

# DYNAMIC GAMES WITH ASYMMETRIC INFORMATION: A FRAMEWORK FOR EMPIRICAL WORK\*

CHAIM FERSHTMAN AND ARIEL PAKES

We develop a framework for the analysis of dynamic oligopolies with persistent sources of asymmetric information that enables applied analysis of situations of empirical importance that have been difficult to deal with. The framework generates policies that are “relatively” easy for agents to use while still being optimal in a meaningful sense, and is amenable to empirical research in that its equilibrium conditions can be tested and equilibrium policies are relatively easy to compute. We conclude with an example that endogenizes the maintenance decisions of electricity generators when the costs states of the generators are private information. *JEL* Codes: L13, C73, D82.

## I. INTRODUCTION

This article develops a framework for the analysis of dynamic oligopolies with persistent sources of asymmetric information that can be used in a variety of situations which are both of empirical importance and have not been adequately dealt with in prior applied work. These situations include competition between producers when there is a producer attribute which is unknown to its competitors and serially correlated over time, investment games where the outcome of the investment is unobserved, or repeated auctions for primary products (e.g., timber) where the capacity available to process the quantity acquired at the auction is private information. Less obviously, but probably more empirically important, the framework also allows us to analyze markets in which the decisions of both producers and consumers have dynamic implications, but consumers make decisions with different information sets than producers do. As will be discussed, this enables applied analysis of dynamics in durable, experience, storeable, and network good industries.

In building the framework we have two goals. First we want a framework that generates policies that are “relatively” easy for agents to use while still being optimal in some meaningful sense

\*We thank two referees; the editor Elhanan Helpman; John Asker, Susan Athey, Adam Brandenburger, Eddie Dekel, Liran Einav, Drew Fudenberg, Phil Haile, Robin Lee, Greg Lewis, and Michael Ostrovsky for constructive comments; and Niyati Ahuja for superb research assistance. We would like to thank the Binational Science Foundation (BSF 2008020) for financial support.

© The Author(s) 2012. Published by Oxford University Press, on behalf of President and Fellows of Harvard College. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

*The Quarterly Journal of Economics* (2012), 1611–1661. doi:10.1093/qje/qjs025.

Advance Access publication on August 30, 2012.

of the word. In particular the framework should not *require* the specification and updating of players' beliefs about their opponents types, as in perfect Bayesian equilibrium, and should not require agents to retain information that it is impractical for them to acquire. Second we want the framework to be usable by empirical researchers; its conditions should be defined in terms of observable magnitudes and it should generate policies which can be computed with relative ease (even when there are many underlying variables that effect the returns to different choices). The twin goals of ease of use to agents and ease of analysis by the applied research work out, perhaps not surprisingly, to have strong complementarities.

To accomplish these tasks we extend the framework in Ericson and Pakes (1995) to allow for asymmetric information.<sup>1</sup> Each agent's returns in a given period are determined by all agents' "payoff relevant" state variables and their actions. The payoff relevant random variables of producers would typically include indexes of their cost function, qualities of the goods they market, and so on, whereas in a durable good market those of consumers would include their current holdings of various goods and the household's demographic characteristics. Neither a player's "payoff relevant" state variables nor its actions are necessarily observed by other agents. Thus producers might not know either the cost positions or the details of supplier contacts of their competitors, and in the durable goods example neither consumers nor producers need know the entire distribution of holdings crossed with household characteristics (even though this will determine the distribution of future demand and prices).

The fact that not all state variables are observed by all agents and that the unobserved states may be correlated over time implies that variables that are not currently payoff relevant but are related to the unobserved past states of other agents will help predict other agent's behavior. Consequently, they will help predict the returns from a given agent's current actions. So in addition to payoff relevant state variables agents have "informationally relevant" state variables. For example, in many markets past prices will be known to agents and will contain information on likely future prices.

1. Indeed our assumptions nest the generalizations to Ericson and Pakes (1995) reviewed in Doraszelski and Pakes (2007).

The “types” of the agents, which are defined by their state variables, are only partially observed by other agents and evolve over time. In the durable goods example, the joint distribution of household holdings and characteristics will evolve with household purchases, and the distribution of producer costs and goods marketed will evolve with the outcomes of investment decisions. As a result each agent continually changes its perceptions of the likely returns from its own possible actions.<sup>2</sup>

Recall that we wanted our equilibrium concept to be testable. This, in itself, rules out basing these perceptions on Bayesian posteriors, as these posteriors are not observed. Instead we assume that the agents use the outcomes they experienced in past periods that had conditions similar to the conditions the agent is currently faced with to form an estimate of expected returns from the actions they can currently take. Agents act so as to maximize the discounted value of future returns implied by these expectations. So in the durable goods example a consumer will know its own demographics and may have kept track of past prices, while the firms might know past sales and prices. Each agent would then choose the action that maximized its estimate of the expected discounted value of its returns conditional on the information at its disposal. We base our equilibrium conditions on the consistency of each agents’ estimates with the expectation of the outcomes generated by the agents’ decisions.

More formally we define a state of the game to be the information sets of all of the players (each information set contains both public and private information). An experience-based equilibrium (EBE) for our game is a triple which satisfies three conditions. The triple consists of (1) a subset of the set of possible states, (2) a vector of strategies defined for every possible information set of each agent, and (3) a vector of values for every state that provides each agent’s expected discounted value of net cash flow conditional on the possible actions that agent can take at that state. The conditions we impose on this triple are as follows. The first condition is that the equilibrium policies ensure that once we visit a state in our subset we stay within that subset in all future periods, visiting each point in that subset repeatedly. That is, the subset of states is a recurrent class of the Markov

2. Dynamic games with asymmetric information have not been used extensively to date, a fact that attests (at least in part) to their complexity. Notable exceptions are Athey and Bagwell (2008), and Cole and Kocherlakota (2001).

process generated by the equilibrium strategies. The second condition is that the strategies are optimal given the evaluations of outcomes. The final condition is that optimal behavior given these evaluations actually generates expected discounted value of future net cash flows that are consistent with these evaluations in the recurrent subset of states.

The conditions defining an EBE do not restrict the players' perceptions of the value of outcomes from nonequilibrium strategies. This is consistent with our focus on equilibrium conditions that use information that could be learned from the outcomes of equilibrium play, as without further restrictions neither the players nor researchers would have access to the experience which would be required to obtain consistent estimates of the outcomes from nonequilibrium strategies. Consequently the EBE conditions, in and of themselves, do not restrict agents' behavior; any profile of strategies can be rationalized as EBE. There are two approaches to choosing restrictions on the value of outcomes from nonequilibrium strategies, and they will be appropriate in different applied problems. One approach is to impose further conditions on the equilibrium concept, an alternative we explore by introducing a restricted EBE. A restricted EBE imposes conditions on the outcome of all feasible strategies from points in the recurrent class. In a restricted EBE any outcome from a feasible strategy that is in the recurrent class is consistent with the expected discounted value of future net cash that would be generated from equilibrium play from that outcome. We provide familiar examples of models used in applied work in which the players can construct consistent estimates of all the values required to satisfy the equilibrium conditions of a restricted EBE from knowledge of the primitives of the game and the outcomes from equilibrium play. Alternatively (or additionally), if data are available on either equilibrium play or the outcomes from that play, then that data and assumptions on the primitives of the game will suffice to restrict nonequilibrium play.

We show that a restricted EBE that is consistent with a given set of primitives can be computed using a simple (reinforcement) learning algorithm. Moreover the equilibrium conditions are testable, and the testing procedure does not require computation of posterior distributions. Neither the iterative procedure which defines the computational algorithm nor the test of the equilibrium conditions have computational burdens which increase at a particular rate as we increase the number of variables that

impact on returns; neither is subject to a curse of dimensionality. At least in principle this should lead to an ability to analyze models that contain many more state variables, and hence are likely to be much more realistic, then could be computed using standard Markov perfect equilibrium concepts.<sup>3</sup>

One could view our reinforcement learning algorithm as a description of how players learn the implications of their actions in a changing environment. This provides an alternative reason for interest in the output of the algorithm. However the learning rule would not, by itself, restrict behavior without either repeated play or prior information on initial conditions. Also the fact that the equilibrium policies from our model can be learned from past outcomes accentuates the fact that those policies are most likely to provide an adequate approximation to the evolution of a game in which it is reasonable to assume that agents' perceptions of the likely returns to their actions can be learned from the outcomes of previous play. Since the states of the game evolve over time and the possible outcomes from each action differ by state, if agents are to learn to evaluate these outcomes from prior play the game needs to be confined to a finite space.

When all the state variables are observed by all the agents, our equilibrium notion is similar to but weaker than the familiar notion of Markov perfect equilibrium as used in Maskin and Tirole (1988, 2001). This is because we only require that the evaluations of outcomes used to form strategies be consistent with competitors' play when that play results in outcomes that are in the recurrent subset of points, and hence are observed repeatedly. We allow for feasible outcomes that are not in the recurrent class, but the conditions we place on the evaluations of those outcomes are weaker; they need only satisfy inequalities that ensure they are not observed repeatedly. In this sense our notion of equilibrium is akin to the notion of self-confirming

3. For alternative computational procedures see the review in Doraszelski and Pakes (2007). Pakes and McGuire (2001), show that reinforcement learning has significant computational advantages when applied to full information dynamic games, a fact which has been used in several applied papers; for example, Goettler, Parlour, and Rajan (2005) and Beresteanu and Ellickson (2006). Goettler, Parlour, and Rajan (2008), use it to approximate optimal behavior in finance applications. We show that a similar algorithm can be used in games with asymmetric information and provide a test of the equilibrium conditions which is not subject to a curse of dimensionality. The test in the original Pakes and McGuire article was subject to such a curse and it made their algorithm impractical for large problems.

equilibrium, as defined by Fudenberg and Levine (1993) (though our application is to dynamic games). An implication of using the weaker equilibrium conditions is that we might admit more equilibria than the Markov perfect concept would. The restrictions used in the restricted EBE reduce the number of equilibria.

The original Maskin and Tirole (1988) article and the framework for the analysis of dynamic oligopolies in Ericson and Pakes (1995) laid the groundwork for the applied analysis of dynamic oligopolies with symmetric information. This generated large empirical and numerical literatures on an assortment of applied problems (see Benkard 2004 or Gowrisankaran and Town 1997 for empirical examples and Doraszelski and Markovich 2007 or Besanko et al. 2010 for examples of numerical analysis). None of these models have allowed for asymmetric information. Our hope is that the introduction of asymmetric information in conjunction with our equilibrium concept helps the analysis in two ways. It enables the applied researcher to use more realistic behavioral assumptions and hence provide a better approximation to actual behavior, and it simplifies the process of analyzing such equilibria by reducing its computational burden.

As noted, this approach comes with its own costs. First, it is most likely to provide an adequate approximation to behavior in situations for which there is a relevant history to learn from. Second, our equilibrium conditions enhance the possibility for multiple equilibria over more standard notions of equilibria. With additional assumptions one might be able to select out the appropriate equilibria from data on the industry of interest, but there will remain the problem of choosing the equilibria for counterfactual analysis.

We conclude with an example that endogenizes the maintenance decisions of electricity generators. We take an admittedly simplified set of primitives and compute and compare equilibria based on alternative institutional constraints. These include asymmetric information equilibria where there are no bounds on agent memory, asymmetric information equilibria where there are such bounds, symmetric information equilibria, and the solutions to the social planner problem in two environments—one with more capacity relative to demand than the other. We show that in this environment the extent of excess capacity relative to demand has economically significant effects on equilibrium outcomes.

The next section describes the primitives of the game. Section II provides a definition of and sufficient conditions for our notion of an EBE. Section III provides an algorithm to compute and test for this equilibrium, and Section IV contains our example.

## II. DYNAMIC OLIGOPOLIES WITH ASYMMETRIC INFORMATION

We extend the framework in Ericson and Pakes (1995) to allow for asymmetric information.<sup>4</sup> In each period there are  $n_t$  potentially active firms, and we assume that with probability one  $n_t \leq \bar{n} < \infty$  (for every  $t$ ). Each firm has payoff relevant characteristics. Typically these will be characteristics of the products marketed by the firm or determinants of their costs. The profits of each firm in every period are determined by their payoff relevant random variables, a subset of the actions of all the firms, and a set of variables which are common to all agents and account for common movements in factor costs and demand conditions, say,  $d \in D$ , where  $D$  is a finite set. For simplicity we assume that  $d_t$  is observable and evolves as an exogenous first-order Markov process.

The payoff relevant characteristics of firm  $i$ , denoted by  $\omega_i \in \Omega_i$ , take values on a finite set of points for all  $i$ . There will be two types of actions—actions that will be observed by the firm's competitors  $m_i^o$ , and those that are unobserved  $m_i^u$ . For simplicity we assume that both take values on a finite state space, so  $m_i = (m_i^o, m_i^u) \in \mathcal{M}_i$ .<sup>5</sup> Notice that, also for simplicity, we limit ourselves to the case where an agent's actions are either known only to itself (they are “private” information), or to all agents (they are “public” information). For example, in an investment game the prices the firm sets are typically observed, but the investments a firm makes in the development of its products may not be. Though both controls could affect current profits or the

4. Indeed our assumptions nest the generalizations to Ericson and Pakes (1995), and the amendments to it introduced in Doraszelski and Satterthwaite (2010), and reviewed in Doraszelski and Pakes (2007). The latter paper also provide more details on the underlying model.

5. As in Ericson and Pakes (1995), we could have derived the assumption that  $\Omega$  and  $\mathcal{M}$  are bounded sets from more primitive conditions. Also the original version of this paper (which is available on request) included both continuous and discrete controls, where investment was the continuous control. It was not observed by agent's opponents and affected the game only through its impact on the transition probabilities for  $\omega$ .

probability distribution of payoff relevant random variables, they need not. A firm might simply decide to disclose information or send a signal of some other form.

Letting  $i$  index firms, realized profits for firm  $i$  in period  $t$  are given by

$$(1) \quad \pi(\omega_{i,t}, \omega_{-i,t}, m_{i,t}, m_{-i,t}, d_t),$$

where  $\pi(\cdot) : \times_{i=1}^n \Omega_i \times \times_{i=1}^n M_i \times D \rightarrow R$ .  $\omega_{i,t}$  evolves over time and its conditional distribution may depend on the actions of all competitors, that is

$$(2) \quad \mathcal{P}_\omega = \{P_\omega(\cdot | m_i, m_{-i}, \omega); (m_i, m_{-i}) \in \times_{i=1}^n \mathcal{M}_i, \omega \in \Omega\}.$$

Some examples will illustrate the usefulness of this structure.

