3-2010

Learning-by-Doing, Organizational Forgetting, and Industry Dynamics

David Besanko

Ulrich Doraszelski
University of Pennsylvania

Yaroslav Kryukov

Mark Satterthwaite

Learning-by-Doing, Organizational Forgetting, and Industry Dynamics

Abstract
Learning-by-doing and organizational forgetting are empirically important in a variety of industrial settings. This paper provides a general model of dynamic competition that accounts for these fundamentals and shows how they shape industry structure and dynamics. We show that forgetting does not simply negate learning. Rather, they are distinct economic forces that interact in subtle ways to produce a great variety of pricing behaviors and industry dynamics. In particular, a model with learning and forgetting can give rise to aggressive pricing behavior, varying degrees of long-run industry concentration ranging from moderate leadership to absolute dominance, and multiple equilibria.

Keywords
dynamic stochastic games, Markov-perfect equilibrium, learning-by-doing, organizational forgetting, industry dynamics, multiple equilibria

Disciplines
Economics | Industrial Organization

Learning-by-Doing, Organizational Forgetting, and Industry Dynamics*

David Besanko† Ulrich Doraszelski‡ Yaroslav Kryukov§
Mark Satterthwaite¶

February 8, 2007

Abstract

Learning-by-doing and organizational forgetting have been shown to be important in a variety of industrial settings. This paper provides a general model of dynamic competition that accounts for these economic fundamentals and shows how they shape industry structure and dynamics. Previously obtained results regarding the dominance properties of firms' pricing behavior no longer hold in this more general setting. We show that organizational forgetting does not simply negate learning-by-doing. Rather, learning-by-doing and organizational forgetting are distinct economic forces. In particular, a model with both learning-by-doing and organizational forgetting can give rise to aggressive pricing behavior, market dominance, and multiple equilibria, whereas a model with learning-by-doing alone cannot.

*We are indebted to Luis Cabral, Jiawei Chen, Stefano Demichelis, Michaela Draganska, Ken Judd, Pedro Marin, Ariel Pakes, Michael Ryall, Karl Schmedders, Chris Shannon, Kenneth Simons, Scott Stern, Michael Whinston, and Huseyin Yildirim for comments and suggestions. We have also benefitted from the comments of participants at the Annual Duke/Northwestern/Texas Industrial Organization Theory Conference (2004), the HBS Strategy Research Conference (2004), the CEPR Conference on Applied Industrial Organization (2005), the International Conference of the Society for Computational Economics (2005), the Econometric Society World Congress (2005), EARIE (2005), the Informs Annual Meeting (2005), and the North American Winter Meetings of the Econometric Society (2006). Besanko and Doraszelski gratefully acknowledge financial support from the NSF under Grant No. 0615615. Doraszelski further benefited from the hospitality of the Hoover Institution during the academic year 2006/07. Kryukov thanks the General Motors Center for Strategy in Management at Northwestern's Kellogg School of Management for support during this project. Satterthwaite acknowledges gratefully that this material is based upon work supported by the National Science Foundation under Grant No. 0121541.

†Kellogg School of Management, Northwestern University, Evanston, IL 60208, d-besanko@kellogg.northwestern.edu.

‡Department of Economics, Harvard University, Cambridge, MA 02138, doraszelski@harvard.edu.

§Department of Economics, Northwestern University, Evanston, IL 60208, kryukov@northwestern.edu.

¶Kellogg School of Management, Northwestern University, Evanston, IL 60208, m-satterthwaite@kellogg.northwestern.edu.

1 Introduction

Empirical studies provide ample evidence that the marginal cost of production decreases with cumulative experience in a variety of industrial settings (see, e.g., Wright 1936, Hirsch 1952, DeJong 1957, Alchian 1963, Levy 1965, Kilbridge 1962, Hirschmann 1964, Preston & Keachie 1964, Baloff 1971, Dudley 1972, Zimmerman 1982, Lieberman 1984, Gruber 1992, Irwin & Klenow 1994, Jarmin 1994, Pisano 1994, Bohn 1995, Hatch & Mowery 1998, Thompson 2001, Thornton & Thompson 2001). This fall in marginal cost is known as learning-by-doing. More recent empirical studies also suggest that organizations can forget the know-how gained through learning-by-doing due to labor turnover, periods of inactivity, and failure to institutionalize tacit knowledge (see, e.g., Argote, Beckman & Epple 1990, Darr, Argote & Epple 1995, Benkard 2000, Shafer, Nembhard & Uzumeri 2001, Thompson 2003). Organizational forgetting has been largely ignored by the theoretical literature.[1] This is especially troubling because Benkard (2004) shows that organizational forgetting is essential to explain the dynamics in the market for wide-bodied airframes in the 1970s and 1980s.

In this paper we provide a general model of dynamic competition based on the Markov-perfect equilibrium framework of Ericson & Pakes (1995). We show how the economic fundamentals of learning-by-doing and organizational forgetting interact to determine the structure and dynamics of an industry. Closest in spirit to our model is the Cabral & Riordan (1994) model with learning-by-doing alone. We build on Cabral & Riordan's (1994) seminal paper and other existing models of learning-by-doing by accounting for organizational forgetting.[2] This seemingly small change has surprisingly large effects. Dynamic competition with learning-by-doing and organizational forgetting is akin to racing down an upward moving escalator. As long as a firm makes sales sufficiently frequently so that the gain in know-how from learning-by-doing outstrips the loss in know-how from organizational forgetting, it moves down its learning curve and its marginal cost decreases. However, if sales slow down or come to a halt, then the firm slides back up its learning curve and its marginal cost increases. This cannot happen in a model with learning-by-doing alone. Due to this qualitative difference, adding organizational forgetting to a model of learning-by-doing leads to a rich array of pricing behaviors and industry dynamics that the existing literature neither imagined nor explained.

It is often said that learning-by-doing promotes market dominance because it gives a more experienced firm the ability to profitably underprice its less experienced rival. As

[1] An exception is Lewis & Yildirim (2005) who study the role of organizational forgetting in the context of a multi-period procurement auction in which a single buyer faces switching costs.

[2] Prior to the infinite-horizon price-setting model of Cabral & Riordan (1994), the literature has studied learning-by-doing using finite-horizon quantity-setting models (e.g., Spence 1981, Fudenberg & Tirole 1983, Ghemawat & Spence 1985, Ross 1986, Dasgupta & Stiglitz 1988, Cabral & Riordan 1997).

Dasgupta & Stiglitz (1988) put it

> ... firm-specific learning encourages the growth of industrial concentration. To be specific, one expects that strong learning possibilities, coupled with vigorous competition among rivals, ensures that history matters ... in the sense that if a given firm enjoys some initial advantages over its rivals it can, by undercutting them, capitalise on these advantages in such a way that the advantages accumulate over time, rendering rivals incapable of offering effective competition in the long run ... (p. 247)

But if learning-by-doing can be "undone" by organizational forgetting, this raises the question whether organizational forgetting is an antidote to market dominance for two reasons. First, to the extent that the leader has more to forget than the follower, organizational forgetting should work to equalize differences between firms. Second, because organizational forgetting makes improvements in competitive position from learning-by-doing more transitory, it should make firms more reluctant to invest in the acquisition of know-how through price cuts in the first place. We reach the opposite conclusion: organizational forgetting tends to make firms more instead of less aggressive. This aggressive pricing behavior, in turn, puts the industry on a path towards market dominance.

In the absence of organizational forgetting, the price that a firm sets reflects two goals. First, by winning a sale, the firm moves down its learning curve. This is the advantage-building motive. Second, the firm prevents its rival from moving down its learning curve. This is the advantage-defending motive. In the presence of organizational forgetting, bidirectional movements through the state space are possible, and this opens up new strategic possibilities for firms that work to enhance the advantage-building and advantage-defending motives. By winning a sale, a firm makes itself less vulnerable to future losses from organizational forgetting, thus enhancing the advantage-building motive. It also makes its rival more vulnerable to future losses from organizational forgetting, thus enhancing the advantage-defending motive. Because these additional benefits are achieved by winning a sale, organizational forgetting creates strong incentives to cut prices. It is thus a source of aggressive pricing behavior.

While the existing literature has mainly focused on the dominance properties of firms' pricing behavior, we find that these properties are neither necessary nor sufficient for economically meaningful market dominance in our more general setting. We therefore go beyond the existing literature and directly examine the industry dynamics implied by firms' pricing behavior. We find that organizational forgetting is a source of—and not an antidote to—market dominance. If organizational forgetting is sufficiently weak, then asymmetries may arise but they cannot persist. If organizational forgetting is sufficiently strong,

then asymmetries cannot arise in the first place because organizational forgetting stifles investment in learning-by-doing altogether. By contrast, for intermediate degrees of organizational forgetting, asymmetries arise and persist. Even extreme asymmetries akin to near-monopoly are possible. This is because organizational forgetting predisposes the leader to defend its position aggressively against imminent and distant threats. This more than offsets the increased vulnerability to organizational forgetting as the stock of know-how grows and therefore makes the leadership position more secure than it would have been in the absence of organizational forgetting.

Organizational forgetting is also a source of multiple equilibria. If the inflow of know-how into the industry due to learning-by-doing is substantially smaller than the outflow of know-how due to organizational forgetting, then it is virtually impossible that both firms reach the bottom of their learning curves. Conversely, if the inflow is substantially greater than the outflow, then it is virtually inevitable that they do. An extreme example is the Cabral & Riordan (1994) model with learning-by-doing alone. In both cases, the primitives of the model tie down the equilibrium. This is no longer the case if the inflow roughly balances the outflow, setting the stage for multiple equilibria. If firms believe that they cannot profitably coexist at the bottom of their learning curves and that instead one firm comes to dominate the market, then both firms cut their prices in the hope of acquiring a competitive advantage early on and maintaining it throughout. This aggressive pricing behavior, in turn, leads to market dominance. However, if firms believe that they can profitably coexist, then neither firm cuts its price, thereby ensuring that the anticipated symmetric industry structure actually emerges. Consequently, in addition to the degree of organizational forgetting, the equilibrium by itself is an important determinant of pricing behavior and industry dynamics.

In our model multiple equilibria do not arise because of the specification of the primitives. In fact, we are able to show that multiple equilibria arise from firms' expectations regarding the value of continued play. In this sense multiplicity is rooted in the dynamics of the model. Our finding of multiplicity is important for two reasons. First, to our knowledge, all applications of Ericson & Pakes's (1995) framework have found a single equilibrium. It indeed is often held that "nonuniqueness does not seem to be a problem" in this setting (Pakes & McGuire 1994, p. 570). It is therefore striking that we obtain up to nine equilibria for some parameterizations. Second, being able to pinpoint the driving force behind multiple equilibria is a first step towards tackling the multiplicity problem that plagues the estimation of dynamic stochastic games and inhibits the use of counterfactuals in policy analysis (see Ackerberg, Benkard, Berry & Pakes (2005) and Pakes (2006) for a discussion of the issue).

In sum, learning-by-doing and organizational forgetting are distinct economic forces. Organizational forgetting, in particular, does not simply negate learning-by-doing. The

unique role played by organizational forgetting comes about because it makes bidirectional movements through the state space possible. As a consequence, a model with both learning-by-doing and organizational forgetting can give rise to aggressive pricing behavior, market dominance, and multiple equilibria, whereas a model with learning-by-doing alone cannot.

We also make two methodological contributions. First, we point out a weakness of the major tool for computing equilibria in the literature following Ericson & Pakes (1995). Specifically, we prove that our dynamic stochastic game has equilibria that cannot be computed by the Pakes & McGuire (1994) algorithm. Roughly speaking, in the presence of multiple equilibria, "in between" two equilibria that can be computed by the Pakes & McGuire (1994) algorithm, there is one equilibrium that cannot. This severely limits the ability of the Pakes & McGuire (1994) algorithm to provide a reasonably complete picture of the set of solutions to the model.

Second, we propose a homotopy or path-following algorithm. The algorithm traces out the equilibrium correspondence by varying the degree of organizational forgetting and allows us to compute equilibria that cannot be computed by the Pakes & McGuire (1994) algorithm. We find that the equilibrium correspondence contains a unique path that starts at the equilibrium of the model with learning-by-doing alone. Whenever this path bends back on itself and then forward again, there are multiple equilibria. In addition, the equilibrium correspondence may contain (one or more) loops that cause additional multiplicity. To our knowledge, our paper is the first to describe in detail the structure of the set of equilibria of a dynamic stochastic game in the tradition of Ericson & Pakes (1995).

The organization of the remainder of the paper is as follows. Sections 2 and 3 describe the model specification and our computational strategy. Section 4 provides an overview of the equilibrium correspondence. Section 5 analyzes industry dynamics and Section 6 characterizes the pricing behavior that drives it. Section 7 describes how organizational forgetting can lead to multiple equilibria. Section 8 undertakes a number of robustness checks. Section 9 summarizes and concludes. All proofs are relegated to the Appendix.

Throughout the paper we distinguish between propositions which are established through formal proofs and results. A result either establishes a possibility through a numerical example or summarizes a regularity through a systematic exploration of the parameter space.

2 Model

For expositional clarity we focus on the basic model of an industry with two firms and neither entry nor exit. The general model is outlined in the Online Appendix.

Firms and states. We consider a discrete-time, infinite-horizon dynamic stochastic game of complete information played by two firms. Firm $n \in \{1, 2\}$ is described by its state $e_n \in \{1, \ldots, M\}$. A firm's state indicates its cumulative experience or stock of know-how. By making a sale, a firm can add to its stock of know-how. Following Cabral & Riordan (1994), we take a period to be just long enough for a firm to make a sale.[3] In contrast to Cabral & Riordan (1994), however, we incorporate organizational forgetting in our model as suggested by the empirical studies of Argote et al. (1990), Darr et al. (1995), Benkard (2000), Shafer et al. (2001), and Thompson (2003). Accordingly, the evolution of firm n's stock of know-how is governed by the law of motion

$$e'_n = e_n + q_n - f_n,$$

where e'_n and e_n is firm n's stock of know-how in the subsequent and current period, respectively, the random variable $q_n \in \{0, 1\}$ indicates whether firm n makes a sale, and the random variable $f_n \in \{0, 1\}$ represents organizational forgetting. If $q_n = 1$, the firm gains a unit of know-how through learning-by-doing, while it loses a unit of know-how through organizational forgetting if $f_n = 1$.

At any point in time, the industry is characterized by a vector of firms' states $\mathbf{e} = (e_1, e_2) \in \{1, \ldots, M\}^2$. We refer to $\mathbf{e}$ as the state of the industry. We use $\mathbf{e}^{[2]}$ to denote the vector (e_2, e_1) found by interchanging the stocks of know-how of firms 1 and 2.

Learning-by-doing. Firm n's marginal cost of production $c(e_n)$ depends on its stock of know-how e_n through a learning curve

$$c(e_n) = \begin{cases} \kappa e_n^{\eta} & \text{if} \quad 1 \leq e_n < m, \\ \kappa m^{\eta} & \text{if} \quad m \leq e_n \leq M, \end{cases}$$

where $\eta = \log_2 \rho$ for a progress ratio of $\rho \in (0, 1]$. Marginal cost decreases by $100(1 - \rho)$ percent as the stock of know-how doubles, so that a lower progress ratio implies a steeper learning curve. The marginal cost of production at the top of the learning curve, $c(1)$, is $\kappa > 0$ and, in line with Cabral & Riordan (1994), m represents the stock of know-how at which a firm reaches the bottom of its learning curve.[4]

[3] A sale may involve a single unit or a batch of units (e.g., 100 aircraft or 10,000 memory chips) that are sold to a single buyer.

[4] While Cabral & Riordan (1994) formally consider the state space to be infinite (i.e., $M = \infty$ in our notation), they make the additional assumption that the price that a firm charges does not depend on how far it is beyond the bottom of its learning curve (p. 1119). This is tantamount to assuming, as we do, that the state space is finite.

Organizational forgetting. We let $\Delta(e_n) = \Pr(f_n = 1)$ denote the probability that firm n loses a unit of know-how through organizational forgetting. We assume that this probability is nondecreasing in the firm's experience level. This has several advantages. First, experimental evidence in the management literature suggests that forgetting by individuals is an increasing function of the current stock of learned knowledge (Bailey 1989). Second, a direct implication of $\Delta(\cdot)$ being increasing is that the expected stock of know-how in the absence of further learning is a decreasing convex function of time.[5] This phenomenon, known in the psychology literature as Jost's second law, is consistent with experimental evidence on forgetting by individuals (Wixted & Ebbesen 1991). Third, in the capital-stock model employed in empirical work on organizational forgetting the amount of depreciation is assumed to be proportional to the stock of know-how. Hence, the additional know-how needed to counteract depreciation must increase with the stock of know-how. Our specification has this feature but, unlike the capital-stock model, is consistent with a discrete state space.[6]

The specific functional form we employ is

$$\Delta(e_n) = 1 - (1 - \delta)^{e_n},$$

where we refer to $\delta \in [0, 1]$ as the forgetting rate.[7] If $\delta > 0$, then $\Delta(e_n)$ is increasing and concave in e_n; $\delta = 0$ corresponds to the absence of organizational forgetting, the special case Cabral & Riordan (1994) analyzed. Other functional forms are plausible, and we explore some of them in Section 8.

