\hat{T}_R\hat{V})(x,\hat{s}) = \sup_{\substack{f\in\mathcal{S}_f(\hat{s})\\ \iota\in\mathcal{I}\\ \rho\geq 0}} \Big\{\pi(x,\hat{s}) + \mathrm{E}\Big[\phi\mathbf{1}\{\phi\geq\rho\} + \mathbf{1}\{\phi<\rho\}\big[-c(x,\iota) + \beta\,\mathrm{E}_{\mu,\lambda}[\hat{V}(x_{i,t+1},\hat{s}_{t+1})\big|x_{it}=x, s_t=(f,d,z),\iota]\big]\Big]\Big\}. \quad (14)$$

The statement of the lemma holds for $\hat{T}_R$ from the results for model (D) in Bertsekas and Shreve (1978) (recall that we assumed bounded profits). The statement follows for T_R since $\hat{T}_R\hat{V} = T_R\hat{V}$ for all bounded $\hat{V}$. □

Proof of Theorem 8.1. Take vectors $\hat{V} \in \hat{\mathcal{V}}$ and V, where $V : \mathcal{X}\times\mathcal{S} \to \Re$, such that $V(x,s) = \hat{V}(x,\hat{s})$, for all moment-based industry state $\hat{s}$ that is consistent with industry state s. Let T^* be the Bellman operator associated with θ^*, that is, with the full Markov best response that keeps track of the entire industry state. It is simple to observe that $T^*V(x,s) \leq \hat{T}\hat{V}(x,\hat{s})$, for all x, and all $\hat{s}$ and s consistent. By the monotonicity of the $\hat{T}$ and T^* operators we conclude that $(T^*)^kV(x,s) \leq \hat{T}^k\hat{V}(x,\hat{s})$ for all $k\geq 1$. Taking k to infinity we get $\hat{T}^k\hat{V} \to \hat{V}$ from Lemma A.1, and $(T^*)^kV \to \overline{V}^*$ by standard dynamic programming arguments. Therefore, $\overline{V}^*(x,s) \leq \hat{V}(x,\hat{s})$ for s and $\hat{s}$ consistent.

□

B Second Approach: Numerical Experiments

This appendix reports the parameters and transition dynamics of the numerical experiments of the second approach as well as some details about the algorithm. The transition structure for dominant firms is based on the investment transitions in Weintraub et al. (2010). For a dominant firm with investment ι,

$$x_{t+1} = \begin{cases} \min(\bar{x}_d, x+1) & \text{w.p. } \frac{\delta a\iota}{1+a\iota} \\ x & \text{w.p. } \frac{(1-\delta')+(1-\delta)a\iota}{1+a\iota} \\ \max(\underline{x}_d, x-1) & \text{w.p. } \frac{\delta'}{1+a\iota}, \end{cases}$$

where $\delta,\delta' \in (0,1)$ and $a>0$ are constants, and $\bar{x}_d$ and $\underline{x}_d$ are the highest and lowest values that dominant firms states can take.

Now, we specify the transitions for fringe firms according to equation (8). Let ζ_f' take values in $\{\zeta_1,\zeta_2,\ldots,\zeta_L\}$ with $\zeta_l < \zeta_{l+1}$ and let $L > \bar{l} > 1$ be some interior index. We define the investment

Notation	Value	Description
β	.95	Discount factor
m	30	Market size
p	.05	Power in fringe transitions
$\mathcal{X}_d$	$\{6, 7, \ldots, 11\}$	Dominant firms' state space
$\mathcal{X}_f$	$[0,\ 10]$	Fringe firms' state space
(c_d, c_f)	(1.1,.1)	Investment cost rate (dominant, fringe)
δ_d ,δ'_d ,a_d	.4, .6, 1.5	Dominant firms' transition parameters
$(\zeta_1, \ldots, \zeta_5)$	$(.93, .96, 1, 1.09, 1.12)$	Fringe firms' transition parameters
$(\delta'_1, \delta'_2, \delta_4, \delta_5)$	$(.2, .6, .6, .2)$	Fringe firms' transition parameters
a_f	3	Fringe firms' transition parameters
x^e	1	Entry state
κ	29	Entry cost
$\bar{\phi}$	9.1	Expected sell-off value (exponential distribution)