A familiar special case occurs when the probability distribution of  $\omega_{i,t+1}$ , or  $P_\omega(\cdot | m_i, m_{-i}, \omega)$ , does not depend on the actions of a firm's competitors, or  $m_{-i}$ . Then we have a "capital accumulation" game. For example in the original Ericson and Pakes (1995) model,  $m$  had two components, price and investment, and  $\omega$  consisted of characteristics of the firm's product or its cost function that the firm was investing to improve. Their  $\omega_{i,t+1} = \omega_{i,t} + \eta_{i,t} - d_t$ , where  $\eta_{i,t}$  was a random outcome of the firm's investment whose distribution was determined by  $P_\omega(\cdot | m_{i,t}, \omega_{i,t})$ , and  $d_t$  was determined by aggregate costs or demand conditions.

Now consider a sequence of timber auctions with capacity constraints for processing the harvested timber. Each period there is a new lot up for auction, firms submit bids (a component of our  $m_i$ ), and the firm that submits the highest bid wins. The quantity of timber on the lot auctioned may be unknown at the time of the auction but is revealed to the firm that wins the lot. The firm's state (our  $\omega_i$ ) is the amount of unharvested timber on the lots the firm owns. Each period each firm decides how much to bid on the current auction (our first component of  $m_i$ ) and how much of its unharvested capacity to harvest (a second component of  $m_i$  which is constrained to be less than  $\omega_i$ ). The timber that is harvested and processed is sold on an international market which has a price that evolves exogenously (our  $\{d_t\}$  process), and revenues equal the amount of harvested timber times this price. Then the firm's stock of unharvested timber in  $t+1$ , our  $\omega_{i,t+1}$  is  $\omega_{i,t}$  minus the harvest during period  $t$  plus the amount on lots for

which the firm won the auction. The latter, the amount won at auction, depends on  $m_{-i,t}$ , that is, the bids of the other firms, as well as on  $m_{i,t}$ .

Finally, consider a market for durable goods. Here we must explicitly consider both consumers and producers. Consumers are differentiated by the type and vintage of the good they own and their characteristics, which jointly define their  $\omega_i$ , and possibly by information they have access to which might help predict future prices and product qualities. Each period the consumer decides whether to buy a new vintage and if so which one (a consumer's  $m_i$ )—a choice that is a determinant of the evolution of their  $\omega_i$ . Producers determine the price of the product marketed and the amount to invest in improving their product's quality (the components of the producer's  $m_i$ ). These decisions are a function of current product quality and its own past sales (both components of the firm's  $\omega_i$ ), as well as other variables that affect the firm's perceptions about demand conditions. Since the price of a firm's competitors will be a determinant of the firm's sales, this is another example where the evolution of the firm's  $\omega_{i,t+1}$  depends on  $m_{-i,t}$  as well as on  $m_{i,t}$ .

The information set of each player at period  $t$  is, in principle, the history of variables that the player has observed up to that period. We restrict ourselves to a class of games in which each agent's strategies are a mapping from a subset of these variables, in particular from the variables that are observed by the agent *and* are either "payoff" or "informationally" relevant, where these two terms are defined as follows. The "payoff relevant" variables are defined, as in Maskin and Tirole (2001), to be those variables that are not current controls and affect the current profits of at least one of the firms. In terms of equation (1), all components of  $(\omega_{i,t}, \omega_{-i,t}, d_t)$  that are observed are payoff relevant. Observable variables that are not payoff relevant will be informationally relevant if and only if either (1) even if no other agent's strategy depends on the variable player  $i$  can improve its expected discounted value of net cash flows by conditioning on it, or (2) even if player  $i$ 's strategy does not condition on the variable there is at least one player  $j$  whose strategy will depend on the variable. For example, say all players know  $\omega_{j,t-1}$  but player  $i$  does not know  $\omega_{j,t}$ . Then even if player  $j$  does not condition its strategy on  $\omega_{j,t-1}$ , since  $\omega_{j,t-1}$  can contain information on the distribution of the payoff relevant  $\omega_{j,t}$  which, in turn, will affect  $\pi_{i,t}(\cdot)$  through

its impact on  $m_{j,t}$ , player  $i$  will generally be able to gain by conditioning its strategy on that variable.<sup>6</sup>

As before, we limit ourselves to the case where information is either known only to a single agent ("private") or to all agents ("public"). The publicly observed component will be denoted by  $\xi_t \in \Omega(\xi)$ , while the privately observed component will be  $z_{i,t} \in \Omega(z)$ . For example  $\omega_{j,t-1}$  may or may not be known to agent  $i$  at time  $t$ ; if it is known  $\omega_{j,t-1} \in \xi_t$ , otherwise  $\omega_{j,t-1} \in z_{j,t}$ . Since the agent's information at the time actions are taken consists of  $J_{i,t} = (\xi_t, z_{i,t}) \in \mathcal{J}_i$ , we assume strategies are functions of  $J_{i,t}$ , that is,

$$m(J_{i,t}) : \mathcal{J}_i \rightarrow \mathcal{M}.$$

Notice that if  $\omega_{j,t}$  is private information and affects the profits of firm  $i$  then we will typically have  $\pi_{i,t} \in z_{i,t}$ .

We use our examples to illustrate. We can embed asymmetric information into the original Ericson and Pakes (1995) model by assuming that  $\omega_{i,t}$  has a product quality and a cost component. Typically quality would be publicly observed, but the cost would not be and so becomes part of the firm's private information. Current and past prices are also part of public information set and contain information on the firms' likely costs, while investment may be public or private. In the timber auction example, the stock of unharvested timber is private information, but the winning bids (and possibly all bids), the published characteristics of the lots auctioned, and the marketed quantities of lumber are public information. In the durable good example the public information is the history of prices, but we need to differentiate between the private information of consumers and that of producers. The private information of consumers consists of the vintage and type of the good it owns and its own characteristics, whereas the firm's private information includes the quantities it sold in prior periods and typically additional information whose contents will depend on the appropriate institutional structure.

Throughout we only consider games where both  $\#\Omega(\xi)$  and  $\#\Omega(z)$  are *finite*. This will require us to impose restrictions on the

6. Note that these definitions will imply that an equilibrium in our restricted strategy space will also be an equilibrium in the general history-dependent strategy space.

structure of informationally relevant random variables, and we come back to a discussion of situations in which these restrictions are appropriate later. To see why we require these restrictions, recall that we want to let agents base decisions on past experience. For the experience to provide an accurate indication of the outcomes of policies we will need a visit a particular state repeatedly; a condition we can only insure when there is a finite state space.

### III. EXPERIENCE-BASED EQUILIBRIUM

This section is in two parts. We first consider our basic equilibrium notion and then consider further restrictions on equilibrium conditions that will sometimes be appropriate.

For simplicity we assume all decisions are made simultaneously so there is no subgame that occurs within a period. In particular we assume that at the beginning of each period there is a realization of random variables and players update their information sets. Then the players decide simultaneously on their policies. The extension to multiple decisions nodes within a period is straightforward.

Let  $s$  combine the information sets of all agents active in a particular period, that is,  $s = (J_1, \dots, J_n)$  when each  $J_i$  has the same public component  $\xi$ . We say that  $J_i = (z_i, \xi)$  is a component of  $s$  if it contains the information set of one of the firms whose information is combined in  $s$ . We can write  $s$  more compactly as  $s = (z_1, \dots, z_n, \xi)$ . So  $S = \{s: z \in \Omega(z)^n, \xi \in \Omega(\xi), \text{ for } 0 \leq n \leq \bar{n}\}$  lists the possible states of the world.

Firms' strategies in any period are a function of their information sets, so they are a function of a component of that period's  $s$ . From equation (2) the strategies of the firms determine the distribution of each firm's information set in the next period, and hence together the firms' strategies determine the distribution of the next period's  $s$ . As a result *any* set of strategies for all agents at each  $s \in S$ , together with an initial condition, defines a Markov process on  $S$ .

We have assumed that  $S$  is a finite set. As a result each possible sample path of any such Markov process will, in finite time, wander into a subset of the states in  $S$ , say  $\mathcal{R} \subset S$ , and once in  $\mathcal{R}$  stay within it forever.  $\mathcal{R}$  could equal  $S$  but typically will not, as the strategies the agents choose will often ensure that some

states will not be visited repeatedly, a point we return to later.<sup>7</sup>  $\mathcal{R}$  is referred to as a recurrent class of the Markov process as each point in  $\mathcal{R}$  will be visited repeatedly.

Note that this implies that the empirical distribution of next period's state given any current  $s \in \mathcal{R}$  will eventually converge to a distribution, and this distribution can be constructed from actual outcomes. This will also be true of the relevant marginal distributions, for example, the joint distribution of the  $J_i$  components of  $s$  that belong to different firms, or that belong to the same firm in adjacent time periods. We use a superscript  $e$  to designate these limiting empirical distributions, so  $p^e(J'_i|J_i)$  for  $J_i \subset s \in \mathcal{R}$  provides the limit of the empirical frequency that firm  $i$ 's next period information set is  $J'_i$  conditional on its current information being  $J_i \in \mathcal{R}$  and so on.<sup>8</sup>

We now turn to our notion of EBE. It is based on the notion that at equilibrium players expected value of the outcomes from their strategies at states which are visited repeatedly are consistent with the actual distribution of outcomes at those states. Accordingly the equilibrium conditions are designed to ensure that at such states (1) strategies are optimal given participants' evaluations, and (2) that these evaluations are consistent with the empirical distribution of outcomes and the primitives of the model.

Notice that this implies that our equilibrium conditions could, at least in principle, be consistently tested.<sup>9</sup> To obtain a consistent test of a condition at a point we must, at least potentially, observe that point repeatedly. So we could only consistently test for conditions at points in a recurrent class. As we shall see this implies that our conditions are weaker than "traditional" equilibrium conditions. We come back to these issues, and

7. Freedman (1983) provides a precise and elegant explanation of the properties of Markov chains used here. Though there may be more than one recurrent class associated with any set of policies, if a sample path enters a particular  $\mathcal{R}$ , a point,  $s$ , will be visited infinitely often if and only if  $s \in \mathcal{R}$ .

8. Formally the empirical distribution of transitions in  $\mathcal{R}$  will converge to a Markov transition matrix, say,  $p^{e,T} \equiv \{p^e(s'|s) : (s', s) \in \mathcal{R}^2\}$ . Similarly the empirical distribution of visits on  $\mathcal{R}$  will converge to an invariant measure, say,  $p^{e,I} \equiv \{p^e(s) : s \in \mathcal{R}\}$ . Both  $p^{e,T}$  and  $p^{e,I}$  are indexed by a set of policies and a particular choice of a recurrent class associated with those policies. Marginal distributions for components of  $s$  are derived from these objects.

9. We say "in principle" here because this presumes that the researcher doing the testing can access the union of the information sets available to the agents that played the game.

their relationship to past work, after we provide our definition of equilibrium.

DEFINITION 1. An EBE consists of

- a subset  $\mathcal{R} \subset \mathcal{S}$ ;
- strategies  $m^*(J_i)$  for every  $J_i$  which is a component of any  $s \in \mathcal{S}$ ;
- expected discounted value of current and future net cash flow conditional on the decision  $m_i$ , say,  $W(m_i|J_i)$ , for each  $m_i \in \mathcal{M}_i$  and every  $J_i$  which is a component of any  $s \in \mathcal{S}$ ,

such that

CONDITION 1 ( $\mathcal{R}$  is a recurrent class). The Markov process generated by any initial condition  $s_0 \in \mathcal{R}$ , and the transition kernel generated by  $\{m^*\}$ , has  $\mathcal{R}$  as a recurrent class (so, with probability 1, any subgame starting from an  $s \in \mathcal{R}$  will generate sample paths that are within  $\mathcal{R}$  forever).

CONDITION 2 (Optimality of strategies on  $\mathcal{R}$ ). For every  $J_i$  which is a component of an  $s \in \mathcal{R}$ , strategies are optimal given  $W(\cdot)$ , that is  $m^*(J_i)$  solves

$$\max_{m_i \in \mathcal{M}_i} W(m_i|J_i).$$

CONDITION 3 (Consistency of values on  $\mathcal{R}$ ). Take every  $J_i$  which is a component of an  $s \in \mathcal{R}$ . Then

$$W(m^*(J_i)|J_i) = \pi^E(m^*(J_i), J_i) + \beta \sum_{J'_i} W(m^*(J'_i)|J'_i) p^e(J'_i|J_i),$$

where

$$\pi^E(m^*(J_i), J_i) \equiv \sum_{J_{-i}} \pi_i(\omega_i, m^*(J_i), \omega_{-i}, m^*(J_{-i}), d_t) p^e(J_{-i}|J_i),$$

and

$$\left\{ p^e(J'_i|J_i) \equiv \frac{p^e(J'_i, J_i)}{p^e(J_i)} \right\}_{J'_i}, \text{ and } \left\{ p^e(J_{-i}|J_i) \equiv \frac{p^e(J_{-i}, J_i)}{p^e(J_i)} \right\}_{J_{-i}} \quad \blacksquare.$$

(3)

Note that the evaluations  $\{W(m_i|J_i)\}$  need not be correct for  $J_i$  not a component of an  $s \in \mathcal{R}$ . Nor do we require correctness of

the evaluations for the  $W(m_i|J_i)$ 's associated with points in  $\mathcal{R}$  but at policies that differ from those in  $m_i^*$ . The only conditions on these evaluations are that choosing an  $m_i \neq m_i^*$  would lead to a perceived evaluation which is less than that from the optimal policy (this is ensured by condition C2).<sup>10</sup> On the other hand, the fact that our equilibrium conditions are limited to conditions on points that are played repeatedly implies that agents are able to learn the values of the outcomes from equilibrium play, and we provide an algorithm that would allow them to form consistent estimates of those outcomes. Further comments on our equilibrium notion follow.

*Beliefs on Types.* Note also that our conditions are not formulated in terms of beliefs about either the play or the "types" of opponents. There are three reasons for this to be appealing. First, as beliefs are not observed, they can not be directly tested. Second, as we will show presently, it implies that we can compute equilibria without ever explicitly calculating posterior distributions. Finally (and relatedly) we will show that an implication of the equilibrium conditions is that agent's can chose optimal strategies based on the agent's own observable experience; indeed the agents need not even know all the primitive parameters of the game they are playing.