Demand. The industry draws its customers from a large pool of potential buyers. In each period, one buyer enters the market and purchases the good from one of the two firms.[8] The utility that the buyer obtains by purchasing good n is $v - p_n + \varepsilon_n$, where p_n is the price of good n, v is a deterministic component of utility, and ε_n is a stochastic component that captures the idiosyncratic preference for good n of this period's buyer. ε_1 and ε_2 are unobservable to firms and are assumed to be independently and identically type 1 extreme value distributed with location parameter 0 and scale parameter $\sigma > 0$. The scale parameter governs the degree of horizontal product differentiation. As $\sigma \to 0$, goods

[5] See the Online Appendix for a proof.

[6] See Benkard (2004) for an alternative approximation to the capital-stock model.

[7] One way to motivate this functional form is to imagine that the stock of know-how is dispersed among a firm's workforce. In particular, assume that e_n is the number of skilled workers and that organizational forgetting is the result of labor turnover. Then, given a turnover rate of δ, $\Delta(e_n)$ is the probability that at least one of the e_n skilled workers leaves the firm.

[8] Since there is a different buyer in each period, buyers are non-strategic. Lewis & Yildirim (2002, 2005) consider models with a single buyer who optimally designs a multi-period procurement auction in order to influence the dynamics of the industry.

become homogeneous.

The buyer purchases the good that gives it the highest utility. Given our distributional assumptions the probability that firm n makes a sale is given by the logit specification

$$D_n(\mathbf{p}) = \Pr(q_n = 1) = \frac{\exp(\frac{v-p_n}{\sigma})}{\sum_{k=1}^{2} \exp(\frac{v-p_k}{\sigma})} = \frac{1}{1+\exp(\frac{p_n-p_{-n}}{\sigma})},$$

where $\mathbf{p} = (p_1, p_2)$ is the vector of prices and we adopt the convention of using p_{-n} to denote the price charged by the other firm. Demand effectively depends on differences in prices because we assume in line with Cabral & Riordan (1994) that the buyer always purchases from one of the two firms in the industry. In Section 8 we include an outside good in the specification.

State-to-state transitions. From one period to the next, a firm's stock of know-how moves up or down or remains constant depending on realized demand $q_n \in \{0,1\}$ and organizational forgetting $f_n \in \{0,1\}$. The transition probabilities are

$$\Pr(e_n'|e_n, q_n) = \begin{cases} 1-\Delta(e_n) & \text{if} \quad e_n' = e_n + q_n, \\ \Delta(e_n) & \text{if} \quad e_n' = e_n + q_n - 1, \end{cases}$$

where, at the upper and lower boundaries of the state space, we modify the transition probabilities to be $\Pr(M|M,1) = 1$ and $\Pr(1|1,0) = 1$, respectively. Note that the firm can increase its stock of know-how only if it makes a sale in the current period, an event that has probability $D_n(\mathbf{e})$; otherwise it runs the risk that its stock of know-how will decrease.

Bellman equation. Define $V_n(\mathbf{e})$ to be the expected net present value of firm n's cash flows if the industry is currently in state $\mathbf{e}$. The value function $\mathbf{V}_n : \{1,\ldots,M\}^2 \to [-\hat{V}, \hat{V}]$, where $\hat{V}$ is a sufficiently large constant, is implicitly defined by the Bellman equation

$$V_n(\mathbf{e}) = \max_{p_n} D_n(p_n, p_{-n}(\mathbf{e}))(p_n - c(e_n)) + \beta \sum_{k=1}^{2} D_k(p_n, p_{-n}(\mathbf{e}))\overline{V}_{nk}(\mathbf{e}), \tag{1}$$

where $p_{-n}(\mathbf{e})$ is the price charged by the other firm in state $\mathbf{e}$, $\beta \in (0,1)$ is the discount factor, and $\overline{V}_{nk}(\mathbf{e})$ is the expectation of firm n's value function conditional on the buyer

purchasing the good from firm $k \in \{1,2\}$ in state $\mathbf{e}$ as given by

$$\overline{V}_{n1}(\mathbf{e}) = \sum_{e_1'=e_1}^{e_1+1} \sum_{e_2'=e_2-1}^{e_2} V_n(\mathbf{e}') \Pr(e_1'|e_1,1) \Pr(e_2'|e_2,0), \tag{2}$$

$$\overline{V}_{n2}(\mathbf{e}) = \sum_{e_1'=e_1-1}^{e_1} \sum_{e_2'=e_2}^{e_2+1} V_n(\mathbf{e}') \Pr(e_1'|e_1,0) \Pr(e_2'|e_2,1). \tag{3}$$

The policy function $\mathbf{p}_n : \{1,\ldots,M\}^2 \to [-\hat{p},\hat{p}]$, where $\hat{p}$ is a sufficiently large constant, specifies the price $p_n(\mathbf{e})$ that firm n sets in state $\mathbf{e}$.[9] Let $h_n(\mathbf{e}, p_n, p_{-n}(\mathbf{e}), \mathbf{V}_n)$ denote the maximand in the Bellman equation (1). Differentiating this so-called return function with respect to p_n and using the properties of logit demand we obtain the first-order condition (FOC):

$$0 = \frac{\partial h_n(\cdot)}{\partial p_n} = \frac{1}{\sigma} D_n(p_n, p_{-n}(\mathbf{e})) \Big(\sigma - (p_n - c(e_n)) - \beta \overline{V}_{nn}(\mathbf{e}) + h_n(\cdot) \Big).$$

Differentiating $h_n(\cdot)$ a second time yields

$$\frac{\partial^2 h_n(\cdot)}{\partial p_n^2} = \frac{1}{\sigma} \frac{\partial h_n(\cdot)}{\partial p_n} \Big(2D_n(p_n, p_{-n}(\mathbf{e})) - 1 \Big) - \frac{1}{\sigma} D_n(p_n, p_{-n}(\mathbf{e})).$$

If the FOC is satisfied, then $\frac{\partial^2 h_n(\cdot)}{\partial p_n^2} = -\frac{1}{\sigma} D_n(p_n, p_{-n}(\mathbf{e})) < 0$. The return function $h_n(\cdot)$ is therefore strictly quasi-concave in p_n, so that the pricing decision $p_n(\mathbf{e})$ is uniquely determined by the solution to the FOC (given $p_{-n}(\mathbf{e})$).

Equilibrium. In our model, firms face identical demand and cost primitives. Asymmetries between firms arise endogenously as a consequence of their pricing decisions for realized demand and organizational forgetting. Hence, we focus attention on symmetric Markov perfect equilibria (MPE). In a symmetric equilibrium the pricing decision taken by firm 2 in state $\mathbf{e}$ is identical to the pricing decision taken by firm 1 in state $\mathbf{e}^{[2]}$, i.e., $p_2(\mathbf{e}) = p_1(\mathbf{e}^{[2]})$, and similarly for the value function. It therefore suffices to determine the value and policy functions of firm 1, and we define $V(\mathbf{e}) = V_1(\mathbf{e})$ and $p(\mathbf{e}) = p_1(\mathbf{e})$ for each state $\mathbf{e}$. Further, we let $\overline{V}_k(\mathbf{e}) = \overline{V}_{1k}(\mathbf{e})$ denote the conditional expectation of firm 1's value function and $D_k(\mathbf{e}) = D_k(p(\mathbf{e}), p(\mathbf{e}^{[2]}))$ the probability that the buyer purchases from firm $k \in \{1,2\}$ in state $\mathbf{e}$.

[9] In what follows we assume that $\hat{p}$ is chosen large enough to not constrain pricing behavior.

Given this notation, the Bellman equation and FOC can be expressed as

$$F^1_{\mathbf{e}}\left(\mathbf{V}^*, \mathbf{p}^*\right) = -V^*(\mathbf{e}) + D^*_1(\mathbf{e})\left(p^*(\mathbf{e}) - c(e_1)\right) + \beta \sum_{k=1}^{2} D^*_k(\mathbf{e})\overline{V}^*_k(\mathbf{e}) = 0, \tag{4}$$

$$F^2_{\mathbf{e}}\left(\mathbf{V}^*, \mathbf{p}^*\right) = \sigma - \left(1 - D^*_1(\mathbf{e})\right)\left(p^*(\mathbf{e}) - c(e_1)\right) - \beta\overline{V}^*_1(\mathbf{e}) + \beta \sum_{k=1}^{2} D^*_k(\mathbf{e})\overline{V}^*_k(\mathbf{e}) = 0, \tag{5}$$

where we use asterisks to denote an equilibrium. The collection of equations (4) and (5) for all states $\mathbf{e} \in \{1, \ldots, M\}^2$ can be written more compactly as

$$\mathbf{F}(\mathbf{V}^*, \mathbf{p}^*) = \begin{bmatrix} F^1_{(1,1)}\left(\mathbf{V}^*, \mathbf{p}^*\right) \\ F^1_{(2,1)}\left(\mathbf{V}^*, \mathbf{p}^*\right) \\ \vdots \\ F^2_{(M,M)}\left(\mathbf{V}^*, \mathbf{p}^*\right) \end{bmatrix} = \mathbf{0}, \tag{6}$$

where $\mathbf{0}$ is a $(2M^2 \times 1)$ vector of zeros. Any solution to this system of $2M^2$ equations in $2M^2$ unknowns $\mathbf{V}^* = (V^*(1,1), V^*(2,1), \ldots, V^*(M,M))$ and $\mathbf{p}^* = (p^*(1,1), p^*(2,1), \ldots, p^*(M,M))$ is a symmetric equilibrium in pure strategies. A slightly modified version of Proposition 2 in Doraszelski & Satterthwaite (2007) establishes that such an equilibrium always exists for our model.

Parameterization. Our focus is on how learning-by-doing and organizational forgetting affect pricing behavior and the industry dynamics implied by that behavior. Accordingly, we explore the full range of values for the progress ratio ρ and the forgetting rate δ. To do so, we proceed as follows: First we specify a grid of 100 equidistant values of $\rho \in (0, 1]$. For each of them, we then use the homotopy algorithm described in Section 3 to trace the equilibrium as δ ranges from 0 to 1. Typically this entails solving the model for a few thousand intermediate values of δ.

Most empirical estimates of progress ratios are in the range of 0.7 to 0.95 (Dutton & Thomas 1984). However, a very steep learning curve with ρ much less than 0.7 may also capture a practically relevant situation. Suppose the first unit of a product is a hand-built prototype and the second unit is a guinea pig for organizing the production line. After this point the gains from learning-by-doing are more or less exhausted and the marginal cost of production is close to zero.[10]

We note that empirical studies have found monthly rates of depreciation ranging from

[10] To avoid a marginal cost of close to zero, shift the cost function $c(e_n)$ by $\tau > 0$. While introducing a component of marginal cost that is unresponsive to learning-by-doing shifts the policy function by τ, the value function and the industry dynamics are left the same.

4 to 25 percent of the stock of know-how (Benkard 2000, Argote et al. 1990). In the Online Appendix we show how to map these estimates that are based on a capital-stock model of learning-by-doing and organizational forgetting into in our specification. The implied values of the forgetting rate δ fall below 0.1.

We fix the values of the remaining parameters until Section 8 where we discuss their influence on the equilibrium and demonstrate the robustness of our conclusions. In our baseline parameterization, we set $M = 30$ and $m = 15$. The marginal cost at the top of the learning curve κ is equal to 10. For a progress ratio of $\rho = 0.85$, this implies that the marginal cost of production declines from a maximum value of $c(1) = 10$ to a minimum value of $c(15) = \ldots = c(30) = 5.30$. For $\rho = 0.15$, we have the case of a hand-built prototype where the marginal cost of production declines very quickly from $c(1) = 10$ over $c(2) = 1.50$ and $c(3) = 0.49$ to $c(15) = \ldots = c(30) = 0.01$.

Turning to demand, we set $\sigma = 1$ in our baseline parameterization. To illustrate, in the Nash equilibrium of a static price-setting game (obtained by setting $\beta = 0$) the own-price elasticity of demand ranges between -8.86 in state $(1, 15)$ and -2.13 in state $(15, 1)$ for a progress ratio of $\rho = 0.85$. The cross-price elasticity of firm 1's demand with respect to firm 2's price is 2.41 in state $(15, 1)$ and 7.84 in state $(1, 15)$. For $\rho = 0.15$ the own-price elasticity ranges between -9.89 and -1.00 and the cross-price elasticity between 1.00 and 8.05. These reasonable elasticities suggest that the results reported below are not artifacts of extreme parameterizations.

We set the discount factor to $\beta = \frac{1}{1.05}$. The discount factor can be thought of as $\beta = \frac{\zeta}{1+r}$, where $r > 0$ is the per-period discount rate and $\zeta \in (0, 1]$ is the exogenous probability that the industry survives from one period to the next. Consequently, our baseline parameterization corresponds to a variety of scenarios that differ in the length of a period. For example, it corresponds to a period length of one year, a yearly discount rate of 5 percent, and certain survival. Perhaps more interesting, it also corresponds to a period length of one month, a monthly discount rate of 1 percent (which translates into a yearly discount rate of 12.68 percent), and a monthly survival probability of 0.96. To put this—our focal scenario—in perspective, technology companies such as IBM and Microsoft had costs of capital in the range of 11 to 15 percent per annum in the late 1990s. Further, an industry with a monthly survival probability of 0.96 has an expected lifetime of 26.25 months. Thus this scenario is consistent with a pace of innovative activity that is expected to make the current generation of products obsolete within two to three years.

3 Computation

In this section we first describe a novel algorithm for computing equilibria that is based on homotopy methods. Then we turn to the Pakes & McGuire (1994) algorithm—the main tool in the literature initiated by Ericson & Pakes (1995)—and show that it is inadequate for characterizing the set of solutions to our model. A reader who is more interested in the economic implications of learning-by-doing and organizational forgetting may skip ahead to Section 4.

3.1 Homotopy algorithm

Our homotopy or path-following algorithm is designed to explore the set of equilibria in a systematic fashion. It is especially useful in models like ours that have multiple equilibria. Starting from a single equilibrium that has already been computed for a given parameterization of the model, the homotopy algorithm traces out an entire path of equilibria by varying a parameter of interest.[11]

In Section 4 we show that the equilibrium is unique if organizational forgetting is either absent ($\delta = 0$) or certain ($\delta = 1$). This makes the forgetting rate the natural choice for the homotopy parameter. The object of interest is therefore the equilibrium correspondence

$$\mathbf{F}^{-1} = \{(\mathbf{V}^*, \mathbf{p}^*, \delta) | \mathbf{F}(\mathbf{V}^*, \mathbf{p}^*, \delta) = \mathbf{0}\},$$

where $\mathbf{F}(\cdot)$ is the system of equations (6) that defines an equilibrium and we make explicit that it depends on δ. Note that we hold fixed all parameters other than the forgetting rate. The homotopy algorithm follows the path that connects the unique equilibrium at $\delta = 0$ with the unique equilibrium at $\delta = 1$. Whenever this path folds back on itself, the homotopy algorithm automatically identifies multiple equilibria.

However, the equilibrium correspondence may consist of more than the path that connects $\delta = 0$ and $\delta = 1$. Through trial-and-error and educated guesses we have been able to identify equilibria off this "main path." Feeding these equilibria as initial conditions to the homotopy algorithm shows that the equilibrium correspondence contains (one or more) loops that are disjoint from the main path (see Section 4 for details). Unfortunately, there is no systematic approach for obtaining an initial condition for a loop of equilibria and, consequently, the homotopy algorithm cannot be guaranteed to find all equilibria.[12] As we

[11] See Zangwill & Garcia (1981) for an introduction to homotopy methods, Schmedders (1998, 1999) for an application to general equilibrium models with incomplete asset markets, and Berry & Pakes (2006) for an application to estimating demand systems.

[12] Unless the system of equations that defines them happens to be polynomial. See Judd & Schmedders (2004) for some early efforts along this line.

show in Section 3.2, it, however, is able to find many more equilibria than the Pakes & McGuire (1994) algorithm.

Example. An example is helpful to explain how the homotopy algorithm works. Consider the equation $F(x, \delta) = 0$, where

$$F(x, \delta) = -15.289 - \frac{\delta}{1 + \delta^4} + 67.500x - 96.923x^2 + 46.154x^3. \tag{7}$$

Equation (7) implicitly relates an endogenous variable x with an exogenous parameter δ. The set of solutions $F^{-1} = \{(x, \delta) | F(x, \delta) = 0\}$ is graphed in Figure 1. There evidently are multiple solutions to equation (7), e.g., $x = 0.610$, $x = 0.707$, and $x = 0.783$ at $\delta = 0.3$. Finding these solutions is trivial with the graph in hand, but the graph is less than straightforward to draw even in this very simple case. Whether one solves $F(x, \delta) = 0$ for x taking δ as given or for δ taking x as given, the result is a multi-valued correspondence, not a single-valued function.