Table 4: Model parameters for the experiments of Subsection 6.1

transitions by

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

with $\delta' = \sum_{l=1}^{\bar{l}-1} \delta'_l < 1$, $\delta = \sum_{l=\bar{l}+1}^{L} \delta_l < 1$, δ'_l is positive for all $l = 1, \ldots, \bar{l} - 1$, δ_l is positive for all $l = \bar{l} + 1, \ldots, L$ 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. In both transition structures for dominant and fringe firms, the optimal investment level has a closed form solution.

The algorithm discretizes the space of moments linearly (equal spaces) and uses linear interpolation when computing the perceived value function between grid points. We use geometric discretization for the set of fringe states $\mathcal{X}_f$ since the transition dynamics are essentially proportional (for p small).

Table 4 reports the values of some parameters.

C Beer Industry Experiments

Denote by x_{it} the goodwill of firm i at time t. The evolution of goodwill is similar to Pakes and McGuire (1994), but with a multiplicative growth model following Roberts and Samuelson (1988):

$$x_{it+1} = \begin{cases} \min(x^{\bar{n}}, x_{it}(1+\rho)) & \text{w.p. } \frac{\delta\psi(x)\iota_{it}}{1+\psi(x)\iota_{it}} \\ x_{it} & \text{w.p. } \frac{1-\delta'+(1-\delta)\psi(x)\iota_{it}}{1+\psi(x)\iota_{it}} \\ \max(x^1, x_{it}/(1+\rho)) & \text{w.p. } \frac{\delta'}{1+\psi(x)\iota_{it}}. \end{cases}$$

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

The moment space is discretized linearly and we use a bicubic spline to interpolate between grid points when computing firms' optimal strategies.

We use the profit function of Example 4.1. We take $\beta = .925$ and $(\delta', \delta) = (1, .55)$ as the transition parameters. After some experimentation we choose the market size $m = 30$. The other parameters are listed in the next table with their relevant sources.

Description	Value	Source
Number of firms (N)	200	This number is chosen to be greater than the maximal number of active firms in this period.
Maximal number of dominant firms ($\overline{D}$)	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.
Production cost per barrel (c)	\$120	Rojas (2008) estimated the markup to be about a third of the price, and the average price is \$165 per barrel in the period studied.
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 plants.
Fixed cost per period fringe (double for dominant)	10^6	We introduce this fixed cost in the single-period profit function.
Profit function parameters (Y and α_2)	200 and 1, resp.	Chosen to match the price elasticity (-.5) for the average price, see (Tremblay and Tremblay, 2005, p. 23).

D Third Approach

D.1 Real Time Algorithm for Empirical Transitions Algorithm

Given strategies (μ, λ) and their associated value functions, it is useful to define

$$W(x, \hat{s}|\mu, \lambda) = \mathrm{E}_{\mu,\lambda}\big[V(x, \hat{s}_{t+1}|\mu, \lambda)\big|\hat{s}_t = \hat{s}\big], \tag{15}$$

where the expectation is taken with respect to the perceived transition kernel. The function 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 that transitions to x will integrate over $(\theta_{t+1}, d_{-i,t+1}, z_{t+1})$ with $d_{t+1} = (x_{i,t+1}, d_{-i,t+1})$. Now, we can write the Bellman equation associated with $\mathcal{C}1$ in the equilibrium definition as follows:

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

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. Based on this formulation we introduce our real-time dynamic programming algorithm to compute MME; see Algorithm 2.