*Relationship to Self-Confirming Equilibria.* EBE, though formulated for dynamic games, is akin to the notion of self-confirming equilibria (Fudenberg and Levine 1993), which has been used in other contexts.<sup>11</sup> Self-confirming equilibria weaken the standard Nash equilibrium conditions. It requires that each player has beliefs about opponents' actions and that the player's actions are best responses to those beliefs. However the players' beliefs need only be correct along the equilibrium path. This ensures that no players observes actions which contradicts its beliefs. Our equilibrium conditions explicitly introduce the evaluations that the agents use to determine their actions. They are similar to the conditions of self-confirming equilibria in that the most they ensure is that these evaluations are consistent with the opponents actions along the equilibrium path. However, we

10. The fact that our conditions do not apply to points outside of  $\mathcal{R}$  or to  $m_i \neq m_i^*$  implies that the conditional probabilities in equation (3) are well defined.

11. See also Dekel, Fudenberg, and Levine (2004) for an anlysis of self-confirming equilibrium in games with asymmetric information.

distinguish between states that are repeated infinitely often and those that are not, and we do not require the evaluations which determine actions at transitory states to be consistent with the play of a firm's opponents.

*Boundary Points.* It is useful to introduce a distinction made by Pakes and McGuire (2001). They partition the points in  $\mathcal{R}$  into interior and boundary points. Points in  $\mathcal{R}$  at which there are feasible (though inoptimal) strategies which can lead to a point outside of  $\mathcal{R}$  are labeled boundary points. Interior points are points that can only transit to other points in  $\mathcal{R}$  no matter which of the feasible policies are chosen (equilibrium or not). At boundary points there are actions which lead to outcomes which can not be consistently evaluated by the information generated by equilibrium play. This because our EBE notion does not restrict perceptions of returns from actions  $m \neq m^*$  for  $J_i \subset s \in \mathcal{R}$ .

*Multiplicity.* Notice that Bayesian perfect equilibria will satisfy our equilibrium conditions, and typically there will be a multiplicity of such equilibria. Since our EBE notion does not restrict perceptions of returns from actions not played repeatedly, it will admit an even greater multiplicity of equilibria. There are at least two ways to select out a subset of these equilibria. One is to impose further conditions on the definition of equilibrium, an alternative that we explore in the next subsection. As explained there, this requires a game form which enables agents to acquire information on outcomes from nonequilibrium play.

Alternatively (or additionally) if data are available we could use it to restrict the set of equilibria. If we observe or can estimate a subset of either  $\{W(\cdot)\}$  or  $\{m^*(\cdot)\}$  we can restrict any subsequent analysis to be consistent with their values. In particular since there are (generically) unique equilibrium strategies associated with any given equilibrium  $\{W(\cdot)\}$ , if we were able to determine the  $\{W(\cdot)\}$  associated with a point (say, through observations on sample paths of profits) we could determine  $m_i^*$  at that point, and conversely if we know  $m_i^*$  at a point we can restrict equilibrium  $\{W(\cdot)\}$  at that point. Similarly we can direct the computational algorithm we are about to introduce to compute an equilibria that is consistent with whatever data are observed. On the other hand were we to change a primitive of the model we could not single out the equilibria that is likely to result without further

assumptions (though one could analyze likely counterfactual outcomes if one is willing to assume a learning rule and an initial condition; see Lee and Pakes 2009).

### III.A. *Restricted EBE*

Our condition (3) only requires correct evaluations of outcomes from equilibrium actions that are observed repeatedly; that is, for  $W(m_i | J_i)$  at  $m_i = m_i^*$  and  $J_i \subset s \in \mathcal{R}$ . There are circumstances when imposing restrictions on equilibrium evaluations of actions off the equilibrium path for states that are observed repeatedly, that is, at  $m_i \neq m_i^*$  for  $J_i \subset s \in \mathcal{R}$ , might be natural, and this subsection explores them.

Barring compensating errors, for agents to have correct evaluations of outcomes from an  $m_i \neq m_i^*$  they will need to know (1) expected profits and the distribution of future states that result from playing  $m_i$ , and (2) the continuation values from the states that have positive probability when  $m_i$  is chosen. Whether or not agents can obtain the information required to compute expected profits and the distribution of future states when an  $m_i \neq m_i^*$  is played depends on the details of the game, and we discuss this further later. For now we assume that they can, and consider what this implies for restricting the evaluations of outcomes from nonoptimal actions.

Consider strengthening the condition (3) to make it apply to all  $m_i \in \mathcal{M}_i$  at any  $J_i \subset s \in \mathcal{R}$ . Then, at equilibrium, all outcomes that are in the recurrent class are evaluated in a way that would be consistent with the expected discounted value of returns that the action would yield were all agents (including itself) to continue playing their equilibrium strategies; this regardless of whether the action that generated the outcome was an equilibrium action. As in an unrestricted EBE outcomes that are not in the recurrent class are evaluated by perceptions which are not required to be consistent with any observed outcome.<sup>12</sup> As a result the restricted EBE ensures that in equilibrium when agents are at interior points they evaluate all feasible actions in a

12. We note that there are cases where it would be natural to require outcomes not in the recurrent class to be consistent with publicly available information on primitives. For example, even if a firm never exited from a particular state the agent might know its selloff value (or a bound on that value), and then it would be reasonable to require that the action of exiting be evaluated in a way that is consistent with that information. It is straightforward to impose such constraints on the computational algorithm introduced in the next section.

way that is consistent with expected returns given equilibrium play. However at boundary points only those actions whose outcomes are in the recurrent class with probability 1 are evaluated in this manner.

**DEFINITION 2 (Restricted EBE).** Let  $\pi^E(m_i, J_i)$  be expected profits and  $\{p(J'_i|J_i, m_i)\}_{J'_i}$  be the distribution of  $J'$ , both conditional on  $(m_i, J_i)$  and  $m_{-i}^*$ . A restricted EBE requires, in addition to C1 and C2, that

$$(4) \quad W(m_i|J_i) = \pi^E(m_i, J_i) + \beta \sum_{J'_i} W(m^*(J'_i)|J'_i)p(J'_i|J_i, m_i)$$

for all  $m_i \in \mathcal{M}_i$  and  $J_i \subset s \in \mathcal{R}$ .

We show how to compute and test for a restricted EBE in the next section. We now point out one of the implications of this definition and then consider situations that enable agents to acquire the information required to consistently evaluate  $W(m_i|J_i)$ , for  $m_i \neq m_i^*$ , and  $J_i \subset s \in \mathcal{R}$ .

Note that in some cases this equilibrium concept imposes a strong restriction on how agents react to nonequilibrium play by their competitors. To see this recall that the outcome is  $J'_i = (\xi', z'_i)$ , where  $\xi'$  contains new public, and  $z'_i$  new private, information. Competitors observe  $\xi'$  and  $\xi$ . Were an agent to play an  $m_i \neq m_i^*$  it may generate a  $\xi'$  which is not in the support of the distribution of  $\xi'$  generated by  $(\xi, m_i^*)$ .<sup>13</sup> Then if we impose the restrictions in equation (4) we impose constraints on the agent's evaluations of outcomes of actions which the agent's competitors would see as inconsistent with their experience from previous play. For the agent to believe such estimates are correct, the agent would have to believe that the competitor's play would not change were the competitor to observe an action off the equilibrium path. An alternative would be to assume that in equilibrium, agents only need to have correct evaluations for the outcomes of actions that competitor's could see as consistent with equilibrium play; that is, actions that generate a support for  $\xi'$  which is contained in the support  $\xi'$  conditional on  $(\xi, m_i^*)$ . Then we would only restrict equilibrium beliefs about outcomes from

13. As an example consider the case where  $m_i$  is observable. Then were the agent to play  $\tilde{m}_i \neq m_i^*$ ,  $\tilde{m}_i$  would be in  $\xi'$  and, provided there does not exist a  $\tilde{J}_i = (\xi, \tilde{z}_i)$  such that  $m^*(\xi, \tilde{z}_i) = \tilde{m}_i$ , the support of  $\xi'$  given  $(\xi, \tilde{m}_i)$  will differ from that given  $(\xi, m_i^*)$ .

actions that no agent perceives as inconsistent with equilibrium play. We do not pursue this further here, but one could modify the computational algorithm to accommodate this definition of a restricted EBE rather than the one in equation (4).

As noted for agents to be able to evaluate actions in a manner consistent with the restricted EBE they must know  $\pi^E(m_i, J_i)$  and  $\{p(J'_i|J_i, m_i)\}_{J'_i}$  for  $m_i \neq m_i^*$  at all  $J_i \subset s \in \mathcal{R}$ . We now consider situations in which these objects can be computed from the information generated by equilibrium play or knowledge of the primitives of the problem.<sup>14</sup> The reader who is not interested in these details can proceed directly to the next section.

We consider a case where  $\pi^E(m_i, J_i)$  can be consistently estimated<sup>15</sup>, and investigate situations in which the agent can calculate  $W(m_i|J_i)$ ,  $\forall m_i \in \mathcal{M}_i$ . To compute  $W(m_i|J_i)$  the agent has to be able to evaluate  $p(J'_i|J_i, m_i) \equiv p(\xi'|z'_i, J_i, m_i)p(z'_i|J_i, m_i)$ ,  $\forall J'_i$  in the support of  $(m_i, J_i)$ ,  $m_i \in \mathcal{M}_i$  and  $J_i \subset s \in \mathcal{R}$ . When the required probabilities can be evaluated, the  $W(m_i|J_i)$  calculated need only be "correct" if the support of  $\{p(J'_i|J_i, m_i)\}$  is in the recurrent class.

Consider a capital accumulation game in which the investment component of  $m_i$ , say,  $m_{I,i}$  is not observed but the pricing component, say,  $m_{P,i}$  is observed, and assume prices are set before the outcome of the current investments is known. If  $z_i$  represent costs which is private information then  $p(\xi'|J_i, m_i) = p(\xi'|J_i, m_{P,i})$ . Assume also that  $\{z\}$  evolves as a controlled Markov process, so that  $p(z'_i|J_i, m_i) = p(z'_i|z_i, m_{I,i})$ , and is known from the primitives of the cost-reducing process. Since costs are not observed and are a determinant of prices, past prices are informationally relevant (they contain information on current costs).

14. Note that even if agents can access the required information, to evaluate actions in the way assumed in a restricted EBE they will have to incur the cost of storing additional information and making additional computations—a cost we return to in the context of the computational algorithm discussed in the next section.

15. Whether  $\pi^E(m_i \neq m_i^*, J_i)$  can be consistently estimated depends on the specifics of the problem, but it frequently can be. For a simple example consider an investment game where the profit function is additively separable in the cost of investment or  $m_i$ , so that  $\pi^E(m_i, J_i) = \pi^E(m_i^*, J_i) + m_i^* - m_i$ . If profits are not additively separable in  $m_i$  but  $m_i$  is observed then it suffices that agents be able to compute profits as a function of  $(J_i, m_i, m_{-i})$ , as in the computational example below and in differentiated product markets in which the source of asymmetric information is costs, equilibrium is Nash in prices, and agents know the demand function. In auctions the agent can compute  $\pi^E(m_i, J_i)$  if the agent can learn the distribution of the winning bid.

In this model  $p(J'_i | J_i, m_i) = p(\xi'_i | J_i, m_{P,i})p(z'_i | z_i, m_{I,i})$ . Since  $\xi'_i$  is set by the firm's decision on  $m_{P,i}$  and  $p(z'_i | z_i, m_{I,i})$  is known, the agent will always be able to evaluate  $W(m_i | J_i), \forall m_i \in \mathcal{M}_i$ . If  $m_{P,i} = m_{P,i}^*$  then these evaluations will be correct if the support  $z'_i$  given  $(z_i, m_{I,i})$  is in the support of  $(z_i, m_{I,i}^*)$ , since then all  $J'$  with positive probability will be in the recurrent class. If the support condition is met but  $m_{P,i} \neq m_{P,i}^*$  then  $W(m_{I,i}, m_{P,i} \neq m_{P,i}^* | J_i)$  will be correct if there is a  $(\tilde{z}_i, \xi) \subset s \in \mathcal{R}$  with the property that the optimal price at that point is  $m_{P,i}$ , i.e.  $m_{P,i}^*(\tilde{z}_i, \xi) = m_{P,i}$ .<sup>16</sup>

### III.B. The Finite State Space Condition

Our framework is restricted to finite state games. We now consider this restriction in more detail. We have already assumed that there was (1) an upper bound to the number of firms simultaneously active, and (2) each firm's physical states (our  $\omega$ ) could only take on a finite set of values. These restrictions ensure that the payoff relevant random variables are finite dimensional, but they do not guarantee this for the informationally relevant random variables, so optimal strategies could still depend on an infinite history.<sup>17</sup> We can ensure that the informationally relevant random variables are finite dimensional either through restrictions on the form of the game, or by imposing constraints on the cognitive abilities of the decision makers.

One example of a game form that can result in a finite dimensional space for the informationally relevant state variables is when there is periodic simultaneous revelation of all variables which are payoff relevant to all agents. Claim 1 of Appendix 1 shows that in this case an equilibrium with strategies restricted to depend on only a finite history is an equilibrium to the game with unrestricted strategies. Claim 2 of Appendix 1 shows that there is indeed a restricted EBE for the game with periodic revelation of information. The numerical analysis in Section IV includes an example in which regulation generates such a structure. Periodic revelation of all information can also result from

16. If the agents did not know the form of the underlying controlled Markov process a priori, it may be estimable using the data generated by the equilibrium process.

17. The conditions would however ensure finiteness in a game with asymmetric information where the sources of asymmetric information are distributed independently over time (as in Bajari, Benkard, and Levin 2007, or Pakes, Ostrovsky, and Berry 2007).

situations in which private information can seep out of firms (say, through labor mobility) and will periodically do so for all firms at the same time, or when the equilibrium has one state which is visited repeatedly at which the states of all players are revealed.

There are other game forms which ensure finiteness. One example is when the institutional structure ensures that each agent only has access to a finite history. For example, consider a sequence of Internet auctions, one every period, for different units of a particular product. Potential bidders enter the auction site randomly and can only bid at finite increments. Their valuation of the object is private information, and the only additional information they observe are the sequence of prices that the product sold at while the bidder was online. If, with probability 1, no bidder remains on the site for more than  $L$  auctions, prices more than  $L$  auctions in the past are not in any bidder's information set, and hence cannot effect bids.<sup>18</sup> Alternatively a combination of assumptions on the functional forms for the primitives of the problem and the form of the interactions in the market that yield finite dimensional sufficient statistics for all unknown variables could also generate our finite state space condition.

A different way to ensure finiteness is through bounded cognitive abilities, say, through a direct bound on memory (e.g., agents cannot remember what occurred more than a finite number of periods prior), or through bounds on complexity, or perceptions. There are a number of reasons such a restriction may be appealing to empirical researchers. First it might be thought to be a realistic approximation to the actual institutions in the market. Second in most applications the available data is truncated so the researcher does not have too long a history to condition on. Moreover in any given application one could investigate the extent to which policies or outcomes depended on particular variables either empirically or computationally.