To apply the homotopy method, we introduce an auxiliary variable s that indexes each point on the graph starting at point A for $s = 0$ and ending at point D for $s = \bar{s}$. The graph is then just the parametric path given by a pair of functions $(x(s), \delta(s))$ satisfying $F(x(s), \delta(s)) = 0$ or, equivalently, $(x(s), \delta(s)) \in F^{-1}$. While there are infinitely many such pairs, there is a simple way to select a member of this family. Differentiate $F(x(s), \delta(s)) = 0$ with respect to s to obtain

$$\frac{\partial F(x(s), \delta(s))}{\partial x} x'(s) + \frac{\partial F(x(s), \delta(s))}{\partial \delta} \delta'(s) = 0. \tag{8}$$

This differential equation in two unknowns $x'(s)$ and $\delta'(s)$ captures the condition that is required to remain "on path." One possible approach for tracing out a path in F^{-1} is thus to solve equation (8) for the ratio $\frac{x'(s)}{\delta'(s)} = -\frac{\partial F(x(s), \delta(s))/\partial \delta}{\partial F(x(s), \delta(s))/\partial x}$ that indicates the direction of the next step along the path from s to $s + ds$. This approach, however, creates difficulties because the ratio may switch from $+\infty$ to $-\infty$, e.g., at point B in Figure 1. So instead of solving for the ratio, we simply solve for each term of the ratio. This insight implies that the graph of F^{-1} in Figure 1 is the solution to the system of differential equations

$$x'(s) = \frac{\partial F(x(s), \delta(s))}{\partial \delta}, \tag{9}$$

$$\delta'(s) = -\frac{\partial F(x(s), \delta(s))}{\partial x}. \tag{10}$$

Equations (9) and (10) are the so-called basic differential equations for our example. Their significance is that they reduce the task of tracing out the set of solutions to solving

a system of differential equations. Given an initial condition this is can be done with a variety of methods (see, e.g., Chapter 10 of Judd 1998). In our example, note that if $\delta = 0$, then $F(x, \delta) = 0$ is easily solved for $x = 0.5$. This provides the initial condition (point A in Figure 1). From there the homotopy algorithm uses the basic differential equations to determine the next step along the path. It continues to follow—step-by-step—the path until it reaches $\delta = 1$ (point D). Whenever $\delta'(s)$ switches sign from negative to positive (point B), the path is bending backward and there are multiple solutions. Conversely, whenever the sign of $\delta'(s)$ switches back from positive to negative (point C), the path is bending forward.[13]

Returning to our model of learning-by-doing and organizational forgetting, let $\mathbf{x} = (\mathbf{V}^*, \mathbf{p}^*)$ denote the $2M^2$ endogenous variables. Our goal is to explore the equilibrium correspondence $\mathbf{F}^{-1} = \{(\mathbf{x}, \delta) | \mathbf{F}(\mathbf{x}, \delta) = \mathbf{0}\}$ that depends on the exogenous parameter δ. Proceeding as in our example, a parametric path is a set of functions $(\mathbf{x}(s), \delta(s)) \in \mathbf{F}^{-1}$. Differentiating $\mathbf{F}(\mathbf{x}(s), \delta(s)) = \mathbf{0}$ with respect to s yields the conditions that are required to remain on path

$$\frac{\partial \mathbf{F}(\mathbf{x}(s), \delta(s))}{\partial \mathbf{x}} \mathbf{x}'(s) + \frac{\partial \mathbf{F}(\mathbf{x}(s), \delta(s))}{\partial \delta} \delta'(s) = \mathbf{0}, \tag{11}$$

where $\frac{\partial \mathbf{F}(\mathbf{x}(s), \delta(s))}{\partial \mathbf{x}}$ is the $(2M^2 \times 2M^2)$ Jacobian, $\mathbf{x}'(s)$ and $\frac{\partial \mathbf{F}(\mathbf{x}(s), \delta(s))}{\partial \delta}$ are $(2M^2 \times 1)$ vectors, and $\delta'(s)$ is a scalar. This system of $2M^2$ differential equations in $2M^2 + 1$ unknowns $x_i'(s)$, $i = 1, \ldots, 2M^2$, and $\delta'(s)$ has a solution that obeys the basic differential equations

$$y_i'(s) = (-1)^{i+1} \det \left(\left(\frac{\partial \mathbf{F}(\mathbf{y}(s))}{\partial \mathbf{y}} \right)_{-i} \right), \quad i = 1, \ldots, 2M^2 + 1, \tag{12}$$

where $\mathbf{y}(s) = (\mathbf{x}(s), \delta(s))$ and the notation $(\cdot)_{-i}$ is used to indicate that the ith column is removed from the $(2M^2 \times 2M^2 + 1)$ Jacobian $\frac{\partial \mathbf{F}(\mathbf{y}(s))}{\partial \mathbf{y}}$. Note that equation (12) reduces to equations (9) and (10) if $\mathbf{x}$ is a scalar instead of a vector. For the general case, a proof that the basic differential equations satisfy the conditions in equation (11) that are required to remain on path can be found in Garcia & Zangwill (1979) and on pp. 27–28 of Zangwill & Garcia (1981).

A closer inspection of the basic differential equations (12) points to a potential difficulty with the homotopy algorithm. If the Jacobian $\frac{\partial \mathbf{F}(\mathbf{y}(s))}{\partial \mathbf{y}}$ has less than full rank, then the determinants of all its submatrices are zero. Thus, $y_i'(s) = 0$, $i = 1, \ldots, 2M + 1$, and the homotopy algorithm stalls. A central condition in the mathematical literature on homotopy

[13]The orientation of the path taken by the homotopy algorithm is arbitrary. Reversing the signs of the basic differential equations implies, perhaps more intuitively, that $\delta'(s)$ switches sign from positive to negative at point B.

methods is that the Jacobian has full rank at all points. If so, the homotopy is called regular and the algorithm is sure to trace out a path. Moreover, regularity rules out both isolated equilibria and continua of equilibria. While we have been unable to prove that our homotopy is regular, we have been able to verify that the Jacobian always had full rank at all points along all paths taken by our homotopy algorithm.[14]

3.2 Pakes & McGuire (1994) algorithm

The Pakes & McGuire (1994) algorithm or some other means for solving a system of nonlinear equations (see, e.g., Judd 1998) is needed in order to compute a starting point for our homotopy algorithm. The Pakes & McGuire (1994) algorithm is the main tool in the literature initiated by Ericson & Pakes (1995). It is intuitively appealing because it combines value function iteration as familiar from dynamic programming with best reply dynamics (akin to Cournot adjustment) as familiar from static games.

Recall that $V_2(\mathbf{e}) = V_1(\mathbf{e}^{[2]})$ and $p_2(\mathbf{e}) = p_1(\mathbf{e}^{[2]})$ for each state $\mathbf{e}$ in a symmetric equilibrium and it therefore suffices to determine $\mathbf{V}$ and $\mathbf{p}$, the value and policy functions of firm 1. The Pakes & McGuire (1994) algorithm is iterative. An iteration cycles through the states in some predetermined order, successively updating $\mathbf{V}$ and $\mathbf{p}$ as it progresses from one iteration to the next.

The strategic situation faced by firms in setting prices in state $\mathbf{e}$ is similar to a static game if the value of continued play is taken as given. The Pakes & McGuire (1994) algorithm computes the best reply of firm 1 against $p(\mathbf{e}^{[2]})$ in this game. The best reply serves to update the value and policy functions of firm 1 in state $\mathbf{e}$. More formally, let $h_1(\mathbf{e}, p_1, p(\mathbf{e}^{[2]}), \mathbf{V})$ be the maximand in the Bellman equation (1) after symmetry is imposed. The best reply of firm 1 against $p(\mathbf{e}^{[2]})$ in state $\mathbf{e}$ is given by

$$G^2_{\mathbf{e}}(\mathbf{V}, \mathbf{p}) = \arg\max_{p_1} h_1(\mathbf{e}, p_1, p(\mathbf{e}^{[2]}), \mathbf{V}) \tag{13}$$

and the value associated with it is

$$G^1_{\mathbf{e}}(\mathbf{V}, \mathbf{p}) = \max_{p_1} h_1(\mathbf{e}, p_1, p(\mathbf{e}^{[2]}), \mathbf{V}). \tag{14}$$

[14] Our programs use Hompack (Watson, Billups & Morgan 1987, Watson, Sosonkina, Melville & Morgan 1997) written in Fortran 90 and are available from the authors upon request.

Write the collection of equations (13) and (14) for all states $\mathbf{e} \in \{1, \ldots, M\}^2$ as

$$\mathbf{G}(\mathbf{V}, \mathbf{p}) = \begin{pmatrix} G^1_{(1,1)}(\mathbf{V}, \mathbf{p}) \\ G^1_{(2,1)}(\mathbf{V}, \mathbf{p}) \\ \vdots \\ G^2_{(M,M)}(\mathbf{V}, \mathbf{p}) \end{pmatrix}. \tag{15}$$

Given an initial guess $\mathbf{x}^0 = (\mathbf{V}^0, \mathbf{p}^0)$, the Pakes & McGuire (1994) algorithm executes the iteration

$$\mathbf{x}^{k+1} = \mathbf{G}(\mathbf{x}^k), \quad k = 0, 1, 2, \ldots.$$

The algorithm aims to compute a fixed point $\mathbf{x} = \mathbf{G}(\mathbf{x})$ by continuing to iterate until the changes in the value and policy functions of firm 1 are deemed small (or a failure to converge is diagnosed). Any fixed point $\mathbf{x} = (\mathbf{V}^*, \mathbf{p}^*)$ of $\mathbf{G}$ is a symmetric equilibrium in pure strategies to our game.

Unlike our homotopy algorithm, the Pakes & McGuire (1994) algorithm does not lend itself to computing multiple equilibria. To identify more than one equilibrium (for a given parameterization of the model), the Pakes & McGuire (1994) algorithm must be restarted from different initial guesses. But different initial guesses may or may not lead to different equilibria. This, however, still understates the severity of the problem here: When there are multiple equilibria, the trial-and-error approach is sure to miss a substantial fraction of them even if an arbitrary number of initial guesses are tried. That is, our dynamic stochastic game has equilibria that cannot computed by the Pakes & McGuire (1994) algorithm.

Recall that the Pakes & McGuire (1994) algorithm continues to iterate until it reaches a fixed point $\mathbf{x} = \mathbf{G}(\mathbf{x})$. A necessary condition for it to converge is that the fixed point is locally stable. Specifically, consider the $(2M^2 \times 2M^2)$ Jacobian $\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}}$ at the fixed point and let $\rho\left(\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}}\right)$ be its spectral radius. The fixed point is locally stable under the Pakes & McGuire (1994) algorithm if $\rho\left(\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}}\right) < 1$, i.e., if all eigenvalues are within the complex unit circle. Given local stability, the algorithm converges provided that the initial guess is close (perhaps very close) to the fixed point. Conversely, the fixed point is unstable and cannot be computed by the Pakes & McGuire (1994) algorithm if $\rho\left(\frac{\partial \mathbf{G}(\mathbf{x})}{\partial \mathbf{x}}\right) \geq 1$.

In the remainder of this section we consider a parametric path $(\mathbf{x}(s), \delta(s)) \in \mathbf{F}^{-1}$ in the equilibrium correspondence, such as the path taken by our homotopy algorithm. Along this path we ask whether the equilibrium $\mathbf{x}(s)$ is locally stable or unstable under the Pakes & McGuire (1994) algorithm when the forgetting rate is set to $\delta(s)$. Proposition 1 identifies a subset of equilibria that the Pakes & McGuire (1994) algorithm is sure to miss.

Proposition 1 *Let* $(\mathbf{x}(s), \delta(s)) \in \mathbf{F}^{-1}$. *(i) If* $\delta'(s) \leq 0$, *then* $\rho\left(\frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}}\Big|_{\delta=\delta(s)}\right) \geq 1$. *(ii)*

Moreover, the equilibrium $\mathbf{x}(s)$ *remains unstable even if either dampening or extrapolation is applied to the Pakes & McGuire (1994) algorithm.*

Part (i) of Proposition 1 establishes that no equilibrium on the part of the equilibrium correspondence where $\delta'(s) \leq 0$ can be computed by the Pakes & McGuire (1994) algorithm. Whenever $\delta'(s)$ switches sign from positive to negative, the path that connects the unique equilibrium at $\delta = 0$ with the unique equilibrium at $\delta = 1$ bends backward and there are multiple equilibria. Conversely, whenever the sign of $\delta'(s)$ switches back from negative to positive, the path bends forward. Hence, holding fixed the forgetting rate, in between two equilibria with $\delta'(s) > 0$, there is one equilibrium with $\delta'(s) \leq 0$ that cannot be computed by the Pakes & McGuire (1994) algorithm. Similarly, a loop is necessarily composed of equilibria with $\delta'(s) > 0$ and equilibria with $\delta'(s) \leq 0$. The latter cannot be computed by the Pakes & McGuire (1994) algorithm.

Dampening and extrapolation are often applied to the Pakes & McGuire (1994) algorithm in the hope of improving its likelihood or speed of convergence. The iteration

$$\mathbf{x}^{k+1} = \omega \mathbf{G}(\mathbf{x}^k) + (1 - \omega)\mathbf{x}^k, \quad k = 0, 1, 2, \ldots,$$

is said to be dampened if $\omega \in (0, 1)$ and extrapolated if $\omega \in (1, \infty)$. Part (ii) of Proposition 1 establishes the futility of these attempts.[15]

The ability of the Pakes & McGuire (1994) algorithm to provide a reasonably complete picture of the set of solutions to the model is limited beyond the scope of Proposition 1. As our computations indicate, some equilibria on the part of the equilibrium correspondence where $\delta'(s) > 0$ also cannot be computed by the Pakes & McGuire (1994) algorithm:

Result 1 *Let* $(\mathbf{x}(s), \delta(s)) \in \mathbf{F}^{-1}$. *If* $\delta'(s) > 0$, *then we may have* $\rho\left(\frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}} \Big|_{\delta = \delta(s)} \right) \geq 1$.

In the Online Appendix we provide a graphic illustration of Proposition 1 and Result 1.

As is well-known, not all Nash equilibria of static games are stable under best reply dynamics (see, e.g., Fudenberg & Tirole 1991).[16] Since the Pakes & McGuire (1994) algorithm incorporates best reply dynamics, it is reasonable to expect that this limits its usefulness. In the Online Appendix we argue that this is not the case. More precisely, we show that, holding fixed the value of continued play, the best reply dynamics are contractive and therefore converge to a unique fixed point irrespective of the initial guess. The value function iteration also is contractive holding fixed the policy function. Hence, each of

[15] Dampening and extrapolation may, of course, still be helpful in computing equilibria with $\delta'(s) > 0$.

[16] More generally, in static games, Nash equilibria of degree -1 are unstable under any Nash dynamics, i.e., dynamics with rest points that coincide with Nash equilibria, including replicator and smooth fictitious play dynamics (Demichelis & Germano 2002).

the two building blocks of the Pakes & McGuire (1994) algorithm "works." What makes it impossible to obtain a large fraction of equilibria is the combination of value function iteration with best reply dynamics.

The Pakes & McGuire (1994) algorithm is also known as a pre-Gauss-Jacobi method. The subsequent literature has sometimes instead used a pre-Gauss-Seidel method (see, e.g., Benkard 2004, Doraszelski & Judd 2004). Whereas a Gauss-Jacobi method replaces the old guesses for the value and policy functions with the new guesses at the end of an iteration after all states have been visited, a Gauss-Seidel method updates after each state. This has the advantage that "information" is used as soon as it becomes available (see Chaps. 3 and 5 of Judd (1998) for an extensive discussion of Gaussian methods). While we have been unable to prove that Proposition 1 carries over to this alternative algorithm, we note that the Stein-Rosenberg theorem asserts, at least for certain systems of linear equations, that if the Gauss-Jacobi algorithm fails to converge then so does the Gauss-Seidel algorithm (see Proposition 6.9 in Section 2.6 of Bertsekas & Tsitsiklis 1997).

4 Equilibrium correspondence

This section provides an overview of the equilibrium correspondence. Section 5 analyzes industry dynamics and Section 6 characterizes the pricing behavior that drives it. Section 7 describes how organizational forgetting can lead to multiple equilibria.

In the absence of organizational forgetting, Cabral & Riordan (1994) show that the equilibrium is unique. The following proposition generalizes their result:

Proposition 2 *If organizational forgetting is either absent ($\delta = 0$) or certain ($\delta = 1$), then there is a unique equilibrium.*

Note that Proposition 2 pertains to both symmetric and asymmetric equilibria.[17] In what follows, we restrict attention to symmetric equilibria.

The cases of $\delta = 0$ and $\delta = 1$ are special in that they ensure that movements through the state space are unidirectional. Specifically, when $\delta = 0$, a firm can never move "backward" to a lower state, and when $\delta = 1$, it can never move "forward" to a higher state. In contrast, when $\delta \in (0, 1)$, a firm can move in either direction. Our computations show that this has a substantive impact on the set of equilibria:

Result 2 *If organizational forgetting is neither absent ($\delta = 0$) nor certain ($\delta = 1$), then there may be multiple equilibria.*

[17] Recall that in a symmetric equilibrium the pricing decision taken by firm 2 in state $\mathbf{e}$ is identical to the pricing decision taken by firm 1 in state $\mathbf{e}^{[2]}$, i.e., $p_2(\mathbf{e}) = p_1(\mathbf{e}^{[2]})$, and similarly for the value function. In an asymmetric equilibrium this is not necessarily the case.