The steps of the algorithm are as follows. We begin with W functions with which firms' optimal decisions can be computed (investment, exit, and entry). Once these are determined, we can simulate the next industry state. The continuation value in the simulated state is used to update the W functions. The update step depends on the number of times the state was visited $e(\hat{s})$ and on the number of rounds (we take $\sigma(n) = \min(n, \bar{n})$ for some integer $\bar{n} > 0$ to allow for quick updating in the early rounds). At the end of the simulation/optimization phase we check whether the W functions have converged, and if so check for convergence with Algorithm 1 as well. If convergence has not occurred we iterate on the simulation/optimization phase.

D.2 Computation of Robust Bound

The reminder of the appendix provides details on the computation of the robust bound under the assumption that there is large number of fringe firms and therefore the one-step transition of the fringe firm state is assumed to be deterministic. When transitions are deterministic, finding the optimal consistent f in (12) is equivalent to choosing the next moment from an *accessibility* set of moments that can be reached from the current industry state. As we will explain in detail now, this considerably simplifies the computation of the inner maximization in (12), since the moment accessibility sets are low dimensional. Moreover, characterizating these accessibility sets can be done efficiently.

We assume throughtout that the set of individual fringe firm states is discrete and univariate, $\mathcal{X}_f = \{x^1, \ldots x^{\bar{n}}\}$, where $x^n \in \Re$, for all $n = 0, 1, \ldots, \bar{n} < \infty$.[20] In addition, for simplicity, we assume that there are no transitions between the dominant and fringe tiers.

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 $\theta = \sum_{x \in \tilde{\mathcal{X}}_f} f(x)(x)^\alpha$ for $\alpha \geq 0$, where $\tilde{\mathcal{X}}_f \subseteq \mathcal{X}_f$. This representation is general enough to include moments and other statistics. The extension to more moments is direct.

Assuming that the next fringe firm state is deterministic and equals to its expected value, we say that

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

Algorithm 2 Equilibrium solver with real-time dynamic programming

Initiate $W(x, \hat{s}) := 0$, for all $(x, \hat{s}) \in \mathcal{X} \times \hat{\mathcal{S}}$;
$e(\hat{s}) = 0$ for all $\hat{s} \in \hat{\mathcal{S}}$;
Initiate industry state (f_0, d_0, z_0) and $\hat{s}_0 := (\theta(f_0), d_0, z_0))$;
$\Delta_w := \delta_w + 1$; $n := 1$
while $\Delta_w > \delta_w$ **do**
  $W'(x, \hat{s}) := W(x, \hat{s})$ for all $(x, \hat{s}) \in \mathcal{X} \times \hat{\mathcal{S}}$;
  $t := 1$;
  while $t \leq T$ **do**
    for all x with $f_t(x) > 0$ or $x \in d_t$, **do**
      Compute optimal strategies using $V(x, \hat{s}_t; W')$ and store them;
    end for;
    Compute optimal entry cutoff from $V(x^e, \hat{s}_t; W)$ and store it;
    Simulate $(f_{t+1}, d_{t+1}, z_{t+1})$ and $\hat{s}_{t+1}$ from these strategies;
    Let $\gamma := \frac{1}{\sigma(n)+e(\hat{s}_t)}$;
    for all $x' \in \mathcal{X}_f$ **do**
      Compute $V(x', \hat{s}_{t+1}; W')$;
      Update $W(x', \hat{s}_t) := \gamma V(x', \hat{s}_{t+1}; W') + (1-\gamma) W'(x', \hat{s}_t)$;
    end for;
    for all Dominant firm i and $x' \in \mathcal{X}_d$ that is accessible in one step from x_{it} **do**
      Define $\hat{s}'_{t+1}$ to be the industry state $\hat{s}_{t+1}$ when firm i transitions to state x';
      Compute $V(x', \hat{s}'_{t+1}; W')$;
      Update $W(x', \hat{s}_t) := \gamma V(x', \hat{s}'_{t+1}; W') + (1-\gamma) W'(x', \hat{s}_t)$;
    end for;
    $e(\hat{s}_t) := e(\hat{s}_t) + 1$, $t := t + 1$;
  end while;
  $\Delta_w := \|W' - W\|_\infty$;
  $e(\hat{s}) := 0$ for all $\hat{s} \in \hat{\mathcal{S}}$;
  $(f_0, d_0, z_0) := (f_{K+1}, d_{K+1}, z_{K+1})$, and $\hat{s}_0 := (\theta(f_0), d_0, z_0)$;
  $n := n + 1$;
end while;
Compute $\mu(x, \hat{s})$ and $\lambda(\hat{s})$ from $V(x, \hat{s}; W)$ for all $(x, \hat{s}) \in \mathcal{X} \times \hat{\mathcal{S}}$;
Run Algorithm 1 with these strategies (used as stopping criteria);