To illustrate, our computational example computes equilibria to finite state games generated by both types of assumptions. One of the questions we address is whether the different assumptions we use to obtain finiteness, all of which seem a priori reasonable, generate equilibria with noticeably different policies.

18. Formally this example requires an extension of our framework to allow for state variables that are known to two or more, but not to all, agents.

## IV. AN ALGORITHM FOR COMPUTING AN EBE

This section shows that we can use a reinforcement learning algorithm to compute an EBE. As a result our equilibria can be motivated as the outcome of a learning process. In the reinforcement learning algorithm players form expectations on the value that is likely to result from the different actions available to them and choose their actions optimally given those expectations. From a given state, those actions, together with realizations of random variables whose distributions are determined by them, lead to a current profit and a new state. Players use this profit together with their expectations of the value they assign to the new state to update their expectation of the continuation values from the starting state. They then proceed to choose an optimal policy for the new state, a policy that maximizes its expectations of the values from that state. This process continues iteratively.

Note that the players' evaluations at any iteration need not be correct. However we would expect that if policies converge and we visit a point repeatedly we will eventually learn the correct continuation value of the outcomes from the policies at that point. Our computational mimic of this process includes a test of whether our equilibrium conditions, conditions that ensure that continuation evaluations are in fact consistent with subsequent play, are satisfied. We note that since our algorithm is a simple reinforcement learning algorithm, an alternative approach would have been to view the algorithm itself as the way players learn the values needed to choose their policies, and justify the output of the algorithm in that way. A reader who subscribes to the latter approach may be less interested in the testing subsection.<sup>19</sup>

We begin with the iterative algorithm for an EBE, then note the modifications required for a restricted EBE, and then move on to the test statistic for both equilibrium concepts. A discussion of the properties of the algorithm, together with its relationship to the previous literature and additional details that can make implementation easier, is deferred until Appendix 2.

The algorithm consists of an iterative procedure and subroutines for calculating initial values and profits. We begin with the

19. On the other hand, there are several issues that arise were one to take the learning approach as an approximation to behavior, among them the question of whether (and how) an agent can learn from the experience of other agents, and how much information an agent gains about its value in a particular state from experience in related states.

iterative procedure. Each iteration, indexed by  $k$ , starts with a location that is a state of the game (the information sets of the players) say,  $L^k = [J_1^k, \dots, J_{n(k)}^k]$ , and the objects in memory, say,  $M^k = \{M^k(J) : J \in \mathcal{J}\}$ . The iteration updates both these objects. We start with the updates for an unrestricted EBE, and then come back to how the iterative procedure is modified when computing a restricted EBE. The rule for when to stop the iterations consists of a test of whether the equilibrium conditions defined in the last section are satisfied, and we describe the test immediately after presenting the iterative scheme.

*Memory.* The elements of  $M^k(J)$  specify the objects in memory at iteration  $k$  for information set  $J$ , and hence the memory requirements of the algorithm. Often there will be more than one way to structure the memory with different ways having different advantages. Here we focus on a simple structure that will always be available (though not necessarily always efficient); alternatives are considered in Appendix 2.

$M^k(J_i)$  contains

- a counter,  $h^k(J_i)$ , which keeps track of the number of times we have visited  $J_i$  prior to iteration  $k$ , and if  $h^k(J_i) > 0$  it contains
- $W^k(m_i|J_i)$  for  $m_i \in \mathcal{M}_i, i = 1, \dots, n$ .

If  $h^k(J_i) = 0$  there is nothing in memory at location  $J_i$ . If we require  $W(\cdot|J_i)$  at a  $J_i$  at which  $h^k(J_i) = 0$  we have an initiation procedure that sets  $W^k(m_i|J_i) = W^0(m_i|J_i)$ . Appendix 2 considers choices of  $\{W^0(\cdot)\}$ . For now we simply note that high initial values tend to ensure that all policies will be explored.

*Policies and Random Draws for Iteration k.* For each  $J_i^k$  which is a component of  $L^k$  call up  $W^k(\cdot|J_i^k)$  from memory and choose  $m^k(J_i^k)$  to

$$\max_{m \in \mathcal{M}_i} W^k(m|J_i^k).$$

With this  $\{m^k(J_i^k)\}$  use equation (1) to calculate the realization of profits for each active agent at iteration  $k$  (if  $d$  is random, then the algorithm has to take a random draw on it before calculating profits). These same policies,  $\{m^k(J_i^k)\}$ , are then substituted into the conditioning sets for the distributions of the next period's state variables (the distributions in equation (2) for payoff relevant random variables and the update of informationally relevant

state variables if the action causes such an update), and they, in conjunction with the information in memory at  $L^k$ , determine a distribution for future states (for  $\{J_i^{k+1}\}$ ). A pseudo random number generator is then used to obtain a draw on the next period's payoff relevant states.

*Updating.* Use  $(J_i^k, m^k(J_i^k), \omega_i^{k+1}, d^{k+1})$  to obtain the updated location of the algorithm

$$L^{k+1} = [J_1^{k+1}, \dots, J_{n(k+1)}^{k+1}].$$

To update the  $W$  it is helpful to define a “perceived realization” of the value of play at iteration  $k$  (that is, the perceived value after profits and the random draws are realized), or

$$(5) \quad V^{k+1}(J_i^k) = \pi(\omega_i^k, \omega_{-i}^k, m_i^k, m_{-i}^k, d^k) + \max_{m \in \mathcal{M}_i} W^k(m | J_i^{k+1}).$$

To calculate  $V^{k+1}(J_i^k)$  we need to first find and call up the information in memory at locations  $\{J_i^{k+1}\}_{i=1}^{n_{k+1}}$ .<sup>20</sup> Once these locations are found we keep a pointer to them, as we will return to them in the next iteration.

For the intuition behind the update for  $W^k(\cdot | J_i^k)$  note that were we to substitute the *equilibrium*  $W^*(\cdot | J_i^{k+1})$  and  $\pi^E(\cdot | J_i^k)$  for the  $W^k(\cdot | J_i^{k+1})$  and  $\pi^k(\cdot | J_i^k)$  in equation (5) and use equilibrium policies to calculate expectations, then  $W^*(\cdot | J_i^k)$  would be the expectation of  $V^*(\cdot | J_i^k)$ . Consequently we treat  $V^{k+1}(J_i^k)$  as a random draw from the integral determining  $W^*(\cdot | J_i^k)$  and update the value of  $W^k(\cdot | J_i^k)$  as we do an average, for example,

$$(6) \quad W^{k+1}(m_i^k | J_i^k) = \frac{1}{h^k(J_i^k)} V^{k+1}(J_i^k) + \left( \frac{h^k(J_i^k) - 1}{h^k(J_i^k)} \right) W^k(m_i^k | J_i^k),$$

where  $m_i^k$  is the policy perceived to be optimal for agent  $i$  at iteration  $k$ . This makes  $W^k(J_i^k)$  the simple average of the  $V^r(J_i^r)$  over the iterations at which  $J_i^r = J_i^k$ . Though use of this simple average will satisfy Robbins and Monroe's (1951) convergence conditions, we will typically be able to improve the precision of our estimates of the  $W(\cdot)$  by using a weighting scheme that

20. The burden of the search for these states depends on how the memory is structured, and the efficiency of the alternative possibilities depend on the properties of the problem analyzed. As a result we come back to this question when discussing our example.

downweights the early values of  $V^r(\cdot)$  as they are estimated with more error than the later values.<sup>21</sup>

*Completing the Iteration.* We now replace the  $W^k(\cdot|J_i^k)$  in memory at location  $J_i^k$  with  $W^{k+1}(\cdot|J_i^k)$  (for  $i = 1, \dots, n_k$ ) and use the pointers obtained above to find the information stored in memory at  $L^{k+1}$ . This completes the iteration as we are now ready to compute policies for the next iteration. The iterative process is periodically stopped to run a test of whether the policies and values the algorithm outputs are equilibrium policies and values. We come back to that test presently.

*Updating When Computing a Restricted EBE.* The algorithm just described only updates  $W^k(m_i|J_i)$  for  $m_i = m_i^k$ , the policy that is optimal given iteration  $k$ 's evaluations. So this algorithm is unlikely to provide correct evaluations of outcomes from actions off the equilibrium path, and a restricted EBE requires correct evaluations of some of those outcomes (the outcomes in  $\mathcal{R}$ ). To compute a restricted EBE we modify this algorithm to update all the  $\{W^k(m|J_i^k)\}_{m \in \mathcal{M}_i}$ , that is, the continuation values for *all possible* actions from a state whenever that state is reached. This ensures that whenever a nonequilibrium action has a possible outcome that is in the recurrent class, it will be evaluated correctly provided all recurrent class points are evaluated correctly.

To update  $W^k(m_i|J_i^k)$  when  $m_i \neq m_i^k$  we take a random draw from the distribution of outcomes conditional on that  $m_i$ , use it and the random draws from the competing agent's *optimal* policies to form what the perceived value realization would have been had the agent implemented policy  $m_i \neq m_i^k$  (substitute  $m_i$  for  $m_i^k$  in the definition  $V^{k+1}(J_i^k)$  in equation (5)), and use it to form  $W^{k+1}(m_i|J_i^k)$  (as in equation (6)). The rest of the algorithm is as above; in particular we update the location using the draws from

21. One simple and surprisingly effective way of doing so is to restart the algorithm using as starting values the values outputted from the first several million draws. The Robbins and Monroe (1951) article is often considered to have initiated the stochastic approximation literature of which reinforcement learning is a special case. Their conditions on the weighting function are that the sum of the weights of each point visited infinitely often must increase without bound while the sum of the weights squared must remain bounded.

the optimal policy. Note that the algorithm to compute a restricted EBE is significantly more computationally burdensome than that for the unrestricted EBE (the computational burden at each point goes up by a factor of  $\times_{i=1}^{n_k} \# \frac{M_i}{n_k}$ ), and is likely to also increase the memory requirements.

#### IV.A. Testing Whether the Output of the Algorithm Constitutes an EBE or a Restricted EBE

Assume we have a  $W$  vector in memory at some iteration of the algorithm, say,  $W^k = \tilde{W}$ , and we want to test whether  $\tilde{W}$  generates an EBE on a recurrent subset of  $S$ . To perform the test we need to check our equilibrium conditions and this requires: (1) a candidate for a recurrent subset determined by  $\tilde{W}$ , say,  $\mathcal{R}(\tilde{W})$ , and checks for both, (2) the optimality of policies and, (3) the consistency of  $\tilde{W}$ , on  $\mathcal{R}(\tilde{W})$ .

To obtain a candidate for  $\mathcal{R}(\tilde{W})$ , start at any  $s^0$  and use the policies implied by  $\tilde{W}$  to simulate a sample path  $\{s^j\}_{j=1}^{J_1+J_2}$ . Let  $\mathcal{R}(J_1, J_2, \cdot)$  be the set of states visited at least once between  $j = J_1$  and  $j = J_2$ . Provided  $J_1, J_2$ , and  $J_1 - J_2$  grow large,  $\mathcal{R}$  will become a recurrent class of the process generated by  $\tilde{W}$ . In practice to determine whether any finite  $(J_1, J_2)$  are large enough, one generates a second sample path starting at  $J_2$  and continuing for another  $J_2 - J_1$  iterations. We then check to see that the set of points visited on the second sample path are the same as those in  $\mathcal{R}(J_1, J_2, \cdot)$ .

The second equilibrium condition specifies that the policies must be optimal given  $\tilde{W}$ . This is satisfied by construction as we chose the policies that maximize  $\tilde{W}(m_i | J_i)$  at each  $J_i$ .

To check the third equilibrium condition we have to check for the consistency of  $\tilde{W}$  with outcomes from the policies generated by  $\tilde{W}$  on the points in  $\mathcal{R}$ . Formally we have to check for the equality in

$$\tilde{W}(m_i^* | J_i) = \pi^E(m_i^*, J_i) + \beta \sum_{J'_i} \tilde{W}(m^*(J'_i) | J'_i) p^e(J'_i | J_i).$$

In principle we could check this by direct summation for the points in  $\mathcal{R}$ . However this is computationally burdensome, and the burden increases exponentially with the number of possible states (generating a curse of dimensionality). So proceeding in this way would limit the types of empirical problems that could be analyzed.

A far less burdensome alternative, and one that does not involve a curse of dimensionality, is to use simulated sample paths for the test. To do this we start at an  $s_0 \in \mathcal{R}$  and forward simulate. Each time we visit a state we compute perceived values, the  $V^{k+1}(\cdot)$  in equation (5), for each  $J_i$  at that state, and keep track of the average and the sample variance of those simulated perceived values across visits to the same state, say,

$$\{(\hat{\mu}(W(m^*(J_i)|J_i)), \hat{\sigma}^2(W(m^*(J_i)|J_i)))\}_{J_i \subset s, s \in \mathcal{R}}.$$

An estimate of the mean square error of  $\hat{\mu}(\cdot)$  as an estimate of  $\tilde{W}(\cdot)$  can be computed as  $(\hat{\mu}(\cdot) - \tilde{W})^2$ . The difference between this mean square error and the sampling variance, or  $\hat{\sigma}^2(W(m^*(J_i)|J_i))$ , is an unbiased estimate of the bias squared of  $\hat{\mu}(\cdot)$  as an estimate of  $\tilde{W}(\cdot)$ . We base our test of the third EBE condition on these bias estimates.