Figure 2 illustrates the extent of multiplicity. It shows the number of equilibria for each combination of forgetting rate δ and progress ratio ρ. Darker shades indicate more equilibria. As can be seen, we have found up to nine equilibria for some values of δ and ρ. Multiplicity is especially pervasive for forgetting rates δ in the empirically relevant range below 0.1; indeed, we always obtained a unique equilibrium for sufficiently large forgetting rates ($\delta \geq 0.15$).

In dynamic stochastic games with finite actions, Herings & Peeters (2004) have shown that generically the number of Markov perfect equilibria is odd. While they consider both symmetric and asymmetric equilibria, in a two-player game with symmetric primitives such as ours, asymmetric equilibria occur in pairs. Hence, their result immediately implies that generically the number of symmetric equilibria is odd in games with finite actions. Figure 2 suggests that this carries over to our setting with continuous actions.

We next take a closer look at the set of equilibria. Let

$$\mathbf{F}^{-1} = \{(\mathbf{V}^*, \mathbf{p}^*, \delta) | \mathbf{F}(\mathbf{V}^*, \mathbf{p}^*, \delta) = \mathbf{0}\},$$

be the equilibrium correspondence, where $\mathbf{F}(\cdot)$ is the system of equations (6) that defines an equilibrium. Our homotopy algorithm traces out an entire path of equilibria by varying the forgetting rate (as explained in Section 3.1). We thus make explicit in our notation that the system of equations (6) depends on δ but, at the most basic level of the analysis, hold fixed all parameters other than δ. To extend the analysis we then vary the remaining parameters. For the progress ratio, in particular, we explore a grid of 100 equidistant values of $\rho \in (0, 1]$. We do not index $\mathbf{F}^{-1}$ by the remaining parameters, however, for notational simplicity.

We have the following result:

Result 3 *The equilibrium correspondence* $\mathbf{F}^{-1}$ *contains a unique path that connects the equilibrium at* $\delta = 0$ *with the equilibrium at* $\delta = 1$. *In addition,* $\mathbf{F}^{-1}$ *may contain (one or more) loops that are disjoint from the above path and from each other.*

Figure 3 illustrates Result 3. An equilibrium is defined in terms of a value and a policy function and is thus an element of a high-dimensional space. To succinctly describe it, we proceed in two steps.

First, we use the policy function to construct the probability distribution over next period's state $\mathbf{e}'$ given this period's state $\mathbf{e}$, i.e., the transition matrix that characterizes the Markov process of industry dynamics. This allows us to use stochastic process theory to analyze the Markov process of industry dynamics rather than rely on simulation. We compute the transient distribution over states in period t, $\mu^t(\cdot)$, starting from state $(1, 1)$. This tells us how likely each possible industry structure is in period t, given that both firms began the game at the top of their learning curves. In addition, we compute the limiting

(or ergodic) distribution over states, $\mu^{\infty}(\cdot)$.[18] The transient distribution captures short-run dynamics and the limiting distribution captures long-run (or steady-state) dynamics.

Second, we use the transient distribution over states in period t, $\mu^t(\cdot)$, to compute the expected Herfindahl index

$$H^t = \sum_{\mathbf{e}} \left(D_1^*(\mathbf{e})^2 + D_2^*(\mathbf{e})^2 \right) \mu^t(\mathbf{e}).$$

The time path of the expected Herfindahl index summarizes the implications of learning-by-doing and organizational forgetting for the dynamics of the industry. To the extent that the industry evolves asymmetrically, the expected Herfindahl index exceeds 0.5. The maximum expected Herfindahl index

$$H^{\wedge} = \max_{t \in \{1,\ldots,100\}} H^t$$

is therefore a summary measure of short-run industry concentration. In addition, we use the limiting distribution over states, $\mu^{\infty}(\cdot)$, to compute the limiting expected Herfindahl index H^{∞}, a summary measure of long-run industry concentration. If $H^{\infty} > 0.5$, then an asymmetric industry structure persists.

We visualize the equilibrium correspondence $\mathbf{F}^{-1}$ for a variety of progress ratios by plotting the maximum expected Herfindahl index $H^{\wedge}$ (dashed line) and the limiting expected Herfindahl index H^{∞} (solid line). As can be seen, there are multiple equilibria whenever the path that connects the equilibrium at $\delta = 0$ with the equilibrium at $\delta = 1$ folds back on itself. Moreover, the equilibrium correspondence contains one loop for $\rho \in \{0.75, 0.65, 0.55, 0.15, 0.05\}$ and two loops for $\rho \in \{0.95, 0.85, 0.35\}$, thus adding further equilibria.

Figure 3 is not necessarily a complete picture of the set of solutions to our model. As discussed in Section 3.1, no algorithm is guaranteed to find all equilibria, and our homotopy algorithm is no exception. We do find all equilibria along the path that connects the equilibrium at $\delta = 0$ with the equilibrium at $\delta = 1$, and we have been successful in finding a number of loops. But other loops may exist because, in order to trace out a loop, we must somehow compute at least one equilibrium on the loop, and doing so is problematic.

Types of equilibria. Despite the multiplicity, the equilibria of our game exhibit four typical patterns. One should recognize that these patterns, helpful as they are in under-

[18] Let $\mathbf{P}$ be the $M^2 \times M^2$ transition matrix. The transient distribution in period t is given by $\mu^t = \mu^0 \mathbf{P}^t$, where μ^0 is the $1 \times M^2$ initial distribution and $\mathbf{P}^t$ the t^{th} matrix power of $\mathbf{P}$. If $\delta \in (0, 1)$, then the Markov process is irreducible because logit demand implies that the probability moving forward is always nonzero. That is, all its states belong to a single closed communicating class and the $1 \times M^2$ limiting distribution μ^{∞} solves the system of linear equations $\mu^{\infty} = \mu^{\infty} \mathbf{P}$. If $\delta = 0$ ($\delta = 1$), then there is also a single closed communicating class, but its sole member is state (M, M) $((1, 1))$.

standing the range of behaviors that can occur, lie on a continuum and thus morph into each other as we change the parameter values.

Figure 4 exemplifies the policy functions of the typical equilibria.[19] The parameter values are $\rho = 0.85$ and $\delta \in \{0, 0.0275, 0.08\}$ and represent the median progress ratio across a wide array of empirical studies combined with the cases of no, low, and high organizational forgetting. The graph in the upper left panel of Figure 4 ($\rho = 0.85$ and $\delta = 0$) is typical for what we call a *flat equilibrium without well*. The policy function is very even over the entire state space. In particular, the price that a firm charges in equilibrium is fairly insensitive to its rival's stock of know-how. In a *flat equilibrium with well*, the policy function continues to be very even over most of the state space. However, price competition is intense in a neighborhood of state $(1, 1)$, which manifests itself as a "well" in the policy function (see the upper right panel of Figure 4 for the case of $\rho = 0.85$ and $\delta = 0.0275$). The graph in the lower left panel of Figure 4 exemplifies a *trenchy equilibrium*. The parameter values are the same ($\rho = 0.85$ and $\delta = 0.0275$), thereby providing an instance of multiplicity. The policy function is more uneven and exhibits a "trench" along the diagonal of the state space. This trench extends from state $(1, 1)$ beyond the bottom of the learning curve in state (m, m) all the way to state (M, M). Hence, in a trenchy equilibrium, price competition between firms with similar stocks of know-how is extremely intense, but price competition abates once firms become asymmetric. Finally, in an *extra-trenchy equilibrium*, the policy function not only has a diagonal trench, but it also has a trench parallel to the edge of the state space. In an extra-trenchy equilibrium, price competition between symmetric firms is extremely intense. Furthermore, due to the sideways trench, there are also parts of the state space where the leader competes aggressively with the follower (see the lower right panel of Figure 4 for the case of $\rho = 0.85$ and $\delta = 0.08$).

Sunspots. For a progress ratio of $\rho = 1$ the marginal cost of production is constant at $c(1) = \ldots = c(M) = \kappa$, and there are no gains from learning-by-doing. It clearly is an equilibrium for firms to disregard their stocks of know-how and set the same prices as in the Nash equilibrium of a static price-setting game (obtained by setting $\beta = 0$). Since firms' marginal costs are constant, so are the static Nash equilibrium prices. Thus, we have an extreme example of a flat equilibrium with $p^*(\mathbf{e}) = \kappa + 2\sigma = 12$ and $V^*(\mathbf{e}) = \frac{\sigma}{1-\beta} = 21$ for all states $\mathbf{e} \in \{1, \ldots, M\}^2$. As Figure 2 shows, however, there are other equilibria for a range of forgetting rates δ below 0.1. Since the state of the industry has no bearing on primitives in case of $\rho = 1$, we refer to these equilibria as sunspots, but we note that they persist for $\rho \approx 1$.

In a sunspot, firms use the state to keep track of their sales. That is, the state serves

[19] The corresponding value functions can be found in the Online Appendix.

merely as a coordination device. One of the sunspots is a trenchy equilibrium while the other one is, depending on the value of the forgetting rate, either a flat or a trenchy equilibrium. In the trenchy equilibrium the industry evolves towards an asymmetric structure where the leader charges a lower price than the follower and enjoys a higher probability of making a sale. Consequently, the net present value of cash flows to the leader exceeds that to the follower. The value in state $(1,1)$, however, is lower than in the static Nash equilibrium, i.e., $V^*(1,1) < 21$.[20] This indicates that value is destroyed as firms fight for dominance. More generally, the existence of sunspots suggests that the concept of Markov perfect equilibrium is richer than one may have thought.

In sum, accounting for organizational forgetting in a model of learning-by-doing leads to multiple equilibria and a rich array of pricing behaviors. In the following section, we explore what this entails for industry dynamics.

5 Industry dynamics

Recall that the transient distribution over states in period t, $\mu^t(\cdot)$, starting from state $(1,1)$, captures short-run dynamics and the limiting distribution, $\mu^\infty(\cdot)$, captures long-run dynamics. Figures 5 and 6 display the transient distribution in period 8 and 32, respectively, and Figure 7 displays the limiting distribution for our four typical cases.[21] In the *flat equilibrium without well* ($\rho = 0.85$, $\delta = 0$, see upper left panels), the transient and limiting distributions are unimodal. The most likely industry structure is symmetric. For example, the modal state is $(5,5)$ in period 8, $(9,9)$ in period 16, $(17,17)$ in period 32, and $(30,30)$ in period 64. Turning from the short run to the long run, the industry is most likely to remain in state $(30,30)$ because, in the absence of organizational forgetting, both firms must eventually reach the bottom of their learning curves. In short, the industry starts symmetric and stays symmetric.

By contrast, in the *flat equilibrium with well* ($\rho = 0.85$, $\delta = 0.0275$, see upper right panels) the transient distributions are first bimodal and then unimodal as is the limiting distribution. The modal states are $(1,8)$ and $(8,1)$ in period 8, $(4,11)$ and $(11,4)$ in period 16, $(9,14)$ and $(14,9)$ in period 32, but the modal state is $(17,17)$ in period 64 and the modal states of the limiting distribution are $(24,25)$ and $(25,24)$. Thus, as times passes, firms compete on equal footing. In sum, the industry evolves first towards an asymmetric

[20] For a forgetting rate of $\delta = 0.0275$, for example, we have $V^*(28,21) = 25.43$ and $p^*(28,21) = 12.33$ for the leader, $V^*(21,28) = 22.39$ and $p^*(21,28) = 12.51$ for the follower, and $V^*(1,1) = 19.36$. For $\delta = 0.08$ we have $V^*(12,6) = 23.41$ and $p^*(12,6) = 11.96$ for the leader, $V^*(6,12) = 18.77$ and $p^*(6,12) = 12.45$ for the follower, and $V^*(1,1) = 15.94$.

[21] To avoid clutter, we do not graph states that have probability of less than 10^{-4}.

structure and then towards a symmetric structure. As we discuss in detail in the following section, the reason is that the well serves to build, but not to defend, a competitive advantage.

While the modes of the transient distributions are more separated and pronounced in the *trenchy equilibrium* ($\rho = 0.85$, $\delta = 0.0275$, see lower left panels) than in the flat equilibrium with well, the dynamics of the industry are similar at first. Unlike in the flat equilibrium with well, however, the industry continues to evolve towards an asymmetric structure. The modal states are $(14, 21)$ and $(21, 14)$ in period 64 and the modal states of the limiting distribution are $(21, 28)$ and $(21, 28)$. Despite cost parity, however, the leader is more secure against future losses from organizational forgetting than the follower. Asymmetries persist as time passes because the diagonal trench serves to build and to defend a competitive advantage.

In the *extra-trenchy equilibrium* ($\rho = 0.85$, $\delta = 0.08$, see lower right panels), one firm never makes it down from the top of its learning curve due to the sideways trench. The transient and limiting distributions are bimodal, and the most likely industry structure is extremely asymmetric. The modal states are $(1, 7)$ and $(7, 1)$ in period 8, $(1, 10)$ and $(10, 1)$ in period 16, $(1, 15)$ and $(15, 1)$ in period 32, and $(1, 19)$ and $(19, 1)$ in period 64. The modal states of the limiting distribution are $(1, 26)$ and $(26, 1)$. In short, one firm acquires a competitive advantage early on and maintains it with an iron hand.

Returning to Figure 3, our summary measures of industry concentration, the maximum expected Herfindahl index $H^{\wedge}$ (dashed line) and the limiting expected Herfindahl index H^{∞} (solid line), illustrate the fundamental economics of organizational forgetting. If organizational forgetting is sufficiently weak ($\delta \approx 0$), then asymmetries may arise but they cannot persist, i.e., $H^{\wedge} \geq 0.5$ and $H^{\infty} \approx 0.5$. Moreover, if asymmetries arise in the short run, they are modest. If organizational forgetting is sufficiently strong ($\delta \approx 1$), then asymmetries cannot arise in the first place, i.e., $H^{\wedge} \approx H^{\infty} \approx 0.5$. The reason is that organizational forgetting stifles investment in learning-by-doing altogether. By contrast, for intermediate degrees of organizational forgetting, asymmetries arise and persist. The asymmetry can be so pronounced that the leader is virtually a monopolist. This is because organizational forgetting predisposes the leader to defend its position aggressively. This more than offsets the increased vulnerability to organizational forgetting as the stock of know-how grows and therefore makes the leadership position more secure than it would have been in the absence of organizational forgetting.[22]

To summarize, contrary to what one might expect, organizational forgetting does not negate learning-by-doing. Rather, as can be seen in Figure 3, over a range of progress

[22]Since the Markov process is irreducible if $\delta \in (0, 1)$, it is inevitable that the follower eventually overtakes the leader. However, as a practical matter, the expected time to a role reversal is so large that this possibility may be disregarded.

ratios ρ above 0.6 and forgetting rates δ below 0.1, learning-by-doing and organizational forgetting reinforce each other. Starting from the absence of both learning-by-doing ($\rho = 1$) and organizational forgetting ($\delta = 0$), a steeper learning curve, i.e., a lower progress ratio, tends to lead to a more asymmetric industry structure just as a higher forgetting rate does. In the following section we analyze in more detail the pricing behavior that drives industry dynamics.

6 Pricing behavior

Re-writing equation (5) shows that firm 1's price in state $\mathbf{e}$ satisfies

$$p^*(\mathbf{e}) = c^*(\mathbf{e}) + \frac{\sigma}{1 - D_1^*(\mathbf{e})}, \tag{16}$$

where the *virtual* marginal cost

$$c^*(\mathbf{e}) = c(e_1) - \beta\phi^*(\mathbf{e})$$

equals the *actual* marginal cost $c(e_1)$ minus the discounted *prize* $\beta\phi^*(\mathbf{e})$ from winning the current period's sale. The prize, to be determined in equilibrium, is given by

$$\phi^*(\mathbf{e}) = \overline{V}_1^*(\mathbf{e}) - \overline{V}_2^*(\mathbf{e})$$

and has two components. First, by winning a sale, firm 1 may move further down its a learning curve. We call this the *advantage-building* motive. Second, firm 1 may prevent firm 2 from moving further down its learning curve. We call this the *advantage-defending* motive. Winning the sale in expectation is worth $\overline{V}_1^*(\mathbf{e})$ to firm 1 and losing it is worth $\overline{V}_2^*(\mathbf{e})$. Pricing behavior thus hinges on the difference between these values of continued play.

The prize $\phi^*(\mathbf{e})$ is the wedge that causes dynamic pricing behavior to differ from static pricing behavior. To see this, recall that the FOC of a static price-setting game can be written as

$$p^\dagger(\mathbf{e}) = c(e_1) + \frac{\sigma}{1 - D^\dagger(\mathbf{e})}, \tag{17}$$

where $D_k^\dagger(\mathbf{e}) = D_k(p^\dagger(\mathbf{e}), p^\dagger(\mathbf{e}^{[2]}))$ denotes the probability that, in the static Nash equilibrium, the buyer purchases from firm $k \in \{1, 2\}$ in state $\mathbf{e}$. Clearly, equation (16) reduces to equation (17) if either the firm is myopic ($\beta = 0$) or its prize is zero ($\phi^*(\mathbf{e}) = 0$). The difference in firms' pricing incentives depends on the difference in their virtual marginal costs. This difference, in turn, depends on the difference in their actual marginal costs and

the difference in their prizes.