moment θ' is accessible from moment θ in industry state $\hat{s}$ if there exists a fringe firm state $f \in \mathcal{S}_f(\hat{s})$ that solves the following system of linear equations,

$$\begin{aligned}
&\sum_{x \in \tilde{\mathcal{X}}_f} f(x)\,(x)^\alpha = \theta \\
&\sum_{x \in \tilde{\mathcal{X}}_f} f(x)\,\mathrm{E}_\mu[(x_{i,t+1})^\alpha | (x_{it}, \hat{s}_t) = (x, \hat{s})] + \mathbf{1}\{x^e \in \tilde{\mathcal{X}}_f\}(x^e)^\alpha \mathcal{P}(\kappa_{it} \leq \lambda(\hat{s}))N^e(f) = \theta' \\
&\sum_{x \in \mathcal{X}_f} f(x) \leq N, \quad f \in \mathbb{N}^{\overline{n}},
\end{aligned} \tag{16}$$

where the first equation states that the current moment is consistent with the fringe firm state, and the second that the expected next moment is θ'. We denote by $N^e(f)$ as the number of potential entrants at fringe firm state f. 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 $\hat{s}$, that is, if there is a fringe firm state consistent with $\hat{s}$ such that the expected next moment is θ'. Note that the accessibility sets depend on the investment and entry MME strategies that control the expected next moment in (16). Due to the integrability constraint $f(x) \in \mathbb{N}^{\overline{n}}$, the computation of $A(\hat{s})$ is demanding. However, we can relax this integrability 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

$$\begin{aligned}
(\hat{T}\hat{V})(x, \hat{s}) = \sup_{\substack{\iota \in \mathcal{I} \\ \rho \geq 0}} \sup_{\theta' \in \hat{A}(\hat{s})} \Big\{ \pi(x, \hat{s}) + \mathrm{E}\Big[\phi \mathbf{1}\{\phi \geq \rho\} \\
+ \mathbf{1}\{\phi < \rho\}\big[-c(x, \iota) + \beta\,\mathrm{E}_{\mu,\lambda}[\hat{V}(x_{i,t+1}, (\theta', d_{t+1}, z_{t+1})) \big| x_{it} = x, \hat{s}_t = \hat{s}, \iota]\big]\Big]\Big\},
\end{aligned}$$

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

Proposition D.1. *Assume that there are no transitions between tiers and that the fringe firm state follows deterministic transitions given by the expected next state (equation* (16)*). Then*

$$(T_R\hat{V}) \leq (\hat{T}\hat{V}),$$

for every $\hat{V} \in \hat{\mathcal{V}}$. Moreover, the operator $\hat{T}$ satisfies the same properties than the operator T_R given in Lemma A.1.

Proof. Without tier transitions the evolution of the moments is independent of the evolution of dominant firms. If fringe firms' transitions are also deterministic each $f \in \mathcal{S}_f(\hat{s})$ maps to an expected next moment and so we can optimize over the set of these moments instead. The inequality follows because we consider the relaxed accessibility sets, so we optimize over a larger set in $\hat{T}$. $\square$

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 $\overline{V}^*$, assuming deterministic fringe transitions (by Theorem 8.1 together with the previous proposition). For problems for which MME is solvable this procedure is generally computationally manageable.