More formally if we let  $E(\cdot)$  take expectations over simulated random draws,  $l$  index information sets, and do all computations as percentages for each  $\tilde{W}_l(\cdot)$  value, the expectation of our estimate of the percentage mean square of  $\hat{\mu}(W_l)$  as an estimate of  $\tilde{W}_l$  is

$$(7) \quad \begin{aligned} MSE_l &\equiv E[\widehat{MSE}_l] \equiv E\left(\frac{\hat{\mu}(W_l) - \tilde{W}_l}{\tilde{W}_l}\right)^2 = \\ &E\left(\frac{\hat{\mu}(W_l) - E[\hat{\mu}(W_l)]}{\tilde{W}_l}\right)^2 + \left(\frac{E[\hat{\mu}(W_l)] - \tilde{W}_l}{\tilde{W}_l}\right)^2 \equiv \sigma_l^2 + (Bias_l)^2. \end{aligned}$$

Let  $(\widehat{MSE}_s, \sigma_s^2, (Bias_s)^2)$  be the average of  $(\widehat{MSE}_l, \sigma_l^2, (Bias_l)^2)$  over the information sets (the  $l$ ) of the agents active at state  $s$ , and  $\hat{\sigma}_s^2$  be the analogous average of  $\hat{\sigma}^2(\frac{W_l}{W^2})$ . Then since  $\hat{\sigma}_s^2$  is an unbiased estimate of  $\sigma_s^2$ , the law of large numbers ensures that an average of the  $\hat{\sigma}_s^2$  at different  $s$  converges to the same average of  $\sigma_s^2$ . Let  $h_s$  be the number of times we visit point  $s$ . We use as our test statistic, say,  $\mathcal{T}$ , an  $h_s$  weighted average of the difference between the estimates of the mean square and that of the variance, so if  $\rightarrow$  indicates (almost sure) convergence, the argument above implies that

$$(8) \quad \mathcal{T} \equiv \sum_s h_s \widehat{MSE}_s - \sum_s h_s \hat{\sigma}_s^2 \rightarrow \sum_s h_s (Bias_s)^2,$$

a weighted average of the sum of squares of the percentage bias. If  $\mathcal{T}$  is sufficiently small we stop the algorithm; otherwise we continue.<sup>22</sup>

*Testing for a Restricted EBE.* Our test for a restricted EBE is similar except that in the restricted case we simulate the mean and the variance of outcomes for every  $m_i \in \mathcal{M}_i$  for each information set  $l$ , say,  $(\hat{\mu}_{m_i, l}, \hat{\sigma}_{m_i, l}^2)$ , for each  $J_l \subset s$  and  $s \in \mathcal{R}$ . We then use the analogue of equation (7) to derive estimates of  $\{\widehat{MSE}_{l, m_i}\}$  and average over  $m_i \in \mathcal{M}_i$  to obtain new estimates of  $(\widehat{MSE}_l, \hat{\sigma}_l^2)$ . The test statistic is obtained by substituting these new estimates into the formula for  $\mathcal{T}$  in equation (8) and will be labeled  $\mathcal{T}_R$ .

## V. EXAMPLE: MAINTENANCE DECISIONS IN AN ELECTRICITY MARKET

The restructuring of electricity markets has focused attention on the design of markets for electricity generation. One issue in this literature is whether the market design would allow generators to make super-normal profits during periods of high demand. In particular the worry is that the facts that currently electricity is not storable and has extremely inelastic demand might lead to sharp price increases in periods of high demand (for a review of the literature on price hikes and an empirical analysis of their sources in California during the summer of 2000, see Borenstein, Bushnell, and Wolak 2002). The analysis of the sources of price increases during periods of high demand typically conditions on whether generators are bid into or withheld from the market, though some of the literature have tried to incorporate the possibility of “forced,” in contrast to “scheduled,” outages (see Borenstein, Bushnell, and Wolak 2002). Scheduled outages are largely for maintenance and maintenance decisions are difficult to incorporate into an equilibrium analysis because, as many authors have noted, they are endogenous.<sup>23</sup>

22. Formally  $\mathcal{T}$  is an  $L^2(\mathcal{P}_{\mathcal{R}})$  norm in the percentage bias, where  $\mathcal{P}_{\mathcal{R}}$  is the invariant measure associated with  $(\mathcal{R}, \bar{W})$ . Appendix 2 comments on alternative possible testing procedures, some of which may be more powerful than the test provided here.

23. There has, however, been an extensive empirical literature on when firms do maintenance (see, for example, Harvey, Hogan, and Schatzki 2004 and the literature reviewed there). Of particular interest are empirical investigations of

Since the benefits from incurring maintenance costs today depend on the returns from bidding the generator in the future, and the latter depend on what the firms' competitors bid at future dates, an equilibrium framework for analyzing maintenance decisions requires a dynamic game with strategic interaction. To the best of our knowledge maintenance decisions of electric utilities have not been analyzed within such a framework to date. Here we provide the details of a simple example that endogenizes maintenance decisions and then compute a restricted EBE for that example.

*Overview of the Model.* In our model the level of costs of a generator evolve on a discrete space in a nondecreasing random way until a maintenance decision is made. In the full information model each firm knows the current cost state of its own generators as well as those of its competitors. In the model with asymmetric information the firm knows the cost position of its own generators, but not those of its competitors.

In any given period firms can hold their generators off the market. Whether they do so is public information. They can, but need not, use the period they are shut down to do maintenance. If they do maintenance, the cost level of the generator reverts to a base state (to be designated as the zero state). If they do not do maintenance the cost state of the generator is unchanged. In the asymmetric information model whether a firm maintains a generator that is not bid into the market is private information.

If they bid the generator into the market, they submit a supply function and compete in the market. If the generator is bid in and operated its costs are incremented by a stochastic shock. There is a regulatory rule ensuring that the firms do maintenance on each of their generators at least once every six periods.

For simplicity we assume that if a firm submits a bid function for producing electricity from a given generator, it always submits the same function (so in the asymmetric information environment the only cost signals sent by the firm is whether it bids in each of its generators). We do, however, allow for heterogeneity in both cost and bidding functions across generators. In particular we allow for one firm that owns only big generators, Firm B, and one firm that only owns small generators, Firm S. Doing

---

the co-ordination of maintenance decisions, see, for example, Patrick and Wolak (1997).

maintenance on a large generator and then starting it up is more costly than doing maintenance on a small generator and starting it up, but once operating the large generator operates at a lower marginal cost. The demand function facing the industry distinguishes between the five days of the work week and the two-day weekend, with demand higher in the work week.

In the full information case the firm's strategy are a function of; the cost positions of its own generators, those of its competitors, and the day of the week. In the asymmetric information case the firm does not know the cost position of its competitors' generators, though it does realize that its competitors' strategy will depend on those costs. As a result any variable that helps predict the costs of a competitors' generators will be informationally relevant.

In the asymmetric information model Firm B's perceptions of the cost states of Firm S's generators will depend on the last time each of Firm S's generators shut down. So the time of the last shutdown decision on each of Firm S's generators are informationally relevant for Firm B. Firm S's last shutdown and maintenance decisions depended on what it thought Firm B's cost states were at the time those decisions were made, and hence on the timing of Firm B's prior shutdown decisions. Consequently Firm B's last shutdown decisions will generally be informationally relevant for itself. As noted in the theory section, without further restrictions this recurrence relationship between one firm's actions at a point in time and the prior actions of the firm's competitors at that time can make the entire past history of shutdown decisions of both firms informationally relevant. Below we consider alternative restrictions each of which have the effect of truncating the relevant past history in a different way.

*Social Planner and Full Information Problem.* To facilitate efficiency comparisons we also present the results generated by the same primitives when (1) maintenance decisions are made by a social planner that knows the cost states of all generators, and (2) a duopoly in which both firms have access to the cost states of all generators (their own as well as their competitors, our "full information" problem). The planner maximizes the sum of the discounted value of consumer surplus and net cash flows to the firms. However since we want to compare maintenance decisions holding other aspects of the environment constant, when the planner decides to bid a generator into the market, we

constrain it to use the same bidding functions used in the competitive environments.

Since the social planner problem is a single-agent problem, we compute it using a standard contraction mapping. The equilibrium concept for the full information duopoly is Markov perfect and an equilibrium can be computed for it using techniques analogous to those used for the asymmetric information duopoly (see Pakes and McGuire 2001).

#### V.A. Details and Parameterization of the Model

Firm B has two generators at its disposal. Each of them can produce up to 25 megawatts of electricity at a constant marginal cost which depends on their cost state ( $mc_B(\omega)$ ) and can produce higher levels of electricity at increasing marginal cost. Firm S has three generators at its disposal each of which can produce 15 megawatts of electricity at a constant marginal cost which depends on their cost state ( $mc_S(\omega)$ ) and higher levels at increasing marginal cost. So the marginal cost function of a generator of type  $k \in \{B, S\}$  is as follows:

$$\begin{aligned} MC_k(\omega) &= mc_k(\omega) & q < \bar{q}_k \\ &= mc_k(\omega) + \beta(q - \bar{q}_k) & q \geq \bar{q}_k \end{aligned}$$

where  $\bar{q}_B = 25$  and  $\bar{q}_S = 15$  and the slope parameter  $\beta = 10$ . For a given  $\omega$  and level of production, Firm B's generator's marginal cost is smaller than those of Firm S at any cost state, but the cost of maintaining and restarting Firm B's generators is two and a half times that of Firm S's generators (see Table I).

The firms bid just prior to the production period and they know the cost of their own generators before they bid. If a generator is bid, it bids a supply curve that is identical to its highest marginal cost at which it can operate. The market supply curve is obtained by the horizontal summation of the individual supply curves. For the parameter values indicated in Table I, if Firm B bids in  $N_b$  number of generators and Firm S bids in  $N_s$  number of generators, the resultant market supply curve is:

$$Q^{MS}(N_b, N_s) = \begin{cases} 0 & p < 100 \\ 25N_b + \left(\frac{p-100}{\beta}\right)N_b & 100 \leq p < 170 \\ 25N_b + \left(\frac{p-100}{\beta}\right)N_b + 15N_s + \left(\frac{p-170}{\beta}\right)N_b & 170 \leq p \leq 600, \end{cases}$$

TABLE I  
PRIMITIVES THAT DIFFER AMONG FIRMS

| Parameter                                           | Firm B         | Firm S           |
|-----------------------------------------------------|----------------|------------------|
| Number of generators                                | 2              | 3                |
| Range of $\omega$                                   | 0–4            | 0–4              |
| Marginal cost constant ( $\omega = (0, 1, 2, 3)$ )* | (20,60,80,100) | (50,100,130,170) |
| Maximum capacity at constant MC                     | 25             | 15               |
| Costs of maintenance                                | 5,000          | 2,000            |

\* At  $\omega = 4$  the generator must shut down.

and supply is infinitely elastic at  $p=600$ . The \$600 price cap is meant to mimic the ability of the independent system operator to import electricity when local market prices are too high.

The market maker runs a uniform price auction; it horizontally sums the generators' bid functions and intersects the resultant aggregate supply curve with the demand curve. This determines the price per megawatt hour and the quantities the two firms are told to produce. The market maker then allocates production across generators in accordance with the bid functions and the equilibrium price.

The demand curve is log-linear

$$\log(Q)^{MD} = D_d - \alpha \log(p),$$

with a price elasticity of  $\alpha = .3$ . In our base case the intercept term  $D_{d=weekday} = 7$  and  $D_{d=weekend} = 6.25$ . We later compare this to a case where demand is lower,  $D_{d=weekday} = 5.3$  and  $D_{d=weekend} = 5.05$ , as we found different behavioral patterns when the ratio of production capacity to demand was higher.

As noted if the generator does maintenance then it can be operated in the next period at the low cost base state ( $\omega = 0$ ). If the generator is shut down but does not do maintenance its cost state does not change during the period. If the generator is operated the state of the generator stochastically decays. Formally if  $\omega_{i,j,t} \in \Omega = \{0, 1, \dots, 4\}$  is the cost state of firm  $i$ 's  $j$ th generator and it is operated in period  $t$ , then

$$\omega_{i,j,t+1} = \omega_{i,j,t} - \eta_{i,j,t},$$

where  $\eta_{i,j,t} \in \{0, 1\}$  with each outcome having probability .5.

The information at the firm's disposal when it makes its shutdown and maintenance decisions, say,  $J_{i,t}$ , always includes the vector of states of its own generators, say,  $\omega_{i,t} = \{\omega_{i,j,t}; j = 1 \dots n_i\} \in \Omega^{n_i}$ , and the day of the week (denoted by  $d \in D$ ). In the full information model it also includes the cost states of its competitors' generators. In the asymmetric information case firms do not know their competitors' cost states and so keep in memory public information sources which may help them predict their competitors' actions. The specification for the public information used differs for the different asymmetric information models we run, so we come back to it when we introduce those models.

The strategy of firm  $i \in \{S, B\}$  is a choice of

$$m_i = [m_{1,i}, \dots, m_{n_i,i}] : J_i \rightarrow (0, 1, 2)^{n_i} \equiv M_i,$$

where  $m = 0$  indicates the generator is shut down and not doing maintenance,  $m = 1$  indicates the generator is shut down and doing maintenance, and  $m = 2$  indicates the firm bids the generator into the market. The cost of maintenance is denoted by  $cm_i$ , and if the firm bids into the market the bid function is the highest marginal cost curve for that type of generator. We imposed the constraint that the firm must do maintenance on a generator whose  $\omega = 4$ .

If  $p(m_{1,t}, m_{2,t}, d_t)$  is the market clearing price while  $y_{i,j,t}(m_{B,t}, m_{S,t}, d_t)$  is the output allocated by the market maker to the  $j$ th generator of the  $i$ th firm, the firm's profits ( $\pi_i(\cdot)$ ) are

$$\begin{aligned} \pi_i(m_{B,t}, m_{S,t}, d_t, \omega_{i,t}) &= p(m_{B,t}, m_{S,t}, d_t) \sum_j y_{i,j,t}(m_{B,t}, m_{S,t}, d_t) \\ &- \sum_j [I\{m_{i,j,t} = 2\}c(\omega_{i,j,t}, y_{i,j,t}(m_{B,t}, m_{S,t}, d_t)) - I\{m_{i,j,t} = 1\}cm_{i,j}], \end{aligned}$$

where  $I\{\cdot\}$  is the indicator function which is 1 if the condition inside the brackets is satisfied and 0 elsewhere,  $c(\omega_{i,j,t}, y_{i,j,t}(\cdot))$  is the cost of producing output  $y_{i,j,t}$  at a generator whose cost state is given by  $\omega_{i,j,t}$ , and  $cm_{j,i}$  is the cost of maintenance (our "investment").

### *V.B. Alternative Informational Assumptions for the Asymmetric Information Model*

We have just described the primitives and the payoff relevant random variables of the models we compute. We now consider the

different information sets that we allow the firm to condition on in those models. As noted the public information that is informationally relevant could, in principle, include all past shutdown decisions of all generators—those owned by the firm as well as those owned by the firms' competitors. To apply our framework we have to ensure that the state space is finite. We present the results from three different assumptions on the information structure of the asymmetric information model, each of which have the effect of ensuring finiteness. In addition we compare these results to both a full information model in which all generator's states are public information, and to those generated by a social planner that maximizes the sum of discounted consumer and producer surplus.