6.1 Price bounds

Comparing equation (16) with equation (17) shows that equilibrium prices $p^*(\mathbf{e})$ and $p^*(\mathbf{e}^{[2]})$ coincide with the prices that obtain in a static Nash equilibrium with costs set to equal virtual marginal costs $c^*(\mathbf{e})$ and $c^*(\mathbf{e}^{[2]})$. Since in the static Nash equilibrium prices are increasing in either firm's cost (Vives 1999, p. 35) it follows that, as long as both firms' prizes are nonnegative, equilibrium prices are bounded above by static Nash equilibrium prices with costs set to equal actual marginal costs $c(e_1)$ and $c(e_2)$. More formally, if $\phi^*(\mathbf{e}) \geq 0$ and $\phi^*(\mathbf{e}^{[2]}) \geq 0$, then $p^*(\mathbf{e}) \leq p^\dagger(\mathbf{e})$ and $p^*(\mathbf{e}^{[2]}) \leq p^\dagger(\mathbf{e}^{[2]})$.

A sufficient condition for $\phi^*(\mathbf{e}) \geq 0$ for each state $\mathbf{e}$ is that the value function $V^*(e_1, e_2)$ is nondecreasing in e_1 and nonincreasing in e_2. Intuitively, it should not hurt firm 1 if it moves down its learning curve and it should not benefit firm 1 if firm 2 moves down its learning curve. This intuition is valid in the absence of organizational forgetting, and equilibrium prices are indeed bounded above by static Nash equilibrium prices:

Result 4 *If organizational forgetting is absent ($\delta = 0$), then we have $p^*(\mathbf{e}) \leq p^\dagger(\mathbf{e})$ for all $\mathbf{e} \in \{1, \ldots, M\}^2$.*

Result 4 highlights the fundamental economics of learning-by-doing: as long as improvements in competitive position are valuable, firms use price cuts as investments to achieve them.

The following proposition complements Result 4 by providing a lower bound on equilibrium prices:

Proposition 3 *If organizational forgetting is absent ($\delta = 0$), then we have (i) $p^*(\mathbf{e}) = p^\dagger(\mathbf{e}) = p^\dagger(m, m) > c(m)$ for all $\mathbf{e} \in \{m, \ldots, M\}^2$ and (ii) $p^*(\mathbf{e}) > c(m)$ for all $e_1 \in \{m, \ldots, M\}$ and $e_2 \in \{1, \ldots, m-1\}$.*

An immediate implication of part (i) of Proposition 3 is that diagonal trenches (and thus trenchy and extra-trenchy equilibria) cannot arise in the absence of organizational forgetting. In this case, prices are flat once both firms reach the bottom of their learning curves. To see this, note that given $\delta = 0$ the prize reduces to $\phi^*(\mathbf{e}) = V^*(e_1+1, e_2) - V^*(e_1, e_2+1)$. But once both firms reach the bottom of their learning curves, no further improvements in competitive position are possible. Hence, as we show in the proof of Proposition 3, we have $V^*(\mathbf{e}) = V^*(\mathbf{e}')$ for all $\mathbf{e}, \mathbf{e}' \in \{m, \ldots, M\}^2$, so that the advantage-building and advantage-defending motives disappear. Consequently, equilibrium prices coincide with prices in the static Nash equilibrium which, in turn, are set above cost.

If the leader but not the follower has reached the bottom of its learning curve, then the leader no longer has an advantage-building motive but he continues to have an advantage-defending motive. This raises the possibility that the leader uses price cuts to delay the follower's progress in moving down its learning curve. However, part (ii) of Proposition 3 shows that there is a limit to how aggressively the leader will defend its advantage: below-cost-pricing is never optimal in the absence of organizational forgetting.

In the presence of organizational forgetting pricing behavior can become much more intricate. To begin with, the intuition that the value function $V^*(e_1, e_2)$ is nondecreasing in e_1 and nonincreasing in e_2 is not always valid:

Result 5 *If organizational forgetting is present ($\delta > 0$), then we may have $p^*(\mathbf{e}) > p^\dagger(\mathbf{e})$ for some $\mathbf{e} \in \{1, \ldots, M\}^2$.*

Figure 8 illustrates Result 5 by plotting the share of equilibria that violate the upper bound on equilibrium prices in Result 4.[23] Darker shades indicate higher shares. As can be seen, Result 4 continues to hold if organizational forgetting is very weak ($\delta \approx 0$) and possibly also if learning-by-doing is very weak ($\rho \approx 1$). Apart from these extremes (and a region around $\delta = 0.25$ and $\rho = 0.45$), at least some, if not all, equilibria entail at least one state where equilibrium prices exceed static Nash equilibrium prices.

At first glance, Result 5 suggests that organizational forgetting makes firms less aggressive. This seems intuitive: After all, why invest in improvements in competitive position when they are transitory? Surprisingly, however, it turns out that organizational forgetting is a source of aggressive pricing behavior:

Result 6 *If organizational forgetting is present ($\delta > 0$), then we may have (i) $p^*(\mathbf{e}) < p^\dagger(\mathbf{e})$ for some $\mathbf{e} \in \{m, \ldots, M\}^2$ or (ii) $p^*(\mathbf{e}) \leq c(m)$ for some $e_1 \in \{m, \ldots, M\}$ and $e_2 \in \{1, \ldots, m-1\}$.*

Figure 9 depicts the share of equilibria that violate the lower bound on equilibrium prices in Proposition 3. As can be seen, unless organizational forgetting or learning-by-doing is very weak ($\delta \approx 0$), at least some, if not all, equilibria fail to obey Proposition 3. That is, the leader may be more aggressive in defending its advantage in the presence of organizational forgetting than in its absence. The most dramatic expression of this aggressive pricing behavior are the wells and trenches in the policy function.

[23] To take into account the limited precision of our computations, we take the upper bound in Result 4 to be violated if $p^*(\mathbf{e}) > p^\dagger(\mathbf{e}) + \epsilon$ for some $\mathbf{e} \in \{1, \ldots, M\}^2$, where ϵ is positive but small. Specifically, we set $\epsilon = 10^{-2}$, so that if prices are measured in dollars, then the upper bound must be violated by more than a cent. Given that the homotopy algorithm solves the system of equations up to a maximum absolute error of about 10^{-12}, Figure 8 therefore almost certainly understates the extent of violations.

	modal	leader				follower			
period	state	cost	prize	price	prob.	cost	prize	price	prob.
0	(1,1)	10.00	6.85	5.48	0.500	10.00	6.85	5.48	0.500
8	(8,1)	6.14	3.95	7.68	0.811	10.00	2.20	9.14	0.189
16	(11,4)	5.70	1.16	7.20	0.616	7.22	1.23	7.68	0.384
32	(14,9)	5.39	0.36	7.16	0.527	5.97	0.64	7.27	0.473
64	(17,17)	5.30	-0.01	7.31	0.500	5.30	-0.01	7.31	0.500
∞	(25,24)	5.30	-0.01	7.30	0.500	5.30	-0.00	7.30	0.500

Table 1: Cost, prize, price, and probability of making a sale. Flat equilibrium with well ($\rho = 0.85$, $\delta = 0.0275$).

6.2 Wells and trenches

In equilibrium the price set by a firm is a best reply to the price set by its rival in each possible state of the world. One might wonder, though, whether actual firms placed in the environment we have modeled behave in such a manner. Benkard's (2004) analysis of the commercial-aircraft market provides a hint that they might. Lockheed sold the L-1011 aircraft at a price below its average variable cost for much of its 14-year lifespan. Given the nontrivial estimates of the forgetting rate in Benkard (2000), Lockheed's actions are consistent with the pricing behavior of a firm in the midst of a well or a trench. This section provides intuition for wells and trenches in order to explore whether the pricing behavior we have characterized is economically plausible and empirically relevant.

Wells. A *well*, as seen in the upper right panel of Figure 4, is a preemption battle that is fought by firms at the top of their learning curves. A well serves to build a competitive advantage as both firms use price cuts in the hope of being the first to move down the learning curve. Once one firm has moved ahead of the other, both the leader and the follower raise their price. The follower, in fact, surrenders by setting a much higher price than the leader. Yet, once the follower starts to move down its learning curve, the leader makes no attempt to defend its position. The competitive advantage is thus of a transitory nature.

A well arises when the first sale has profound consequences for the evolution of the industry. Table 1 provides details on firms' competitive positions at various points in time for our leading example of a flat equilibrium with well ($\rho = 0.85$, $\delta = 0.0275$).[24] Being the first to move down the learning curve, the leader has a lower cost and a higher prize and therefore charges a lower price and enjoys a higher probability of making a sale than the

[24] In the remainder of this section we assume, without loss of generality, that firm 1 is the leader and firm 2 the follower.

	leader				follower			
state	cost	prize	price	prob.	cost	prize	price	prob.
$(21, 21)$	5.30	2.14	5.26	0.50	5.30	2.14	5.26	0.50
$(21, 20)$	5.30	3.53	5.57	0.72	5.30	0.14	6.54	0.28
$(22, 20)$	5.30	3.22	6.44	0.76	5.30	-1.04	7.60	0.24
$(20, 20)$	5.30	2.16	5.24	0.50	5.30	2.16	5.24	0.50

Table 2: Cost, prize, price, and probability of making a sale. Trenchy equilibrium ($\rho = 0.85$, $\delta = 0.0275$).

follower in the modal state $(8, 1)$ in period 8. As time passes and the follower moves down its learning curve, the competitive advantage of the leader begins to erode (see the modal state $(11, 4)$ in period 16) and eventually vanishes completely (see the modal state $(17, 17)$ in period 64). This erosion of the competitive advantage of the leader is reflected in the prize: While the leader's prize in state $(8, 1)$ is higher than the follower's (3.95 vs. 2.20), in state $(11, 4)$ the leader's prize is lower than the follower's (1.16 vs. 1.23). Although the competitive advantage is transitory, it is surely worth having: The prize in state $(1, 1)$ is 6.85 and justifies charging a price of 5.48 that is well below the marginal cost of 10. The well is therefore an investment in building a competitive advantage. It is deep to the extent that the competitive advantage can be sustained for at least some time.

Diagonal trenches. A *diagonal trench*, as seen in the lower panels of Figure 4, is a price war between symmetric or nearly symmetric firms. Like a well, a diagonal trench serves to build a competitive advantage. Unlike a well, however, a diagonal trench also serves to defend it. A diagonal trench is about acquiring and maintaining a permanent competitive advantage. Aggressive pricing is not confined to the top of the learning curve. On the contrary, it takes place all along the diagonal of the state space as each firm uses price cuts to push the state to "its" side of the diagonal and keep it there. A curious feature of a diagonal trench is that firms compete fiercely even though they have already exhausted the gains from learning-by-doing.

We can employ backward-induction-like logic, as illustrated in Figure 10, to gain intuition about the link between organizational forgetting and diagonal trenches. Consider state (e, e), where $e \geq m$, on the diagonal of the state space at or beyond the bottom of the learning curve. From part (i) of Proposition 3, without organizational forgetting, the advantage-building and advantage-defending motives disappear and equilibrium prices coincide with prices in the static Nash equilibrium. However, with organizational forgetting, the advantage-building and advantage-defending motives continue to operate. The advantage-building motive operates in state (e, e) because by winning a sale, the firm creates a "buffer

stock" of know-how against future losses from organizational forgetting. The advantage-defending motive operates because by winning a sale, the firm increases the likelihood that its rival slides back up its learning curve. Thus, organizational forgetting predisposes firms to compete fiercely even though they have already exhausted the gains from learning-by-doing. Table 2 illustrates this point by providing details on firms' competitive positions in various states for our leading example of a trenchy equilibrium ($\rho = 0.85$, $\delta = 0.0275$). As can be seen, the prize in state $(21, 21)$ is 2.14 and justifies charging a price of 5.26 that is a little below the marginal cost of 5.30 and a lot below the static Nash equilibrium price of 7.30.

Next consider state $(e, e-1)$ where firm 1 has a slight lead over firm 2 (see Figure 10). With organizational forgetting, the leader's prize from winning a sale is likely larger than the follower's. In our leading example, the leader's prize in state $(21, 20)$ is almost 25 times larger than the follower's. This indicates that winning a sale is considerably more valuable to the leader than the follower.

To see why, suppose the follower wins. In this case the follower may leapfrog the leader if the industry moves to state $(e-1, e)$, but the odds are against this possibility (by about $3:1$ in our leading example). The most likely possibility is that the industry moves back to state (e, e) where there is brutal price competition. In our leading example, such a move to state $(21, 21)$ happens with a probability of 0.32 and, if so, then the follower's expected cash flow in the next period decreases to $-0.02 = 0.50 \times (5.26 - 5.30)$ compared to $0.34 = 0.28 \times (6.54 - 5.30)$ if the industry were to remain in state $(21, 20)$.

By contrast, if the leader wins, then the industry is guaranteed to not move back to state (e, e), thereby avoiding the brutal price competition in this state. The leader also, of course, avoids the possibility of losing its competitive advantage by precluding movements to states such as $(e-1, e)$ and $(e-1, e-1)$. And the leader may even enhance its competitive advantage by inducing movements to states such as $(e, e-2)$, $(e+1, e-1)$, and $(e+1, e-2)$. The most likely possibility is that the industry moves to state $(e+1, e-1)$. In our leading example, such a move to state $(22, 20)$ happens with a probability of 0.32 and, if so, the leader's expected cash flow in the next period increases to $0.87 = 0.76 \times (6.44 - 5.30)$ compared to $0.20 = 0.72 \times (5.57 - 5.30)$ if the industry were to remain in state $(21, 20)$.

To the extent that the leader's prize from winning a sale is larger than the follower's, the leader has a large advantage over the follower in terms of virtual marginal cost. Hence, the leader substantially underprices the follower. As a result the leadership position in state $(e, e-1)$ is more secure in the presence of organizational forgetting than in its absence. In our leading example, the probability that the leader defends its position is 0.79 (compared to 0.62 in the absence of organizational forgetting).

To complete the induction, consider finally state $(e-1, e-1)$ on the diagonal of the

state space just below state (e, e) (see again Figure 10). Because the leadership position in state $(e, e-1)$ is so secure, both firms fight hard to attain it, thereby intensifying price competition in state $(e-1, e-1)$. In our leading example, the prize in state $(20, 20)$ is 2.16 (compared to 2.14 in state $(21, 21)$) and justifies charging a price of 5.24 (compared to 5.26). These effects cascade through the state space and give rise to diagonal trenches.

By fighting the price war in the diagonal trench the leader is able to acquire and maintain a permanent competitive advantage. In our leading example, the modal states of the limiting distribution are $(21, 28)$ and $(28, 21)$ where both firms are safely beyond the bottom of their learning curve and cost parity obtains. Yet, the leader underprices the follower (7.63 vs. 7.81) and thus enjoys higher probability of making a sale (0.55 vs. 0.45). The follower recognizes that to attain a competitive advantage it will have to "cross over" the diagonal trench. This discouraging prospect weakens the follower's motivation to improve its competitive position and, in turn, strengthens the leader's motivation to preserve or strengthen its competitive position. What is critical about this price war is that it is a part of a Markov perfect equilibrium and, as such, is a credible threat the follower cannot ignore.

Interestingly enough, our leading example violates both the upper and the lower bound on equilibrium prices in Result 4 and Proposition 3, respectively, that obtain in the absence of organizational forgetting: Whereas both firms price higher in the most likely long-run industry structure than in the static Nash equilibrium (e.g., in states $(21, 28)$ and $(28, 21)$), "in the trench" they price lower (e.g., in states $(20, 20)$ and $(21, 21)$).[25] This illustrates Results 5 and 6 and, in turn, shows that in the presence of organizational forgetting the equilibrium may entail lackluster competition in some states as well as price wars in other states.

In sum, diagonal trenches are self-reinforcing mechanisms that lead to market dominance: if the leadership position is aggressively defended, symmetric firms fight a price war to attain it. This provide all the more reason to aggressively defend a competitive advantage because if it is lost and the industry moves back to the diagonal of the state space, then another price war ensues.

Sideways trenches. A sideways trench, as seen in the lower right panel of Figure 4, is a price war between very asymmetric firms. It is triggered when the follower starts to move down its learning curve. Similar to a diagonal trench, a sideways trench is about acquiring and maintaining a permanent competitive advantage. However, while a diagonal trench is about fighting an imminent threat, a sideways trench is about fighting a distant threat. Put informally, a sideways trench is the equilibrium manifestation of former Intel CEO Andy Grove's dictum "Only the paranoid survive." One can think of a sideways trench as an

[25] Recall that the static Nash equilibrium price is 7.30 is these states.

	leader				follower			
state	cost	prize	price	prob.	cost	prize	price	prob.
$(26, 1)$	5.30	6.43	8.84	0.90	10.00	0.12	11.00	0.10
$(26, 2)$	5.30	6.21	7.48	0.88	8.50	0.21	9.44	0.12
$(26, 3)$	5.30	5.16	6.94	0.85	7.73	0.27	8.65	0.15
$(26, 4)$	5.30	4.28	6.66	0.82	7.23	0.31	8.15	0.18
$(26, 5)$	5.30	3.70	6.49	0.79	6.86	0.34	7.81	0.21
$(26, 6)$	5.30	3.36	6.34	0.76	6.57	0.39	7.51	0.24
$(26, 7)$	5.30	3.04	6.14	0.73	6.34	0.58	7.15	0.27
$(26, 8)$	5.30	2.33	5.99	0.66	6.14	1.08	6.64	0.34
$(26, 9)$	5.30	1.17	6.24	0.51	5.97	1.71	6.29	0.49
$(26, 10)$	5.30	0.16	6.83	0.40	5.83	1.96	6.44	0.60

Table 3: Cost, prize, price, and probability of making a sale. Extra-trenchy equilibrium ($\rho = 0.85$, $\delta = 0.08$).

endogenously arising mobility barrier in the sense of Caves & Porter (1977). If the follower crashes through this mobility barrier—an unlikely event—he moves from being a docile competitor to being a viable threat. To prevent this from happening, the leader stalls the follower.