All three asymmetric information (AsI) models that we compute assume  $(\omega_{i,t}, d_t) \in J_{i,t}$ . The only factor that differentiates the three is the public information kept in memory to help the firm assess the likely outcomes of its actions. In one case there is periodic full revelation of information; it is assumed that a regulator inspects all generators every  $T$  periods and announces the states of all generators just before period  $T+1$ . In this case we know that if one agent uses strategies that depend only on the information it has accumulated since the states of all generators were revealed, the other agent can do no better than doing so also. We computed the equilibria for this model for  $T=3, 4, 5, 6$  to see the sensitivity of the results to the choice of  $T$ . The other two cases restrict the memory used in the first case; in one a firm partitions the history it uses more finely than in the other. In these cases it may well be that the agents would have profitable deviations if we allowed them to condition their strategies on more information.

The public information kept in memory in the three AsI models is as follows.

- (1) In the model with periodic full revelation of information the public information is the state of all generators at the last date information was revealed, and the shutdown decisions of all generators since that date (since full revelation occurs every  $T$  periods, no more than  $T$  periods of shutdown decisions are ever kept in memory).
- (2) In finite history  $s$  the public information is just the shutdown decisions made in each of the last  $T$  periods on each generator.

- (3) In finite history  $\tau$  the public information is only the time since the last shutdown decision of each generator.

The information kept in memory in each period in the third model is a function of that in the second; a comparison of the results from these two models provides an indication on whether the extra information kept in memory in the second model has any impact on behavior. The first model, with full revelation every six periods, is the only one whose equilibrium is ensured to be an equilibrium to the game where agents can condition their actions on the indefinite past. That is, there may be unexploited profit opportunities when employing the equilibrium strategies of the last two models. On the other hand the cardinality of the state space in the model with periodic full revelation of information is an order of magnitude larger than in either of the other two models.<sup>24</sup>

### V.C. Computational Details

We compute a restricted EBE using the algorithm provided in Section III. The full information (FI) equilibrium is computed using analogous reinforcement learning algorithms (see Pakes and McGuire 2001), and the social planner is computed using a standard iterative technique (as it is a contraction mapping with a small state space). This section describes two model-specific details needed for the computation: (1) starting values for the  $W(\cdot|\cdot)$ 's and the  $\pi^E(\cdot|\cdot)$ , and (2) the information storage procedures.

To ensure experimentation with alternative strategies we used starting values which, for profits, were guaranteed to be higher than their true equilibrium values, and for continuation values, that we were quite sure would be higher. Our initial values for expected profits are the actual profits the agent would receive were its competitor not bidding at all, or

$$\pi_i^{E, k=0}(m_i, J_i) = \pi_i(m_i, m_{-i} = 0, d, \omega_i).$$

For the initial condition for the expected discounted values of outcomes given different strategies we assumed that the profits were the other competitor not producing at all could be

24. However, there is no necessary relationship between the size of the recurrent classes in the alternative models, and as a result no necessary relationship between either the computational burdens or the memory requirements of those models. The memory requirements and computational burdens generated by the different assumptions have to be analyzed numerically.

obtained forever with zero maintenance costs and no depreciation, that is,

$$W^{k=0}(m_i|J_i) = \frac{\pi_i(m_i, m_{-i} = 0, d, \omega_i)}{1 - \beta}.$$

The memory was structured first by public information, and then for each given public information node, by the private information of each agent. We used a tree structure to order the public information and a hash table to allocate the private information conditional on the public information. To keep the memory manageable, every 50 million iterations we performed a “clean-up” operation that dropped all those points that were not visited at all in the last 10 million iterations.

#### *V.D. Computational Properties of the Results*

The results reported below are from runs in which we ran the model 500 million iterations and then printed out test statistics for each firm. The test statistics gave us an  $R^2 \approx 1$  (to five significant digits; for example, for  $T=5$ , the  $R^2$  was .9995 and .9996 for Firms B and S, respectively).<sup>25</sup>

Table II considers the sensitivity of the output from the AsI model with full revelation every  $T$  periods to the choice of  $T$ . As we increase  $T$  the difference in these variables' values becomes progressively smaller, with the difference between  $T=6$  and  $T=5$  not large enough to impact any of our conclusions. Consequently we focus on the  $T=5$  case for all the rest of our calculations.

Next we asked how well we approximate the AsI model with periodic full revelation with our AsI models with restricted state spaces. Table III compares summary statistics from the full revelation model to models in which all a firm remembers about its competitors is (1) whether the competitors' generators were bid into the market in each of the last five periods (the finite history  $s$  information structure), or (2) the last time each of its competitors' generators was shut down (our finite history  $\tau$  model). The table shows that the finite history  $\tau$  information structure does not approximate the periodic full revelation model well, but the finite history  $s$  structure does much better. Indeed, it would be

25. As a check of our programs, we also checked to see that our results from the program that computed the restricted EBE model also constituted an equilibrium for the program that ran the unrestricted EBE model, which it did.

TABLE II  
PERIODIC FULL REVELATION DIFFERENT T

| Summary statistics                            | T = 3 | T = 4 | T = 5 | T = 6 |
|-----------------------------------------------|-------|-------|-------|-------|
| Consumer surplus ( $\times 10^{-3}$ ) 58,000+ | 550   | 572   | 581   | 580   |
| Profit B ( $\times 10^{-3}$ )                 | 393   | 389   | 384   | 383   |
| Profit S ( $\times 10^{-3}$ )                 | 334   | 324   | 322   | 324   |
| Maintenance cost B ( $\times 10^{-3}$ )       | 25.9  | 21.6  | 20.2  | 19.4  |
| Maintenance cost S ( $\times 10^{-3}$ )       | 12.1  | 11.8  | 11.8  | 11.8  |
| Production cost B ( $\times 10^{-3}$ )        | 230.2 | 235.3 | 235.1 | 234.3 |
| Production cost S ( $\times 10^{-3}$ )        | 230.4 | 226.9 | 228.1 | 229.2 |

hard to tell the difference from this and the periodic full revelation model with the kind of data sets we usually have. We use the output from the model with periodic full revelation in our analysis of results in the next subsection, but if we were to compute models with larger state spaces the finite history  $s$  model would become increasingly attractive.

We are interested in the models with the restricted information structures because they generate smaller state spaces and hence are likely to impose less of a computational burden both on the researcher, and perhaps more important, on the agents' actual decision-making process. For both these reasons the restricted information structures may be more suitable for applied work. Table IV provides the sizes of the recurrent classes and compute times for these models (including the test time).<sup>26</sup> We note that the compute times for the AsI models are the compute times for the restricted EBE. If one were to suffice with the weaker notion of an (unrestricted) EBE, the compute time would go down dramatically (e.g., its compute time per 100 million iterations for the periodic full revelation model when  $T=5$  was just under two hours).

The first thing to notice from the table is that the compute time per 100 million iterations increases with the size of the recurrent class, though at a decreasing rate. The size of the

26. All computations in this article were run on the Odyssey cluster supported by the FAS Science Division Research Computing Group at Harvard University. For a description of the machine used see <http://rc.fas.harvard.edu>. However the memory requirements for all runs was well within 1 GB, so the runs could be done on a laptop.

TABLE III  
THREE ASYMMETRIC INFORMATION MODELS (T = 5)

| Summary statistics                      | Finite history |       | Periodic Revelation |
|-----------------------------------------|----------------|-------|---------------------|
|                                         | $\tau$         | $s$   |                     |
| Consumer surplus ( $\times 10^{-3}$ )   | 58,000+        | 270   | 580                 |
| Profit B ( $\times 10^{-3}$ )           | 414            | 384.7 | 384.5               |
| Profit S ( $\times 10^{-3}$ )           | 439            | 323.5 | 322.8               |
| Maintenance cost B ( $\times 10^{-3}$ ) | 28.5           | 20.0  | 20.2                |
| Maintenance cost S ( $\times 10^{-3}$ ) | 18.0           | 11.7  | 11.8                |
| Production cost B ( $\times 10^{-3}$ )  | 226.8          | 235.5 | 235.1               |
| Production cost S ( $\times 10^{-3}$ )  | 254.6          | 228.4 | 228.1               |

TABLE IV  
COMPUTATIONAL COMPARISONS

|                                                                 | AsI; finite hist. $\tau$ | AsI; finite hist. $s$ | AsI; full revel. | Full info. |
|-----------------------------------------------------------------|--------------------------|-----------------------|------------------|------------|
| Compute times per 100 million iterations (hours; includes test) |                          |                       |                  |            |
| Hours                                                           | 3:04                     | 11:08                 | 17:14            | 1:05       |
| Cardinality of recurrent class                                  |                          |                       |                  |            |
| Firm B                                                          | 5650                     | 38,202                | 67,258           | 3,553      |
| Firm S                                                          | 5519                     | 47,304                | 137,489          | 3,553      |

recurrent class for the finite history  $\tau$  model is only 5% of that for the periodic full revelation model, and apparently this is not a rich enough partition of the state space to provide an adequate approximation. The size of the recurrent class from the finite history  $s$  model that does approximate quite well is about 42% of that of the periodic full revelation model. The relative simplicity of the FI model is reflected in its much smaller recurrent class. As a result it computes much quicker than any of the other models.

Finally, to figure out compute times one needs to know how the test statistic behaves as we increase the number of iterations. To ensure that the  $R^2$  statistic was above .99 we needed as much as 200 million iterations, and in all our runs the  $R^2$  flattened out between 250 and 350 million iterations at values that were one to at least four significant digits (see Figure I for an example).

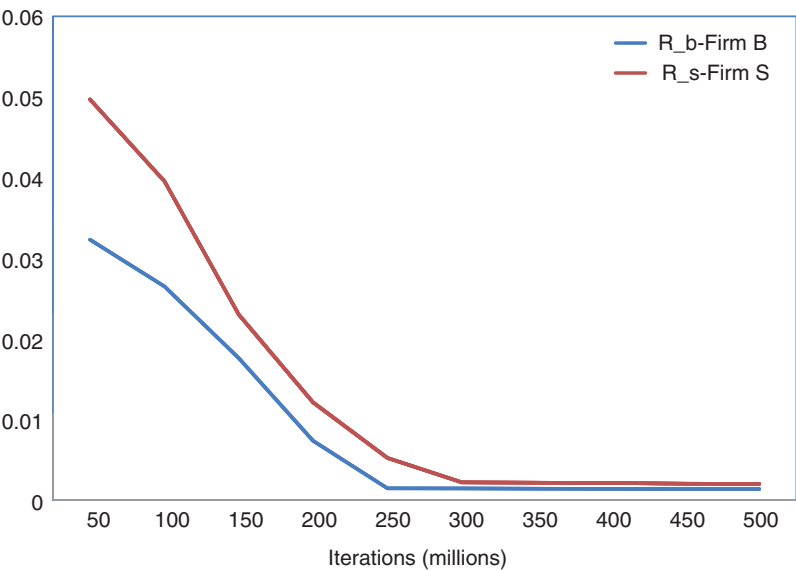


FIGURE I  
Convergence Test Statistics

*V.E. The Economics of the Alternative Environments*

The output of the algorithm includes; strategies, quantities produced and prices by day of the week, realized costs (both operational and maintenance), profits, and consumer welfare. We start with a comparison of the base case AsI model with the base case social planner (the second and third columns in Table V).

*Strategies.* Panel A of Table V is rather striking. The social planner never shuts down without doing maintenance, and does more maintenance on both its big and its small generators than do the AsI competitors. During the week, when demand is high, the planner operates both its large generators and its small generators at almost full capacity. On average it typically does maintenance on one, and sometimes on both, large generators on Sunday, ensuring that those generators are at a low-cost state when going into the work week. The planner typically does maintenance on one small generator on Saturday and another on

TABLE V  
QUANTITIES AND COSTS

|                                                 | Base case |       |       | Excess capacity |       |
|-------------------------------------------------|-----------|-------|-------|-----------------|-------|
|                                                 | Planner   | AsI   | FI    | AsI             | FI    |
| <b>Panel A: Strategies</b>                      |           |       |       |                 |       |
| Firm B: Shutdown and maintenance                |           |       |       |                 |       |
| Shutdown percentage                             | 14.52     | 19.96 | 12.31 | 41.97           | 43.75 |
| Maintenance percentage                          | 14.52     | 10.10 | 10.90 | 6.47            | 6.25  |
| Firm S: Shutdown and maintenance                |           |       |       |                 |       |
| Shutdown Percentage                             | 16.85     | 21.48 | 20.74 | 53.1            | 56.4  |
| Maintenance Percentage                          | 16.85     | 9.83  | 9.91  | 5.22            | 4.84  |
| Firm B: Operating generators by day of the week |           |       |       |                 |       |
| Saturday                                        | 1.41      | 1.08  | 1.72  | 1.03            | 1.0   |
| Sunday                                          | 0.88      | 1.21  | 1.65  | 1.03            | 1.0   |
| Weekday ave.                                    | 1.93      | 1.78  | 1.78  | 1.03            | 1.0   |
| Firm S: Operating generators by day of the week |           |       |       |                 |       |
| Saturday                                        | 1.55      | 1.56  | 2.03  | 1.21            | 0.48  |
| Sunday                                          | 1.89      | 1.75  | 1.86  | 1.20            | 0.44  |
| Weekday ave.                                    | 2.80      | 2.64  | 2.55  | 1.25            | 1.44  |
| <b>Panel B: Costs</b>                           |           |       |       |                 |       |
| Maintenance cost B ( $\times 10^{-3}$ )         | 29        | 20.2  | 21.95 | 12.9            | 12.5  |
| Maintenance cost S ( $\times 10^{-3}$ )         | 20.2      | 11.8  | 11.9  | 6.3             | 5.8   |
| Production cost B ( $\times 10^{-3}$ )          | 211.1     | 235.1 | 240.4 | 48.3            | 48.4  |
| Production cost S ( $\times 10^{-3}$ )          | 174.8     | 228.1 | 215.9 | 13.6            | 11.8  |
| Total cost/quantity                             | 0.389     | 0.452 | 0.444 | 0.290           | 0.282 |
| <b>Panel C: Quantities and prices</b>           |           |       |       |                 |       |
| Average quantity wkend                          | 93.5      | 92.0  | 98.6  | 33.6            | 33.1  |
| Average price wkend                             | 303       | 325   | 260   | 168             | 175.6 |
| Average quantity wkday                          | 185.7     | 181.8 | 181.2 | 42.50           | 42.43 |
| Average price wkday                             | 374       | 401   | 411   | 177             | 177   |

Sunday, and if it requires more maintenance of small generators than that it will maintain two small generators on Saturday.

The AsI equilibrium generates about 30% *more* shutdown of large generators and 25% *more* shutdown of small generators than the social planner, but actually does about 30% *less* maintenance on both types of generators than does the social planner. That is, about half the time generators are not operating in the

AsI equilibrium they are shut down without doing maintenance. The shutdown decision results in higher prices for the firm's operating generator(s). The number of generators operated on weekends by the AsI equilibrium is about the same as the social planner operates, so the AsI equilibrium is operating fewer generators than the social planner on the high-demand weekdays (though it still operates more generators on weekdays than on weekends).

The social planner does more maintenance than the AsI equilibrium generates, and almost all its maintenance is done during the low-demand weekends. This enables the planner to operate more generators on the high-demand weekdays, pushing down price on those days and adding to consumer surplus. The social planner internalizes this increase in consumer surplus, while the firms operating in the AsI equilibrium would not.

*Costs.* The fact that the social planner does more maintenance, and that the planner can optimize maintenance jointly over the large and small generators, results in much lower production costs for the planner than is generated by the AsI equilibrium. Indeed, the planner has lower total (maintenance plus production) costs per unit quantity. This is despite the fact that our model has increasing costs, and the social planner produces more quantity (particularly on the high demand weekdays).

*Prices and Quantities.* Recall that we model an electricity market with relatively inelastic demand. So the fact that the planner produces about 2% more output than the AsI equilibrium on weekdays causes the planners' prices to be about 10% lower on those days. This implies that the AsI equilibrium produces a larger difference in prices between weekdays and weekends than does the social planner. However even the social planner's weekday prices are 20% higher than weekend prices; that is, prices "spike" on high-demand days. Apparently we need to change the institutional setting to get the price discrepancy between weekdays and weekends to under 20%.

*Consumer Surplus and Profits.* It is not a surprise that the planner generates higher total surplus (see Table VI) than does the AsI equilibrium, but it is somewhat surprising that the planner also generates more profits. This is largely because the planner does more maintenance than either of the duopolies, and this reduces total costs.

TABLE VI  
CONSUMER SURPLUS AND PROFITS

|                                               | Base Case |       |       | Excess Capacity |       |
|-----------------------------------------------|-----------|-------|-------|-----------------|-------|
|                                               | Planner   | AsI   | FI    | AsI             | FI    |
| Cons. surplus ( $\times 10^{-3}$ ) 58,000+    | 662       | 581.5 | 595   | 1,316           | 1,311 |
| Total profits (Firm B+S) ( $\times 10^{-3}$ ) | 716.2     | 707.3 | 706.7 | 58.1            | 61.9  |
| Firm B profits ( $\times 10^{-3}$ )           | 385.3     | 384.5 | 388.1 | 53.2            | 54.5  |
| Firm S profits ( $\times 10^{-3}$ )           | 331.0     | 322.9 | 318.8 | 4.9             | 7.4   |
| Total surplus ( $\times 10^{-3}$ ) 590,000+   | 378.1     | 288.9 | 301.4 | 1,374           | 1,373 |

*AsI versus FI Equilibria: Base Case and "Excess" Capacity.*

The comparison of the AsI equilibrium to the FI equilibrium strategies depends on the extent of generating capacity relative to demand. In the base case the FI equilibrium generates less shutdown and more maintenance than does the AsI, but when there is more capacity relative to demand the AsI equilibrium does less shutdown and more maintenance.

The differences are most noticeable in the comparative behavior of the firms during weekends. In the base case the AsI equilibrium generates noticeably less operation of both large and small generators during the weekend than does the FI equilibrium. The weekend shutdowns in the AsI equilibrium enables the firms to signal that their generators will be bid in on the weekdays to follow, and in the base case weekday prices are more than 20% higher than weekend prices. There are no signaling incentives in the FI equilibrium and in that equilibrium more output is produced on weekends. When we increase capacity relative to demand the difference between weekday and weekend prices drops dramatically (to 5.4% in the AsI and 1% in the FI equilibrium) and now both firms operate more on weekends in the AsI equilibrium than in the FI equilibrium (only slightly more for Firm B, but noticeably more for Firm S).

The second noticeable change when we add capacity relative to demand is that the average cost (maintenance plus production cost divided by quantity) is quite a bit lower when there is relatively more capacity. In both capacity environments the average cost in the AsI equilibrium is similar to that in the FI equilibrium, but average costs falls by over 30% when the ratio of capacity to demand increases. Of course firms would have to

weigh any reduction in average costs against the cost of installing the capacity before engaging in capacity-expanding investments, and we cannot compute the private value of capacity expansion without a more complicated dynamic model than the one used here.

Moving to Table VI, we see that the differences in consumer and total surplus between the AsI and the FI equilibrium are always small and differ in sign in the two environments. The major difference between the environment with more capacity relative to demand is that with a higher capacity to demand ratio we see a large increase in consumer surplus and a large (but smaller in absolute value) decrease in producer surplus. As a result total surplus is noticeably larger when the ratio of capacity to demand is higher. This is largely a consequence of prices falling when the ratio of capacity to demand increases, particularly weekday prices. Indeed, it seems that one way to decrease the weekend/weekday price differential is to increase the ratio of capacity to demand.

## VI. CONCLUDING REMARK

We have presented a simple framework for analyzing finite state dynamic games with asymmetric information. It consists of a set of equilibrium conditions which, at least in principle, are empirically testable, and an algorithm capable of computing policies that satisfy those conditions for a given set of primitives. Its advantages are twofold. First by choosing alternative information structures we can approximate behavior by agents in complex institutional settings without requiring those agents to have unrealistically excessive information retention and computational abilities. Second the algorithm we use for analyzing the equilibria is relatively efficient in that it does not require storage and updating of posterior distributions, explicit integration over possible future states to determine continuation values, or storage and updating of information at all possible points in the state space. The hope is that this will enable us to approximate behavior and analyze outcomes in markets which have been difficult to deal with to date. This includes markets with dynamic consumers as well as dynamic producers, and markets where accounting for persistent sources of asymmetric information is crucial to the analysis of outcomes.

## APPENDIX I: CLAIMS FOR PERIODIC REVELATION

CLAIM 1 (Periodic revelation). If for any initial  $s_t \in \mathcal{R}$  there is a  $T^* < \infty$  and a random  $\tau$  (whose distribution may depend on  $s_t$ ) which is less than or equal to  $T^*$  with probability 1, such that all payoff relevant random variables are revealed at  $t + \tau$ , then if we construct an equilibrium to a game whose strategies are restricted to not depend on information revealed more than  $\tau$  periods prior to  $t$ , it is an equilibrium to a game in which strategies are unrestricted functions of the entire history of the game. Moreover there will be optimal strategies for this game which, with probability 1, only take distinct values on a finite state space, so  $\#\mathcal{R}$  is finite. ■

*Sketch of Proof.* Let  $h_{i,t}$  denote the entire history of variables observed by agent  $i$  by time  $t$ , and  $J_{i,t}$  denote that history truncated at the last point in time when all information was revealed. Let  $(W^*(\cdot|J_i), m^*(J_i), p^e(\cdot|J_i))$  be EBE (or restricted EBE) valuations, strategies, and resulting probability distributions when agents condition both their play and their evaluations on  $J_i$  (so they satisfy C1, C2, C3 of Section II). Fix  $J_i = J_{i,t}$ . What we must show is that

$$(W^*(\cdot|J_{i,t}), m^*(J_{i,t}))$$

satisfy C1, C2, C3 if the agents' condition their expectations on  $h_{i,t}$ .

For this it suffices that if the  $*$  strategies are played then for every possible  $(J'_i, J_{-i})$ ,

$$p^e(J'_i|J_{i,t}) = Pr(J'_i|h_{i,t}), \text{ and } p^e(J_{-i}|J_{i,t}) = Pr(J_{-i}|h_{i,t}).$$

If this is the case, strategies which satisfy the optimality conditions with respect to  $\{W^*(\cdot|J_{i,t})\}$  will satisfy the optimality conditions with respect to  $\{W(\cdot|h_{i,t})\}$ , where it is understood that the latter equal the expected discounted value of net cash flows conditional on all history.

We prove the second equality by induction (the proof of the first is similar and simpler). For the initial condition of the inductive argument use the period in which all information is revealed. Then  $p^e(J_{-i}|J_i)$  puts probability 1 at  $J_{-i} = J_{-i,t}$  as does  $Pr(J_{-i}|h_i)$ . For the inductive step, assume  $Pr(J_{-i,t_0}|h_{i,t_0}) = p^e(J_{-i}|J_{i,t_0})$ . What we must show is that if

agents use the  $*$  policies then the distribution of  $J_{-i,t_0+1}$  conditional on  $h_{i,t_0+1}$  depends only on  $J_{i,t_0+1}$ .

Let a bar over a set of variables indicate its complement in  $\cup_i J_{i,t}$  for any  $t$ , and

$$\mu_i \equiv J_{i,t_0+1} \cap \bar{J}_{-i,t_0+1} \cap \bar{J}_{i,t_0}, \quad \text{while} \quad \epsilon \equiv \cap_i J_{i,t+1} \cap \bar{J}_{i,t}$$

so that  $\mu_i$  is the new private, and  $\epsilon$  is the new public, information in  $J_{i,t_0+1}$ . We assume that

$$(A1) \quad \begin{aligned} P(\mu_i | h_{i,t}) &= P(\mu_i | J_{i,t}, m_{i,t}) \text{ and} \\ P(\epsilon | \cup_i h_{i,t}) &= P(\epsilon | \cup_i J_{i,t}, \cup_i m_{i,t}) \end{aligned}$$

so that the distribution of the new private and public information depend only on agents' policies and the information in  $\cup_i J_{i,t}$ . The fact that (A1) allows the distribution of  $\epsilon$  to depend on policies generates the possibility of sending signals or revealing information on events that have occurred since all information was revealed. What (A1) rules out is models where the interpretation of those signals depends on information that occurred prior to the period when all states were revealed.

Since for any events  $(A, B, C)$ ,  $Pr(A|B, C) = \frac{Pr(A, B|C)}{Pr(B|C)}$

$$\begin{aligned} Pr(J_{-i,t_0+1} | h_{i,t_0+1}) &= Pr(\mu_{-i}, \epsilon, J_{-i,t_0} | \mu_i, \epsilon, h_{-i,t_0}) \\ &= \frac{Pr(\mu_{-i}, \mu_i, \epsilon, J_{-i,t_0} | h_{i,t_0})}{Pr(\mu_i, \epsilon | h_{i,t_0})}. \end{aligned}$$

From (A1) and the  $*$  policies, the numerator in this expression can be rewritten as

$$\begin{aligned} &Pr(\mu_{-i}, \mu_i, \epsilon, J_{-i,t_0} | h_{i,t_0}) \\ &= Pr(\mu_{-i}, \mu_i, \epsilon, J_{-i,t_0} | \cup_i J_{i,t_0}, \cup_i m^*(J_{i,t_0})) Pr(J_{-i,t_0} | h_{i,t_0}), \end{aligned}$$

and from the hypothesis of the inductive argument  $Pr(J_{-i,t_0} | h_{i,t_0}) = p^e(J_{-i,t_0} | J_{i,t_0})$ . A similar calculation for the denominator concludes the proof. ■

**CLAIM 2.** There exists a restricted EBE if there is periodic revelation of information. ■

*Sketch of a Proof.* In our existence proof we consider only games for which there is a periodic revelation of all

private information. That is, games in which every  $\tau$  periods all private information is revealed. Note that at that period  $J_{i,t} = \xi_t$  and it contains only the payoff relevant variables  $\omega$  for all the firms. Our proof will follow Maskin and Tirole (2001) that showed the existence of Markov perfect equilibrium (MPE) for complete information dynamic game with finite action space with the proper adjustments. Their proof established first that there is an MPE in the finite period game and then used the fact that the game is continuous at infinity to establish the existence of MPE in the infinite horizon game (see Fudenberg and Levine 1983).

To prove existence we must allow for behavioral strategies. In our main setup we consider only pure strategies because we believe that this is the right framework for empirical analysis, but to guarantee the existence of restricted EBE we extend our framework and allow for mixing as well. Having behavioral strategies will not change our setting and our equilibrium conditions much. Each firm will still have the evaluation  $W(m|J_i)$  according to which it determines its optimal actions and our consistency requirement would be the same. We will just allow the firm to mix but we do not need to have a valuation for the mixed strategy, only for the actions  $m$  themselves. Note also that whenever the firm mixes between different  $m$ 's these  $m$ 's should have the same valuations.<sup>27</sup>

Our concept of restricted EBE is relevant for an infinite game. We define a cycle as a game that starts at a particular initial conditions  $\omega_0$ , which is the list of all payoff relevant random variables in that period, and is being played for  $\tau$  periods. A one-cycle game would be an infinite game that starts with the initial conditions  $\omega_0$  being played for one cycle and then starts all over with the same initial condition  $\omega_0$  and continue to repeat itself. We construct a  $T$ -cycle game in the same manner. A game that starts at a particular  $\omega_0$  being played for  $T$  cycles (or for  $T\tau$  periods) and then starts over with the same initial conditions  $\omega_0$  and continues to repeat itself in this manner. Our approach would be to establish a restricted EBE for this  $T$ -cycle game and then let  $T \rightarrow \infty$  and show that the limit would be a restricted EBE equilibrium for our dynamic game.

27. Clearly whenever the firm mixes, all the  $m$ 's in the support of such a mix are chosen with positive probability in the equilibrium play and all resultant outcomes are in the recurrent class.

Consider now a one-cycle game that starts at a particular  $\omega_0$ . This is a finite game and it has a perfect Bayesian equilibrium (see Kreps and Wilson 1982). That is, for this game we can define (possibly mixed) strategies  $m^*(J_{i,t}, \omega_0)$  and beliefs  $B^*(z_{-i,t}|J_{i,t}, \omega_0)$  which are probability distributions over  $z_{-i,t}$  (the types of the other players at period  $t$ ,  $t = 1, \dots, \tau$ ) that specifies the beliefs of player  $i$  regarding the types of other players.<sup>28</sup> The conditions that are satisfied are that  $m^*(J_{i,t}, \omega_0)$  is optimal given  $B^*(z_{-i,t}|J_{i,t}, \omega_0)$  and that the beliefs are consistent with strategies whenever possible. We can now use this PBE to construct a restricted EBE for our one-cycle game. The strategies would be the same strategies as in PBE and the construction of  $W(m|J_{i,t}, \omega_0)$  is straightforward given the equilibrium strategies. Such a construction can be done for every  $\omega_0$ . In a similar way we can construct a restricted EBE for the  $T$ -cycle game, that is, constructing the values  $W(m|J_{i,t}, \omega_0, T)$  and the strategies  $m(J_{i,t}, \omega_0, T)$  where  $t \in \{1, \dots, T\tau\}$  (as afterward the game will replicate itself). Claim 1 implies that we need only ensure that  $J_{i,t}$  includes the values of the payoff relevant random variables at the beginning of the cycle, and all observable variables since that time. Both  $W(m|J_{i,t}, \omega_0, T)$  and  $m(J_{i,t}, \omega_0, T)$  depend on  $T$  which defines the number of cycles we are having before restarting the game.