Table 3 provides details on firms' competitive positions in various states for our leading example of an extra-trenchy equilibrium ($\rho = 0.85$, $\delta = 0.08$). Recall that the modal state of the limiting distribution is $(26, 1)$. The sideways trench is evident in the decrease in the price charged by the leader between states $(26, 1)$ and $(26, 8)$ and the increase between states $(26, 8)$ and $(26, 10)$. To see how the leader can stall the follower, note that the follower has little chance of making it down its learning curve as long as the probability of winning a sale is less than the probability of losing a unit of know-how through organizational forgetting. Indeed, while $D_2^*(26, 1) = 0.10 > 0.08 = \Delta(1)$, we have $D_2^*(26, 2) = 0.12 < 0.15 = \Delta(2)$ and $D_2^*(26, 3) = 0.15 < 0.22 = \Delta(3)$. The ability of the leader to stall the follower at the top of its learning curve is reflected in the large prize in state $(26, 1)$. In state $(26, 2)$ the prize is almost as large because by winning a sale the leader may move the industry back to state $(26, 1)$ in the next period. The prize falls in case the follower moves further down its learning curve because it takes progressively longer for the leader to force the follower back up its learning curve and because the lower cost of the follower makes it harder for the leader to do so. In the unlikely event that the follower crashes through the sideways trench in state $(26, 8)$, the prize of the leader falls sharply. At the same time the prize of the follower rises sharply because the follower turns from a docile competitor into a viable threat with an equal or larger share of the market.

Again, sideways trenches are self-reinforcing mechanisms that lead to market dominance:

ρ	0.95	0.85	0.75	0.65	0.55	0.35	0.15	0.05
$\bar{\delta}(\rho)$	0.55	0.60	0.62	0.71	0.78	0.81	0.88	0.90

Table 4: Critical value $\bar{\delta}(\rho)$ for investment stifling.

If organizational forgetting enables the leader to stall the follower in state (e_1, e_2), then this creates strong incentives for the leader to cut prices in states like (e_1, e_2+1) and (e_1, e_2+2) in order to deny the follower a sale and force it back up its learning curve. But because being in states like (e_1, e_2+1) and (e_1, e_2+2) is not especially attractive, the follower has little reason to cut price in state (e_1, e_2). This makes it even easier for the leader to stall the follower.

In sum, the four types of equilibria that we have identified in Section 4 give rise to distinct yet plausible pricing behaviors and, in turn, industry dynamics as discussed in Section 5. Rather than impeding it, organizational forgetting facilitates aggressive behavior. In the absence of organizational forgetting, the equilibria in our computations have always been flat either without or with well depending on the progress ratio. Moreover, the lower bound on equilibrium prices in part (i) of Proposition 3 rules out diagonal trenches (and thus trenchy and extra-trenchy equilibria). Indeed, in the absence of organizational forgetting, prices are flat once both firms reach the bottom of their learning curves. Similarly, if the leader but not the follower has reached the bottom of its learning curve, then part (ii) of Proposition 3 shows that the leader will not be overly aggressive in defending its advantage. This limits the scope for sideways trenches. In the presence of organizational forgetting, in contrast, firms may compete fiercely even though they have already exhausted the gains from learning-by-doing. Generally speaking, organizational forgetting is associated with "trenchier" equilibria, more aggressive behavior, and more concentrated industries both in the short run and in the long run.

There is clearly a limit to the enhancement of price-cutting incentives through organizational forgetting. If the forgetting rate δ is very large, then organizational forgetting stifles investment in learning-by-doing altogether. In particular, if δ exceeds the critical value $\bar{\delta}(\rho)$ listed in Table 4, then firms cannot expect to make it down their learning curves.[26] Hence, equilibrium prices at the top of the learning curve are close to prices in the static Nash equilibrium. Even then, however, price competition at the bottom of the learning curve is extremely intense as both firms seek to reduce the chance of being the first to slide back up the learning curve.

[26] We take $\bar{\delta}(\rho)$ to be the smallest forgetting rate such that state $(1,1)$ is the mode of the limiting distribution.

6.3 Dominance properties

Traditional intuition suggests that learning-by-doing leads by itself to market dominance by giving a more experienced firm the ability to profitably underprice its less experienced rival. This enables the leader to widen its competitive advantage over time, thereby further enhancing its ability to profitably underprice the follower. Cabral & Riordan (1994) formalize this idea with "two concepts of self-reinforcing market dominance" (p. 1115), increasing dominance (ID) and increasing increasing dominance (IID). An equilibrium exhibits ID if $p^*(\mathbf{e}) < p^*(\mathbf{e}^{[2]})$ whenever $e_1 > e_2$ and IID if $p^*(\mathbf{e}) - p^*(\mathbf{e}^{[2]})$ is decreasing in e_1. If ID holds, the leader charges a lower price than the follower and therefore enjoys a higher probability of making a sale. If IID holds, the gap between the leader's price and the follower's price widens with the length of the lead. Athey & Schmutzler (2001) extend this idea to dynamic games with deterministic state-to-state transitions. Their notion of weak increasing dominance describes the relationship between players' states and their actions and coincides with Cabral & Riordan's (1994) notion of ID.[27]

In the absence of organizational forgetting, Cabral & Riordan (1994) show that ID and IID hold provided that the discount factor β is sufficiently close to 1 (or, alternatively, sufficiently close to 0). Their main result carries over to our parameterization with $\beta = \frac{1}{1.05}$:

Result 7 *(i) If organizational forgetting is absent ($\delta = 0$), then IID holds. (ii) Thus, ID holds.*

Even though the equilibrium satisfies ID and IID, it is not clear that the industry is inevitably progressing towards monopolization. If the price gap is small, then the effect of ID and IID may be trivial.[28] In such a scenario, the leader charges a *slightly* lower price than the follower and this gap widens *a bit* over time. However, with even a modest degree of horizontal product differentiation, the firms still split sales more or less equally and thus move down the learning curve in tandem. Consequently, ID and IID may have no discernible impact on industry structure and dynamics, either in the short run or in the long run. This is exactly what happens in the absence of organizational forgetting. For example, the flat equilibrium without well ($\rho = 0.85$, $\delta = 0$) satisfies IID and thus ID. Yet, the industry is likely to be a symmetric duopoly at all times. More generally, as Figure 3 shows, in the absence of organizational forgetting asymmetries are modest if they arise at all. In fact, although ID and IID hold, the maximum expected Herfindahl index is 0.67

[27] Similar notions of increasing dominance have also been used by Vickers (1986) and Budd, Harris & Vickers (1993) in dynamic investment games.

[28] Indeed, Cabral & Riordan (1994) show that $p^*(\mathbf{e}) \to p^\dagger(m, m)$ for all $\mathbf{e} \in \{1, \ldots, M\}^2$ as $\beta \to 1$, i.e., both firms price *as if* at the bottom of their learning curves. This suggests that the price gap may be small for "reasonable" discount factors.

(attained at $\rho = 0.65$). Hence, ID and IID are not sufficient for economically meaningful market dominance.

ID and IID are also not necessary for market dominance. To give an example, the extra-trenchy equilibrium ($\rho = 0.85$, $\delta = 0.08$) violates ID and thus IID because the leader coasts by charging a higher price if it is far ahead of the follower. Yet, the industry is likely to be a near-monopoly at all times. More generally, while the empirical studies of Argote et al. (1990), Darr et al. (1995), Benkard (2000), Shafer et al. (2001), and Thompson (2003) warrant the inclusion of organizational forgetting in a model of learning-by-doing, ID and IID may fail in its presence:

Result 8 *If organizational forgetting is present ($\delta > 0$), then (i) IID may fail and (ii) ID may fail.*

Figure 11 illustrates Result 8 by plotting the share of equilibria that violate IID (upper panel) and ID (lower panel). As can be seen, all equilibria fail to obey IID unless organizational forgetting or learning-by-doing is very weak. Even violations of ID are extremely common, especially for forgetting rates δ in the empirically relevant range below 0.1.

Of course, we do not mean to argue that the concepts of ID and IID have no place in the analysis of industry dynamics. It appears, however, that caution is warranted: Since ID and IID are neither necessary nor sufficient for economically meaningful market dominance, making inferences about the evolution of the industry on the basis of ID and IID alone may be misleading. We therefore suggest that these concepts be augmented by a search for tell-tale signs of market dominance such as wells and trenches and, perhaps even more important, by a direct examination of the industry dynamics implied by firms' pricing behavior.

6.4 Summary

Table 5 summarizes the broad patterns of pricing behavior and industry dynamics. Acknowledging that the know-how gained through learning-by-doing can be lost through organizational forgetting is evidently important. Generally speaking, organizational forgetting is associated with "trenchier" equilibria, more aggressive behavior, and more concentrated industries both in the short run and in the long run. Moreover, the dominance properties of firms' pricing behavior can break down in the presence of organizational forgetting.

The key difference between a model with learning-by-doing and organizational forgetting and a model with learning-by-doing alone is that in the former a firm can move both forward to a higher state and backward to a lower state whenever $\delta \in (0, 1)$. The possibility of bidirectional movements through the state space opens up new strategic possibilities for firms that work to enhance the advantage-building and advantage-defending motives. By winning

	flat eqbm. without well	flat eqbm. with well	trenchy eqbm.	extra-trenchy eqbm.
leading example	$\rho = 0.85$, $\delta = 0$	$\rho = 0.85$, $\delta = 0.0275$	$\rho = 0.85$, $\delta = 0.0725$	$\rho = 0.85$, $\delta = 0.08$
preemption battle (well)	no	yes	no	no
price war triggered by imminent threat (diagonal trench)	no	no	yes	yes
price war triggered by distant threat (sideways trench)	no	no	no	yes
short-run market dominance	no	yes	yes	yes
long-run market dominance	no	no	yes, modest	yes, extreme
dominance properties	yes	no, mostly	no, mostly	no, mostly

Table 5: Pricing behavior and industry dynamics.

a sale, a firm makes itself less vulnerable to future losses from organizational forgetting, thus enhancing the advantage-building motive. At the same time, it makes its rival more vulnerable to future losses from organizational forgetting, thus enhancing the advantage-defending motive. Because these additional benefits (like the benefits from learning-by-doing) are achieved by winning a sale, organizational forgetting can create strong incentives to cut prices. Thus, rather than impeding it, organizational forgetting can facilitate aggressive pricing behavior that manifests itself in wells and trenches.

7 Organizational forgetting and multiple equilibria

While the equilibrium is unique if organizational forgetting is either absent ($\delta = 0$) or certain ($\delta = 1$), multiple equilibria are common for intermediate forgetting rates. Surprisingly, these equilibria range from "peaceful coexistence" to "trench warfare." Consequently, in addition to the degree of organizational forgetting, the equilibrium by itself is an important determinant of pricing behavior and industry dynamics.

Why do multiple equilibria arise in our model? To explore this question, think about the strategic situation faced by firms in setting prices in state $\mathbf{e}$. The value of continued play to firm n is given by the conditional expectation of its value function, $\overline{V}_{n1}(\mathbf{e})$ and $\overline{V}_{n2}(\mathbf{e})$. Holding the value of continued play fixed, the strategic situation in state $\mathbf{e}$ is thus akin to a static game. If the reaction functions in this game intersect more than once, then multiple equilibria arise. On the other hand, we say that the model satisfies stagewise uniqueness if

the reactions functions of the two firms intersect once irrespective of the value of continued play. This is indeed the case:

Proposition 4 *Statewise uniqueness holds.*

Note that the proof of Proposition 4 relies on the functional form of demand. This is reminiscent of the restrictions on demand (e.g., log-concavity) that Caplin & Nalebuff (1991) set forth to guarantee uniqueness of Nash equilibrium in their analysis of static price-setting games.

Given that the model satisfies statewise uniqueness, multiple equilibria must arise from firms' expectations regarding the value of continued play. To see this, consider again state $\mathbf{e}$. The intersection of the reaction functions constitutes a Nash equilibrium in prices in a subgame in which firm n believes that its value of continued play is given by $\overline{V}_{n1}(\mathbf{e})$ and $\overline{V}_{n2}(\mathbf{e})$. If firms have rational expectations, i.e., if the conjectured value of continued play is actually attained, then these prices constitute an equilibrium of our dynamic stochastic game. In our model, taking the value of continued play as given, the reaction functions intersect once because we have statewise uniqueness, but there may be more than one value of continued play that is consistent with rational expectations. In this sense multiplicity is rooted in the dynamics of the model.

The key driver of multiplicity is organizational forgetting. Dynamic competition with learning-by-doing and organizational forgetting is like racing down an upward-moving escalator. Unless a firm makes sales at a rate that exceeds the rate at which it loses know-how through organizational forgetting, its marginal cost is bound to increase. The inflow of know-how into the industry is one unit per period whereas in expectation the outflow in state $\mathbf{e}$ is $\Delta(e_1) + \Delta(e_2)$. Consider state (e, e), where $e \geq m$, on the diagonal of the state space at or beyond the bottom of the learning curve. If $1 \ll 2\Delta(e)$, then it is impossible that both firms reach the bottom of their learning curves and remain there. Knowing this, firms have no choice but to price aggressively. The result is trench warfare as each firm uses price cuts to push the state to its side of the diagonal and keep it there. If, however, $1 \gg 2\Delta(e)$, then it is virtually inevitable that both firms reach the bottom of their learning curves, and firms may as well price softly. In both cases, the primitives of the model tie down the equilibrium.

This is no longer the case if $1 \approx 2\Delta(e)$, setting the stage for multiple equilibria as diverse as peaceful coexistence and trench warfare. If firms believe that they cannot peacefully coexist at the bottom of their learning curves and that one firm will come to dominate the market, then both firms will cut their prices in the hope of acquiring a competitive advantage early on and maintaining it throughout. This naturally leads to trench warfare and market dominance. If, however, firms believe that they can peacefully coexist at the

bottom of their learning curves, then neither firm cuts its price. Soft pricing, in turn, ensures that the anticipated symmetric industry structure actually emerges. A back-of-the-envelope calculation is reassuring here. Recall that $m = 15$ and $M = 30$ in our parameterization and observe that $1 = 2\Delta(15)$ implies $\delta \approx 0.05$, $1 = 2\Delta(20)$ implies $\delta \approx 0.03$, and $1 = 2\Delta(30)$ implies $\delta \approx 0.02$. This range of forgetting rates for which the inflow of know-how approximately equals the outflow is indeed where multiplicity prevails (see again Figure 2).

In general, a sufficient condition for uniqueness of equilibrium in a dynamic stochastic game with a finite state space is that the model satisfies statewise uniqueness and the movements through the state space are unidirectional. Statewise uniqueness precludes players' actions from giving rise to multiple equilibria and unidirectional movements preclude their expectations from doing so. To illustrate, recall that in the game at hand, a firm can never move backward to a lower state if $\delta = 0$. Hence, once the industry reaches state (M, M), it remains there forever, so that the value of future play in state (M, M) coincides with the value of being in this state *ad infinitum*. In conjunction with statewise uniqueness, this uniquely determines the value of being in state (M, M). Next consider states $(M-1, M)$ and $(M, M-1)$. The value of future play in states $(M-1, M)$ and $(M, M-1)$ depends on the value of being in state (M, M). Statewise uniqueness ensures that firms' prices in states $(M-1, M)$ and $(M, M-1)$ as well as the value of being in these states are uniquely determined. Continuing to work backwards in this fashion establishes that the equilibrium is unique. If $\delta = 1$, then a firm can never move forward to a higher state and a similar argument anchored on state $(1, 1)$ establishes uniqueness of equilibrium.[29]

8 Robustness checks

In this section, we discuss how the model specification and parameterization affect our results. In the interest of brevity, we confine ourselves to summarizing our robustness checks; the underlying figures and tables can be found in the Online Appendix.

8.1 Product differentiation

Our baseline parameterization gives rise to a moderate degree of horizontal product differentiation. In the static Nash equilibrium, the own-price elasticity of demand ranges between -8.86 in state $(1, 15)$ and -2.13 in state $(15, 1)$ for a progress ratio of $\rho = 0.85$. The cross-price elasticity of firm 1's demand with respect to firm 2's price is 2.41 in state $(15, 1)$ and 7.84 in state $(1, 15)$. As σ is decreased from 1 to 0.2, the respective elasticities become -102.00, -0.00, 0.00, and 55.0. As σ increases from 1 over 2 to 10, the respective

[29] Proposition 2 provides an example of this type of proof.

elasticities become -4.38, -1.86, 2.10, and 3.88 in case of $\sigma = 2$ and -1.54, -1.24, 1.32, and 1.45 in case of $\sigma = 10$.

In case of weaker product differentiation with $\sigma = 0.2$, trenchier equilibria lead to more asymmetric industry structures. Conversely, in case of stronger product differentiation with $\sigma = 2$, we obtain more symmetric industry structures. With $\sigma = 10$, firms hardly compete any more and sales are split more or less equally between them. Multiple equilibria no longer arise because firms are essentially monopolists that do not interact strategically with each other. Finally, a higher degree of horizontal product differentiation makes the market that a firm has effective access to smaller, so that it becomes easier for organizational forgetting to stifle investment in learning-by-doing altogether.

8.2 Outside good

We allow the buyer to choose an alternative made from a substitute technology (outside good 0) instead of purchasing from one of the two firms (inside goods 1 and 2). The probability that firm n makes a sale becomes

$$D_n(\mathbf{p}) = \frac{\exp(\frac{v-p_n}{\sigma})}{\exp(\frac{v_0-c_0}{\sigma}) + \sum_{k=1}^{2} \exp(\frac{v-p_k}{\sigma})},$$

where we assume that the outside good is supplied under conditions of perfect competition with price equal to marginal cost, $p_0 = c_0$. As $v_0 - c_0 \to -\infty$, $D_0(\mathbf{p}) = 1 - \sum_{n=1}^{2} D_n(\mathbf{p}) \to 0$ and we revert to the Cabral & Riordan (1994) setting in which the buyer always purchases from one of the two firms in the industry.

If we set $v = 10$ and $v_0 - c_0 = 0$, then $v - c(1) = v_0 - c_0$ and a firm at the top of its learning curve is on par with the outside good. The share of the outside good is quite small in general. In the static Nash equilibrium, as the marginal cost of production declines, the share of the outside good declines from 0.63 in state $(1, 1)$ over 0.33 in state $(2, 2)$ and 0.15 in state $(4, 4)$ to 0.03 in state $(15, 15)$ for a progress ratio of $\rho = 0.85$. To further increase the attractiveness of the outside good we set $v_0 - c_0 \in \{3, 5, 10\}$. If $v_0 - c_0 = 5$, then the share of the outside good is quite large in general and declines from 1.00 in state $(1, 1)$ over 0.99 in state $(2, 2)$ and 0.93 in state $(4, 4)$ to 0.69 in state $(15, 15)$.

Multiple equilibria continue to arise in the presence of an outside good, although less frequently as the outside good becomes more attractive. In particular, we no longer have sunspots for a progress ratio of $\rho = 1$ because the outside good sufficiently constrains firms' pricing behavior even with $v_0 - c_0 = 0$. With $v_0 - c_0 = 10$ the equilibrium is unique because almost all consumers choose the outside good, so that the inflow of know-how into the industry is much smaller than the outflow. Finally, a more attractive outside good *de facto* makes the market smaller, so that it becomes easier for organizational forgetting to stifle

investment in learning-by-doing altogether.

8.3 Choke price

In the absence of organizational forgetting the equilibria in our computations have always been flat either without or with well depending on the progress ratio. As in Cabral & Riordan (1994) our logit specification for demand ensures that a firm always has a positive probability of making a sale and therefore must eventually reach the bottom of its learning curve. A natural concern is whether this stacks the deck against long-run market dominance to occur in the absence of organizational forgetting. To explore this issue, we assume that the probability that firm n makes a sale is given by the linear specification

$$D_n(\mathbf{p}) = \min\left(\max\left(\frac{1}{2} - \frac{1}{4\sigma}(p_n - p_{-n}), 0\right), 1\right).^{30}$$

Note that due to the choke price in the linear specification, a firm is able to surely deny its rival a sale by pricing sufficiently aggressively.

For linear demand with $\sigma = 1$ our computations show that the industry evolves towards a symmetric structure in the absence of organizational forgetting ($\delta = 0$). With $\sigma = 0.2$, in contrast, firms at the top of their learning curves fight a preemption battle. The industry remains in an asymmetric structure as the winning firm takes advantage of the choke price to stall the losing firm at the top of its learning curve. Yet, we never found a trenchy or extra-trenchy equilibrium with linear demand in the absence of organizational forgetting. We are therefore confident that the flat equilibria that arise in the absence of organizational forgetting are not an artifact of the lack of a choke price with logit demand. At the same time, we continued to find trenchy and extra-trenchy equilibria with linear demand in the presence of organizational forgetting. Organizational forgetting has thus the same dramatic effect on firms' pricing behavior whether demand is logit or linear.

8.4 Frequency of sales

Following Cabral & Riordan (1994) we take a period to be just long enough for a firm to make a sale. One might wonder whether the insights of the analysis are sensitive to this assumption. To explore this issue without fundamentally departing from our modeling framework, we divide a period into $K > 1$ subperiods. Assuming that one sale occurs in a subperiod, K sales occur in a period. If r is the discount rate per period, then $\frac{r}{K}$ is the discount rate per subperiod and $\beta = \frac{1}{1+\frac{r}{K}}$ the discount factor. We have to be careful not to inadvertently change the properties of learning-by-doing and organizational

[30] To allow for a fair comparison between linear and logit demand, we choose the slope parameter so that in the static Nash equilibrium the own-price elasticity of demand in state $(1, 1)$ is the same.

period	subperiod	flat eqbm. without well	flat eqbm. with well	trenchy eqbm.	extra-trenchy eqbm.
8	16	$(5,5)$	$(3,6.5)$, $(6.5,3)$	$(3,6.5)$, $(6.5,3)$	$(1,6.5)$, $(6.5,1)$
16	32	$(9,9)$	$(5.5,10)$, $(10,5.5)$	$(6,9.5)$, $(9.5,6)$	$(1,10)$, $(10,1)$
32	64	$(17,17)$	$(11.5,14)$, $(14,11.5)$	$(11,14.5)$, $(14.5,11)$	$(1,14)$, $(14,1)$
64	128	$(30,30)$	$(19.5,20)$, $(20,19.5)$	$(18,21.5)$, $(21.5,18)$	$(1,18)$, $(18,1)$
∞	∞	$(30,30)$	$(29,30)$, $(30,29)$	$(26,30)$, $(30,26)$	$(1,24)$, $(24,1)$

Table 6: Frequency of sales with $K = 2$. Modal states of transient and limiting distributions.

forgetting by changing the frequency of sales. For example, the reduction in marginal cost that is achievable by a period's worth of sales in the original specification has to be comparable to the reduction that is achievable by K subperiods' worth of sales in the alternative specification. To accomplish this, we take the state space to be $\{1,\ldots,K(M-1)+1\}^2$. The marginal cost and probability of forgetting of firm n in the alternative specification are given by $c\left(\frac{e_n-1}{K}+1\right)$ and $\Delta\left(\frac{e_n-1}{K}+1\right)$. Finally, we take $K(m-1)+1$ to be the stock of know-how at which a firm reaches the bottom of its learning curve.

We have computed equilibria for a progress ratio of $\rho = 0.85$ while doubling the frequency of sales by setting $K = 2$. We obtain a flat equilibrium without well ($\delta = 0$), a flat equilibrium with well ($\delta = 0.02$), a trenchy equilibrium ($\delta = 0.02$), and an extra-trenchy equilibrium ($\delta = 0.09$), similar to the four typical cases in Figure 4. Table 6 lists the modal states of the implied transient and limiting distributions. As can be seen, industry structure and dynamics are comparable to those in Figures 5–7.[31] Overall, it appears that our results are not sensitive to the frequency of sales.

8.5 Learning-by-doing

Following Cabral & Riordan (1994) we assume that $m < M$ represents the stock of know-how at which a firm reaches the bottom of its learning curve. To check the robustness of our results, we instead assume $m = M$. In this *bottomless learning* specification, we obtain another type of equilibrium in addition to the four typical cases in Figure 4. Figure 12 exemplifies the policy function of this *plateau equilibrium* (upper left panel), the transient distribution in period 8 and 32 (upper right and lower left panels), and the limiting distribution (lower right panel). The parameter values are $\rho = 0.9$ and $\delta = 0.04$.

As can be seen, the plateau equilibrium is similar to a trenchy equilibrium except that the diagonal trench is interrupted by a region (around state $(17,17)$) of very soft price

[31] With the possible exception of the extra-trenchy equilibrium: While the modal states of the limiting distribution are $(1,24)$ and $(24,1)$ with a probability of 0.0086 each, the limiting distribution also has secondary peaks at states $(5.5,9.5)$ and $(9.5,5.5)$ with a probability of 0.0068 each. That is, there is some chance that the industry does not become extremely asymmetric as it does in our baseline parameterization with $K = 1$.

	asymmetric		symmetric	
period	state	prob.	state	prob.
8	$(2,7)$, $(7,2)$	0.0782	–	–
16	$(4,10)$, $(10,4)$	0.0357	–	–
32	$(6,14)$, $(14,6)$	0.0192	–	–
64	$(8,20)$, $(20,8)$	0.0143	$(15,15)$	0.0017
∞	$(11,25)$, $(25,11)$	0.0111	$(17,17)$	0.0072

Table 7: Bottomless learning. Most-likely asymmetric and symmetric states of transient and limiting distributions. Plateau equilibrium ($\rho = 0.9$, $\delta = 0.04$).

	asymmetric		symmetric	
period	state	prob.	state	prob.
8	–	–	$(4,5)$, $(5,4)$	0.0803
16	–	–	$(6,7)$, $(7,6)$	0.0434
32	$(4,17)$, $(17,4)$	0.0007	$(10,10)$	0.0312
64	$(8,21)$, $(21,8)$	0.0013	$(13,14)$, $(14,13)$	0.0247
∞	$(15,21)$, $(21,15)$	0.0075	–	–

Table 8: Bottomless learning. Most-likely asymmetric and symmetric states of transient and limiting distributions. Plateau equilibrium ($\rho = 0.9$, $\delta = 0.04$).

competition. On this plateau both firms charge prices well above cost. This "cooperative" behavior contrasts markedly with the price war of the diagonal trench. While the most-likely industry structure is asymmetric in the long run in this example, there is also a substantial probability that the industry becomes symmetric: The modal states of the limiting distribution are $(11,25)$ and $(25,11)$ with a probability of 0.0111 each. Yet, the limiting distribution also has a secondary peak at state $(17,17)$ with a probability of 0.0072. Table 7 summarizes the dynamics of the industry by providing the most-likely asymmetric and symmetric states of the transient and limiting distributions. As can be seen, the likelihood of cooperation goes up with time.

Table 8 provides another example of a plateau equilibrium. The parameter values are the same ($\rho = 0.9$ and $\delta = 0.04$), thereby providing another instance of multiplicity. In this case the most-likely industry structure is symmetric in the short run and asymmetric in the long run (see the Online Appendix for details). That is, the likelihood of cooperation goes down with time.

8.6 Organizational forgetting

We take the probability $\Delta(e_n)$ that firm n loses a unit of know-how through organizational forgetting to be $1-(1-\delta)^{e_n}$, an increasing and concave function (as long as $\delta > 0$), to

capture the idea that a firm with more know-how is more vulnerable to organizational forgetting. We alternatively take $\Delta(e_n)$ to be δ, a constant. This may be appropriate in situations in which there is a leading edge of know-how which, if not continually applied, is at risk of being lost.

Our results carry over to this *constant forgetting* specification. If organizational forgetting is sufficiently weak, then asymmetries may arise but they cannot persist. If organizational forgetting is sufficiently strong, then asymmetries cannot arise in the first place because organizational forgetting stifles investment in learning-by-doing altogether. By contrast, for intermediate degrees of organizational forgetting, asymmetries arise and persist.

Multiple equilibria continue to arise in the constant forgetting specification. Multiplicity is especially pervasive for forgetting rates δ around or somewhat below 0.5. This reaffirms our notion that the primitives of the model tie down the equilibrium unless the inflow of know-how into the industry balances the outflow. The latter happens for forgetting rates around 0.5, and the nature of the equilibrium is therefore governed by firms' expectations regarding to value of continued play.

8.7 Entry and exit

So far we have assumed that the industry is composed of a fixed number of firms. It is straightforward to extend the model to allow for entry and exit. The Online Appendix formally derives the general model; here, we briefly sketch it.

We assume that at any point in time there is a total of N firms, each of which can be either an incumbent firm or a potential entrant. Thus, if N^* is the number of incumbent firms, $N - N^*$ is the number of potential entrants. Once an incumbent firm exits the industry, it perishes and a potential entrant automatically takes its "slot" and has to decide whether or not to enter the industry. Potential entrants are drawn from a large pool. Hence, if a potential entrant chooses not to enter the industry in the current period, it disappears and its slot is given to another potential entrant in the subsequent period. In what follows we focus on the case of $N = 2$. Though alternatives are possible, we specify that an entrant comes into the industry at the top of the learning curve. Since the analysis of entry and exit requires a well-posed monopoly problem, we include an outside good.

To ensure the existence of an equilibrium, we use the approach in Doraszelski & Satterthwaite (2007). In each period, each potential entrant receives a privately observed draw S_n from a uniform distribution of possible set-up costs with support $[3, 6]$ and each incumbent firm receives a privately observed draw X_n from a uniform distribution of possible salvage values with support $[0, 3]$.[32] It is convenient to summarize the entry and exit deci-

[32]This implies that some portion of set-up costs is sunk, thereby eliminating the possibility that a firm enters the industry merely because it hopes to draw a salvage value that exceeds its set-up cost.

sions of firm n using an operating probability $\lambda_n(\mathbf{e})$, where $\mathbf{e} \in \{0, 1, \ldots, M\}^N$ is the state of the industry. If $e_n = 0$, firm n is a potential entrant and $\lambda_n(\mathbf{e})$ is the probability that it enters the industry in state $\mathbf{e}$; if $e_n \neq 0$, firm n is an incumbent firm and $\lambda_n(\mathbf{e})$ is the probability that it remains in the industry. A symmetric and anonymous Markov perfect equilibrium consists of a value function $V^*(\mathbf{e}) = V_1(\mathbf{e})$, a pricing function $p^*(\mathbf{e}) = p_1(\mathbf{e})$, and an operating probability $\lambda^*(\mathbf{e}) = \lambda_1(\mathbf{e})$ for firm 1.

Entry and exit do not alter the thrust of our results. Organizational forgetting remains a source of aggressive pricing behavior. Indeed, allowing for exit adds another component to the prize from winning a sale because by winning a sale, a firm may move the industry to a state in which its rival is likely to exit. But if the rival exits, then it may be replaced by an entrant that comes into the industry at the top of its learning curve or it may not be replaced at all. As a result, pricing behavior is more aggressive than in the basic model without entry nor exit. This leads to more pronounced asymmetries both in the short run and in the long run. It is even possible that the industry is monopolized.

Multiple equilibria continue to arise in the general model. Entry and exit exacerbate the multiplicity problem. Strikingly, in contrast to Proposition 2, there may be multiple equilibria even in the absence of organizational forgetting ($\delta = 0$). For a progress ratio of $\rho = 0.75$, for example, we found three equilibria. While these equilibria are flat either without or with well, the implied long-run industry structures range from symmetric (with the model state of the limiting distribution being $(30, 30)$) to monopolistic (with the modal states being $(0, 30)$ and $(30, 0)$). In the former equilibrium, once both firms have entered the industry, there may not be exit in the future (we have $\lambda^*(\mathbf{e}) = 1.00$ for all $\mathbf{e} \in \{1, \ldots, M\}^2$). Knowing this, firms may as well price softly, so that, in turn, the incentive to enter the industry is strong even if an incumbent must be faced (we have $\lambda^*(0, 1) = 0.84$). In the latter equilibrium, each firm uses price cuts to induce its rival to exit (we have $p^*(1, 1) = -36.95$ and $\lambda^*(2, 1) = 1.00$ but $\lambda^*(1, 2) = 0.76$). Given that post-entry pricing behavior is "predatory", the incentive to enter the industry is weak in the first place (we have $\lambda^*(0, 1) = 0.08$), thereby ensuring that the most-likely industry structure is monopolistic not only in the long run but also in the short run (the modal states of the transient distribution are $(0, 8)$ and $(8, 0)$ in period 8 and $(0, 30)$ and $(30, 0)$ in period 32).

9 Conclusions

Learning-by-doing and organizational forgetting have been shown to be important in a variety of industrial settings. Using the Markov-perfect equilibrium framework of Ericson & Pakes (1995) this paper provides a general model of dynamic competition that accounts for these economic fundamentals and shows how they shape industry structure and dynamics.

We enhance the methodological foundations of this literature in two ways. First, we show that there are equilibria that cannot be computed by the Pakes & McGuire (1994) algorithm. Second, we propose a homotopy algorithm that allows us to describe in detail the structure of the set of equilibria of our dynamic stochastic game.

In contrast to the present paper, the theoretical literature on learning-by-doing has largely ignored organizational forgetting. Moreover, it has mainly focused on firms' pricing behavior. By directly examining industry dynamics, we are able to show that ID and IID may not be sufficient for economically meaningful market dominance. By generalizing the existing models of learning-by-doing through the addition of organizational forgetting, we are able to show that these dominance properties of firms' pricing behavior break down in the presence of even a small degree of organizational forgetting. Yet, it is precisely in the presence of organizational forgetting that market dominance ensues both in the short run and in the long run.