The last stage is to establish existence of restricted EBE for the infinite horizon game. Let's look at the values and the strategies only of the first cycle (the first  $\tau$  periods). Let's define by  $m^1(J_{i,t}, \omega_0, T)$  as the restricted EBE strategies of the first cycle (the first  $\tau$  periods) when the game is a  $T$ -cycle game and when the starting point is  $\omega_0$ . There is an equilibrium in which these strategies are identical for every first cycle in the  $T$ -cycle game. We now let  $T \rightarrow \infty$  and examine the strategies  $m^1(J_{i,t}, \omega_0, T)$ . We can construct a converging sequence (subsequence if needed) such that  $m^1(J_{i,t}, \omega_0, T) \rightarrow m^\infty(J_{i,t}, \omega_0)$ . We can construct such a converging sequence for every possible  $\omega_0$  and define the strategies  $m^\infty(J_{i,t})$  for the infinite game. We now claim that  $m^\infty(J_{i,t})$  together with the valuation  $W(m|J_{i,t})$  that it generates constitute a restricted EBE.

To do that, we follow similar arguments as in Fudenberg and Levine (1983). First note that that for every  $T'$  there is a

28. Note that  $\omega_0$  is formally part of  $J_{i,t}$  but we write it separately here to indicate the starting point of any cycle.

sufficiently large  $T''$  such that if we look at a  $T$ -cycle game where  $T > T''$  then  $m(J_{i,t}, \omega_0, T)$  would be sufficiently close to  $m^\infty(J_{i,t}, \omega_0)$  for  $t < T''\tau$ . That is, the strategies of the first  $T'$  cycles converge to  $m^\infty(J_{i,t}, \omega_0)$  as  $T \rightarrow \infty$  (note that we can view each of the first  $T'$  cycles as the first cycle). Assume now that  $m^\infty(J_{i,t})$  (together with  $W(m|J_{i,t})$ ) is not a restricted EBE for the infinite game. Then there is a player  $i$  and an information set  $J'_{i,t}$  such that there is a strategy  $m'$  such that  $m' \succ_i m^\infty(J'_{i,t})$ , that is,  $W(m'|J'_{i,t}) > W(m^\infty(J'_{i,t})|J'_{i,t})$  (this is with a slight abuse of notation such that  $W(m'|J_{i,t})$  would be the expected evaluation when player  $i$  plays the [possibly mixed] strategy  $m'$ ). We now claim that if  $W(m'|J'_{i,t}) > W(m^\infty(J'_{i,t})|J'_{i,t})$  then there is a  $T$  and restricted EBE for the  $T$ -cycle game that starts at  $\omega_0$  such that at the first cycle we have  $W(m'|J'_{i,t}, \omega_0, T) > W(m(J'_{i,t}, \omega_0, T)|J'_{i,t}, \omega, T)$  which contradicts the fact that  $m(J'_{i,t}, \omega_0, T)$  is the equilibrium play for the  $T$ -cycle game. This inequality exists because  $W(m'|J'_{i,t}, \omega_0, T)$  (respectively  $W(m(J'_{i,t}, \omega_0, T)|J'_{i,t}, \omega, T)$ ) can be as close as we wish to  $W(m'|J'_{i,t})$  (respectively  $W(m^\infty(J'_{i,t})|J'_{i,t})$ ) when we let  $T$  be sufficiently large (and using the continuity at infinity).

## APPENDIX II: ALGORITHMIC DETAILS

We begin with a brief review of the properties of the algorithm, and then move to some notes on how one might usefully amend the algorithm to be more efficient when different primitives are appropriate.

The advantages of using a stochastic algorithm to compute equilibria in full information games relative to standard iterative techniques like those used in Pakes and McGuire (1994) were explored by Pakes and McGuire (2001). These advantages are even larger in asymmetric information games that use the EBE conditions. This because those conditions do not require us to form beliefs about players' *types*, and the stochastic algorithm neither computes posterior beliefs nor tests for their consistency with the actual distribution of types.

Pakes and McGuire (2001) noted that, at least formally, their stochastic algorithm does away with all aspects of the curse of dimensionality but the one in computing their test statistic. Accordingly as they increased the dimension of the state space

in their examples the computation of the test statistic quickly became the dominant computational burden. We circumvent this problem by substituting simulation for explicit integration in the construction of the test statistic, thereby eliminating the curse of dimensionality entirely.

However as is typical in algorithms designed to compute equilibria for (nonzero sum) dynamic games, there is no guarantee that our algorithm will converge to equilibrium values and policies; that is, all we can do is test whether the algorithm outputs equilibrium values, we can not guarantee convergence to an equilibrium a priori. Moreover there may be more than one equilibria which is consistent with a given set of primitives. There are, however, both choices in implementation and amendments to the algorithm that will influence which equilibrium is computed.

One choice is that for the initial evaluations, that is, our  $W^0$ . High initial values are likely to encourage experimentation and lead to an equilibrium in which players have explored many alternatives. An alternative way of ensuring experimentation is to amend the algorithm as follows. Instead of having agents chose the "greedy" policy at each iteration, that is, the policy that maximizes  $W^k$ , use choice procedure that has an exogenous probability of choosing each possible action at each early iteration, but let that probability go to 0 for all but the greedy policy as the number of iterations grows large. Though both these procedures will insure experimentation, they will also tend to result in longer computational times.

As noted in a particular applied context one may be more interested in directing the algorithm to compute an equilibrium which is consistent with observed data, say, by introducing a penalty function that penalizes deviations from the exogenous information available, than in computing an equilibria that ensures experimentation. Relatedly note that our estimates of the  $\tilde{W}$  are sample averages and will be more accurate at a given location the more times we visit that location. If one is particularly interested in policies and values at a given point, for example, at a point that is consistent with the current data on a given industry, one can increase the accuracy of the relevant estimates by restarting the algorithm repeatedly from that point.

Both the structure of memory provided and the test given in the text are always available, but that memory structure need not be computationally efficient, and the test need not be the most

powerful test. A brief discussion of alternative memory structures and testing procedures follows.

*Alternative Memory Structures.* It is useful to work with the distribution of the increment in  $\omega$  between two periods, that is, defining  $\eta_{t+1} \equiv \omega_{t+1} - \omega_t$ , we work with

$$\mathcal{P}_\eta = \{P_\eta(\cdot|m_i, m_{-i}, \omega); (m_i, m_{-i}) \in \mathcal{M}^n, \omega \in \Omega\},$$

where  $\mathcal{P}_\eta$  is derived from the family of distributions in equation (2).

We begin with the case where  $m$  is observed by the agent's competitors. Then we could hold in memory either estimates of  $W(m|J_i)$  or estimates of  $W(\eta, m|J_i)$ . If the latter we would choose  $m$  at iteration  $k$  to maximize  $\sum_\eta W^k(\eta, m|J_i)p(\eta|m, m_{-i}^{k-1}, \omega)$ . The trade-off here is clear. By holding estimates of  $W(\eta, m)$  instead of estimates of  $W(m)$  in memory, we increase both memory requirements and the number of summations we need to do at each iteration. However, we are likely to decrease the number of iterations needed until convergence, as explicit use of the primitive  $p(\eta|\cdot)$  allows us to integrate out the variance induced by  $\eta$  conditional  $(m, J_i)$  rather than relying on averaging the simulation draws to do so. The  $W(\eta, m|J_i)$  memory structure is particularly easy to use when the probability of  $\eta$  conditional on  $m_i$  is independent of  $m_{-i}$  (i.e., in capital accumulation games), and we used it in our electric utility example.

When  $m$  is unobservable there is an even simpler memory structure that can be used in capital accumulation games. We can then hold in memory estimates of  $W(\eta|J_i)$  and chose  $m$  at iteration  $k$  to maximize  $\sum_\eta W^k(\eta|J_i)p(\eta|m, \omega)$  (we cannot do this when  $m$  is observable because then  $m$  is a signal and will have an effect on next period's state that is independent of  $\eta$ ). Then the memory requirements may be larger when we hold estimates of  $W(m|J_i)$  in memory relative to holding estimates of  $W(\eta|J_i)$ , and will be if the cardinality of the choice set (of  $M$ ) is greater than the cardinality of the the support of the family  $\mathcal{P}_\eta$ . Notice that the model that holds estimates of  $W(\eta|J_i)$  in memory is a natural way of dealing with continuous controls (continuous  $m$ ) whose values are unobserved by competitors, and that we may well have some controls observed and some unobserved, in which case hybrids of the structures introduced above would be possible. As for computational burden, the model that holds estimates of  $W(\eta|J_i)$  in memory has the advantage that it explicitly integrates out over

the uncertainty in  $\eta$  and hence should require fewer iterations until convergence.

*Alternative Testing Procedures.* Several aspects of the test provided in the text can be varied. First the test provided in the text ensures that the  $\bar{W}$  outputted by the algorithm is consistent with the distribution of *current* profits and the discounted evaluations of the *next period's* state. We could have considered a test based on the distribution of discounted profits over  $\tau$  periods and the discounted evaluation of states reached in the  $\tau$ th period. We chose  $\tau = 1$  because it generates the stochastic analogue of the test traditionally used in iterative procedures to determine whether we have converged to a fixed point. It may well be that a different  $\tau$  provides a more discerning test, and with our testing algorithm it is not computational burdensome to increase  $\tau$ .

Second we used an informal stopping rule, stopping the algorithm when the norm of the bias in the estimates of  $\{W(\cdot)\}$  was sufficiently small. Instead, we could have used a formal statistical test of the null hypothesis that there was no bias (i.e., test the null  $H^0 : T = 0$ ). Notice that if we did proceed in this way we could, by increasing the number of simulation draws, increase the power of any given alternative to one. This suggests that we would want to formalize the trade-off between size, power, and the number of simulation draws, and explicitly incorporate allowance for imprecision in the computer's calculations. These are tasks we leave to future research.

TELAVIV UNIVERSITY AND CEPR  
HARVARD UNIVERSITY AND NBER

## REFERENCES

- Athey, Susan, and Kyle Bagwell, "Collusion with Persistent Cost Shocks," *Econometrica*, 76, no. 3 (2008), 493–540.
- Bajari, Patrick, Lanier Benkard, and Jonathan Levin, "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75 (2007), 1331–1370.
- Benkard, Lanier, "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft," *Review of Economic Studies*, 71, no. 3 (2004), 581–611.
- Beresteanu, Arie, and Paul Ellickson, "The Dynamics of Retail Oligopoly," (Mimeo: Department of Economics, Duke University, 2006).
- Besanko, David, Ulrich Doraszelski, Yaroslav Kryukov, and Mark Satterthwaite, "Learning by Doing, Organizational Forgetting and Industry Dynamics," *Econometrica*, 78, no. 2 (2010), 453–508.
- Borenstein, Severin, James B. Bushnell, and Frank A. Wolak, "Measuring Market Inefficiencies in California's Restructured Wholesale Electricity Market," *American Economic Review*, 92 (2002), 1376–1405.

- Cole, Harold L., and Narayana Kocherlakota, "Dynamic Games with Hidden Actions and Hidden States," *Journal of Economic Theory*, 98 (2001), 114–126.
- Dekel, Eddie, Drew Fudenberg, and David K. Levine, "Learning to Play Bayesian Games," *Games and Economic Behavior*, 46 (2004), 282–303.
- Doraszelski, Ulrich, and Sarit Markovich, "Advertising Dynamics and Competitive Advantage," *RAND Journal of Economics*, 38, no. 3 (2007), 1–36.
- Doraszelski, Ulrich, and Ariel Pakes "A Framework for Applied Dynamic Analysis in IO," In *The Handbook of Industrial Organization*, ed. Armstrong, M., and Porter, R. (Amsterdam: North Holland, 2007), 1889–1966.
- Doraszelski, Ulrich, and Mark Satterthwaite, "Computable Markov-Perfect Industry Dynamics," *RAND Journal of Economics*, 41 (2010), 215–243.
- Ericson, Richard, and Ariel Pakes, "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62 (1995), 53–82.
- Freedman, David, *Markov Chains* (New York: Springer Verlag, 1983).
- Fudenberg, Drew, and David K. Levine, "Subgame Perfect Equilibrium of Finite and Infinite Horizon Games," *Journal of Economic Theory*, 31 (1983), 227–256.
- , "Self Confirming Equilibrium," *Econometrica*, 61, no. 3 (1993), 523–545.
- Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, "Equilibrium in a Dynamic Limit Order Market," *Journal of Finance*, 60, no. 5 (2005), 2149–2192.
- , "Informed Traders and Limit Order Markets," (Mimeo: The of Chicago, the Graduate School of Business, 2008).
- Gowrisankaran, Gautam, and Robert J. Town, "Dynamic Equilibrium in the Hospital Industry," *Journal of Economics and Management Strategy*, 6, no. 1 (1997), 45–74.
- Harvey, Scott M., William W. Hogan, and Todd Schatzki, "A Hazard Rate Analysis of Mirantis Generating Plant Outgaes in California," (Mimeo: Kennedy School of Government, 2004).
- Kreps, David M., and Robert Wilson, "Sequential Equilibria," *Econometrica*, 50, no. 4 (1982), 863–894.
- Lee, R., and Ariel Pakes, "Multiple Equilibria and Selection by Learning in an Applied Setting," *Economics Letters*, 104 (2009), 13–16.
- Maskin, Eric, and Jean Tirole, "A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs," *Econometrica*, 56 (1988), 549–570.
- , "Markov Perfect Equilibrium: Observable Actions," *Journal of Economic Theory*, 100 (2001), 191–219.
- Pakes, Ariel, and Paul McGuire, "Computing Markov Perfect Nash Equilibrium: Numerical Implications of a Dynamic Differentiated Product Model," *RAND Journal of Economics*, 25, no. 4 (1994), 555–589.
- , "Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the 'Curse' of Dimensionality," *Econometrica*, 69 (2001), 1261–1281.
- Pakes, Ariel, Michael Ostrovsky, and Steve Berry, "Simple Estimators for the Parameters of Discrete Dynamic Games with Entry Exit Examples," *RAND Journal of Economics*, 38, no. 2 (2007), 373–399.
- Patrick, Robert H., and Frank A. Wolak, "The Impact of Market Rules and Market Structure on the Price Determination Process in the England and Wales Electricity Market," POWER working paper PWP-047, University of California Energy Institute, 1997.
- Robbins, Herbert, and Sutton Monroe, "A Stochastic Approximation Technique," *Annals of Mathematics and Statistics*, 22 (1951), 400–407.

This page intentionally left blank