Our analysis of the role of organizational forgetting reveals that learning-by-doing and organizational forgetting are distinct economic forces. Organizational forgetting, in particular, does not simply negate learning-by-doing. The unique role played by organizational forgetting comes about because it makes bidirectional movements through the state space possible. Hence, dynamic competition with learning-by-doing and organizational forgetting is akin to racing down an upward moving escalator. As a consequence, a model with both learning-by-doing and organizational forgetting can give rise to aggressive pricing behavior, market dominance, and multiple equilibria, whereas a model with learning-by-doing alone cannot.

Diagonal and sideways trenches are part and parcel to the self-reinforcing mechanisms that lead to market dominance. Since the leadership position is aggressively defended, firms fight a price war to attain it. This provides all the more reason to aggressively defend the leadership position because if it is lost, then another price war ensues. This seems like a good story to tell. Our computations show that this is not just an intuitively sensible story but also a logically consistent one that—perhaps—plays out in real markets.

Appendix

Proof of Proposition 1. *Part (i):* The basic differential equations (12) set

$$\delta'(s) = \det\left(\frac{\partial \mathbf{F}(\mathbf{x}(s), \delta(s))}{\partial \mathbf{x}}\right).$$

The Jacobian $\frac{\partial \mathbf{F}(\mathbf{x}(s), \delta(s))}{\partial \mathbf{x}}$ is a $(2M^2 \times 2M^2)$ matrix and therefore has an even number of eigenvalues. Its determinant is the product of its eigenvalues. Hence, if $\delta'(s) \leq 0$, then there exists at least one real nonnegative eigenvalue. (Suppose to the contrary that all eigenvalues

are either complex or real and negative. Since the number of complex eigenvalues is even, so is the number of real eigenvalues. Moreover, the product of a conjugate pair of complex eigenvalues is positive, as is the product of an even number of real negative eigenvalues.)

To relate the Pakes & McGuire (1994) algorithm to our homotopy algorithm, consider a parametric path $(\mathbf{x}(s), \delta(s)) \in \mathbf{F}^{-1}$ in the equilibrium correspondence. We show in the Online Appendix that

$$\left.\frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}}\right|_{\delta=\delta(s)} = \frac{\partial \mathbf{F}(\mathbf{x}(s), \delta(s))}{\partial \mathbf{x}} + \mathbf{I}, \tag{18}$$

where $\mathbf{I}$ denotes the $(2M^2 \times 2M^2)$ identity matrix.

The proof is completed by recalling a basic result from linear algebra: Let $\mathbf{A}$ be an arbitrary matrix and $\sigma(\mathbf{A})$ its spectrum. Then $\sigma(\mathbf{A}+\mathbf{I}) = \sigma(\mathbf{A}) + 1$ (see Proposition A.17 in Appendix A of Bertsekas & Tsitsiklis 1997). Hence, because $\frac{\partial \mathbf{F}(\mathbf{x}(s),\delta(s))}{\partial \mathbf{x}}$ has at least one real nonnegative eigenvalue, it follows from equation (18) that $\left.\frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}}\right|_{\delta=\delta(s)}$ has at least one real eigenvalue equal to or bigger than unity. Hence, $\rho\left(\left.\frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}}\right|_{\delta=\delta(s)}\right) \geq 1$.

Part (ii): Consider the iteration $\mathbf{x}^{k+1} = \tilde{\mathbf{G}}(\mathbf{x}^k) = \omega \mathbf{G}(\mathbf{x}^k) + (1-\omega)\mathbf{x}^k$, where $\omega > 0$. Using equation (18) its Jacobian at $(\mathbf{x}(s), \delta(s)) \in \mathbf{F}^{-1}$ is

$$\left.\frac{\partial \tilde{\mathbf{G}}(\mathbf{x}(s))}{\partial \mathbf{x}}\right|_{\delta=\delta(s)} = \omega \left.\frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}}\right|_{\delta=\delta(s)} + (1-\omega)\mathbf{I} = \omega \frac{\partial \mathbf{F}(\mathbf{x}(s), \delta(s))}{\partial \mathbf{x}} + \mathbf{I}.$$

As before it follows that $\rho\left(\left.\frac{\partial \tilde{\mathbf{G}}(\mathbf{x}(s))}{\partial \mathbf{x}}\right|_{\delta=\delta(s)}\right) \geq 1$. ■

Proof of Proposition 2. We rewrite the Bellman equations and FOCs in state $\mathbf{e}$ as

$$V_1 = D_1(p_1, p_2)\left(p_1 - c(e_1) + \beta\left(\overline{V}_{11} - \overline{V}_{12}\right)\right) + \beta \overline{V}_{12}, \tag{19}$$

$$V_2 = D_2(p_1, p_2)\left(p_2 - c(e_2) + \beta\left(\overline{V}_{22} - \overline{V}_{21}\right)\right) + \beta \overline{V}_{21}, \tag{20}$$

$$0 = \frac{\sigma}{D_2(p_1, p_2)} - \left(p_1 - c(e_1) + \beta\left(\overline{V}_{11} - \overline{V}_{12}\right)\right), \tag{21}$$

$$0 = \frac{\sigma}{D_1(p_1, p_2)} - \left(p_2 - c(e_2) + \beta\left(\overline{V}_{22} - \overline{V}_{21}\right)\right), \tag{22}$$

where, to simplify the notation, V_n is shorthand for $V_n(\mathbf{e})$, $\overline{V}_{nk}$ for $\overline{V}_{nk}(\mathbf{e})$, p_n for $p_n(\mathbf{e})$, etc. and we use the fact that $D_1(p_1, p_2) + D_2(p_1, p_2) = 1$.

Case (i): First suppose $\delta = 0$. The proof proceeds in a number of steps. In step 1, we establish that the equilibrium in state (M, M) is unique. In step 2a, we assume that there is a unique equilibrium in state $(e_1 + 1, M)$, where $e_1 \in \{1, \ldots, M-1\}$, and show that this implies that the equilibrium in state (e_1, M) is unique. In step 2b, we assume that there is a unique equilibrium in state $(M, e_2 + 1)$, where $e_2 \in \{1, \ldots, M-1\}$, and show that this implies that the equilibrium in state (M, e_2) is unique. By induction, steps 1, 2a, and 2b establish uniqueness along the upper edge of the state space. In step 3, we assume that there is a unique equilibrium in states $(e_1 + 1, e_2)$ and $(e_1, e_2 + 1)$, where

$e_1 \in \{1, \ldots, M-1\}$ and $e_2 \in \{1, \ldots, M-1\}$, and show that this implies that the equilibrium in state (e_1, e_2) is unique. Hence, uniqueness in state $(M-1, M-1)$ follows from uniqueness in states $(M, M-1)$ and $(M-1, M)$, uniqueness in state $(M-2, M-1)$ from uniqueness in states $(M-1, M-1)$ and $(M-2, M)$, etc. Working backwards gives uniqueness in states $(e_1, M-1)$, where $e_1 \in \{1, \ldots, M-1\}$. This, in turn, gives uniqueness in states $(e_1, M-2)$, where $e_1 \in \{1, \ldots, M-1\}$, etc.

Step 1: Consider state $\mathbf{e} = (M, M)$. From the definition of the state-to-state transitions in Section 2, we have

$$\overline{V}_{11} = \overline{V}_{12} = V_1, \quad \overline{V}_{21} = \overline{V}_{22} = V_2.$$

Imposing these restrictions and solving equations (19) and (20) for V_1 and V_2, respectively, yields

$$V_1 = \frac{D_1(p_1, p_2)(p_1 - c(e_1))}{1 - \beta}, \tag{23}$$

$$V_2 = \frac{D_2(p_1, p_2)(p_2 - c(e_2))}{1 - \beta}. \tag{24}$$

Simplifying equations (21) and (22) yields

$$0 = \frac{\sigma}{D_2(p_1, p_2)} - (p_1 - c(e_1)) = F_1(p_1, p_2), \tag{25}$$

$$0 = \frac{\sigma}{D_1(p_1, p_2)} - (p_2 - c(e_2)) = F_2(p_1, p_2). \tag{26}$$

The system of equations (25) and (26) determines equilibrium prices. Once we have established that there is a unique solution for p_1 and p_2, equations (23) and (24) immediately ascertain that V_1 and V_2 are unique.

Let $p_1^\natural(p_2)$ and $p_2^\natural(p_1)$ be defined by

$$F_1(p_1^\natural(p_2), p_2) = 0, \quad F_2(p_1, p_2^\natural(p_1)) = 0$$

and set $F(p_1) = p_1 - p_1^\natural(p_2^\natural(p_1))$. The p_1 that solves the system of equations (25) and (26) is the solution to $F(p_1) = 0$, and this solution is unique provided that $F(p_1)$ is strictly monotone. The implicit function theorem yields

$$F'(p_1) = 1 - \frac{\left(-\frac{\partial F_1}{\partial p_2}\right)}{\frac{\partial F_1}{\partial p_1}} \frac{\left(-\frac{\partial F_2}{\partial p_1}\right)}{\frac{\partial F_2}{\partial p_2}}.$$

Straightforward differentiation shows that

$$\frac{\left(-\frac{\partial F_1}{\partial p_2}\right)}{\frac{\partial F_1}{\partial p_1}} = \frac{-\frac{D_1(p_1,p_2)}{D_2(p_1,p_2)}}{-\frac{1}{D_2(p_1,p_2)}} = D_1(p_1,p_2) \in (0,1),$$
$$\frac{\left(-\frac{\partial F_2}{\partial p_1}\right)}{\frac{\partial F_2}{\partial p_2}} = \frac{-\frac{D_2(p_1,p_2)}{D_1(p_1,p_2)}}{-\frac{1}{D_1(p_1,p_2)}} = D_2(p_1,p_2) \in (0,1).$$

It follows that $F'(p_1) > 0$.

Step 2a: Consider state $\mathbf{e} = (e_1, M)$, where $e_1 \in \{1, \ldots, M-1\}$. We have

$$\overline{V}_{12} = V_1, \quad \overline{V}_{22} = V_2.$$

Imposing these restrictions and solving equations (19) and (20) for V_1 and V_2, respectively, yields

$$V_1 = \frac{D_1(p_1,p_2)(p_1 - c(e_1) + \beta\overline{V}_{11})}{1 - \beta D_2(p_1,p_2)}, \tag{27}$$
$$V_2 = \frac{D_2(p_1,p_2)(p_2 - c(e_2) - \beta\overline{V}_{21}) + \beta\overline{V}_{21}}{1 - \beta D_2(p_1,p_2)}. \tag{28}$$

Substituting equations (27) and (28) into equations (21) and (22) and dividing through by $\frac{1-\beta}{1-\beta D_2(p_1,p_2)}$ and $\frac{1}{1-\beta D_2(p_1,p_2)}$, respectively, yields

$$0 = \frac{(1 - \beta D_2(p_1,p_2))\sigma}{(1-\beta)D_2(p_1,p_2)} - \big(p_1 - c(e_1) + \beta\overline{V}_{11}\big) = G_1(p_1,p_2), \tag{29}$$
$$0 = \frac{(1 - \beta D_2(p_1,p_2))\sigma}{D_1(p_1,p_2)} - \big(p_2 - c(e_2) - \beta(1-\beta)\overline{V}_{21}\big) = G_2(p_1,p_2). \tag{30}$$

The system of equations (29) and (30) determines equilibrium prices as a function of $\overline{V}_{11}$ and $\overline{V}_{21}$. These are given by $V_1(e_1+1, M)$ and $V_2(e_1+1, M)$, respectively, and are unique by hypothesis. As in step 1, once we have established that there is a unique solution for p_1 and p_2, equations (27) and (28) immediately ascertain that, in state $\mathbf{e} = (e_1, M)$, V_1 and V_2 are unique.

Proceeding as in step 1, set $G(p_1) = p_1 - p_1^\natural(p_2^\natural(p_1))$, where $p_1^\natural(p_2)$ and $p_2^\natural(p_1)$ are defined by $G_1(p_1^\natural(p_2), p_2) = 0$ and $G_2(p_1, p_2^\natural(p_1)) = 0$, respectively. We have to show that $G(\cdot)$ is strictly monotone. Straightforward differentiation shows that

$$\frac{\left(-\frac{\partial G_1}{\partial p_2}\right)}{\frac{\partial G_1}{\partial p_1}} = \frac{-\frac{D_1(p_1,p_2)}{(1-\beta)D_2(p_1,p_2)}}{-\frac{1-\beta D_2(p_1,p_2)}{(1-\beta)D_2(p_1,p_2)}} = \frac{D_1(p_1,p_2)}{1-\beta D_2(p_1,p_2)} \in (0,1),$$
$$\frac{\left(-\frac{\partial G_2}{\partial p_1}\right)}{\frac{\partial G_2}{\partial p_2}} = \frac{-\frac{(1-\beta)D_2(p_1,p_2)}{D_1(p_1,p_2)}}{-\frac{1-\beta D_2(p_1,p_2)}{D_1(p_1,p_2)}} = \frac{(1-\beta)D_2(p_1,p_2)}{1-\beta D_2(p_1,p_2)} \in (0,1).$$

It follows that $G'(p_1) > 0$.

Step 2b: Consider state $\mathbf{e} = (M, e_2)$, where $e_2 \in \{1, \ldots, M-1\}$. We have

$$\overline{V}_{11} = V_1, \quad \overline{V}_{21} = V_2.$$

The argument is completely symmetric to the argument in step 2a and therefore omitted.

Step 3: Consider state $\mathbf{e} = (e_1, e_2)$, where $e_1 \in \{1, \ldots, M-1\}$ and $e_2 \in \{1, \ldots, M-1\}$. The system of equations (21) and (22) determines equilibrium prices as a function of $\overline{V}_{11}$, $\overline{V}_{12}$, $\overline{V}_{21}$, and $\overline{V}_{22}$. These are given by $V_1(e_1+1, e_2)$, $V_1(e_1, e_2+1)$, $V_2(e_1+1, e_2)$, and $V_2(e_1, e_2+1)$, respectively, and are unique by hypothesis. As in step 1, once we have established that there is a unique solution for p_1 and p_2, equations (19) and (20) immediately ascertain that, in state $\mathbf{e} = (e_1, e_2)$, V_1 and V_2 are unique.

Let $H_1(p_1, p_2)$ and $H_2(p_1, p_2)$ denote the RHS of equation (21) and (22), respectively. Proceeding as in step 1, set $H(p_1) = p_1 - p_1^\natural(p_2^\natural(p_1))$, where $p_1^\natural(p_2)$ and $p_2^\natural(p_1)$ are defined by $H_1(p_1^\natural(p_2), p_2) = 0$ and $H_2(p_1, p_2^\natural(p_1)) = 0$, respectively. We have to show that $H(\cdot)$ is strictly monotone. Straightforward differentiation shows that

$$\frac{\left(-\frac{\partial H_1}{\partial p_2}\right)}{\frac{\partial H_1}{\partial p_1}} = \frac{-\frac{D_1(p_1,p_2)}{D_2(p_1,p_2)}}{-\frac{1}{D_2(p_1,p_2)}} = D_1(p_1, p_2) \in (0,1),$$
$$\frac{\left(-\frac{\partial H_2}{\partial p_1}\right)}{\frac{\partial H_2}{\partial p_2}} = \frac{-\frac{D_2(p_1,p_2)}{D_1(p_1,p_2)}}{-\frac{1}{D_1(p_1,p_2)}} = D_2(p_1, p_2) \in (0,1).$$

It follows that $H'(p_1) > 0$.

Case (ii): Next suppose $\delta = 1$. A similar induction argument as in the case of $\delta = 0$ can be used to establish the claim except that in the case of $\delta = 1$ we anchor the argument in state $(1,1)$ rather than state (M, M). ■

Proof of Proposition 3. *Part (i):* Consider the static Nash equilibrium. The FOCs in state $\mathbf{e}$ are

$$p_1^\dagger(\mathbf{e}) = c(e_1) + \frac{\sigma}{1 - D_1(p_1^\dagger(\mathbf{e}), p_2^\dagger(\mathbf{e}))}, \tag{31}$$
$$p_2^\dagger(\mathbf{e}) = c(e_2) + \frac{\sigma}{1 - D_2(p_1^\dagger(\mathbf{e}), p_2^\dagger(\mathbf{e}))}. \tag{32}$$

Equations (31) and (32) imply $p_1^\dagger(\mathbf{e}) > c(e_1)$ and $p_2^\dagger(\mathbf{e}) > c(e_2)$ and thus in particular $p^\dagger(m,m) > c(m)$. In addition, $p^\dagger(\mathbf{e}) = p^\dagger(m,m)$ because $c(e_1) = c(e_2) = c(m)$ for all $\mathbf{e} \in \{m, \ldots, M\}^2$.

Turning to our dynamic stochastic game, suppose that $\delta = 0$. The proof of part (i) proceeds in a number of steps, similar to the proof of Proposition 2. In step 1, we establish that equilibrium prices in state (M, M) coincide with the static Nash equilibrium. In step 2a, we assume that the equilibrium in state (e_1+1, M), where $e_1 \in \{m, \ldots, M-1\}$, coincides with the equilibrium in state (M, M) and show that this implies that the equilibrium in state (e_1, M) does the same. In step 2b, we assume that the equilibrium in state (M, e_2+1),

where $e_2 \in \{m, \ldots, M-1\}$, coincides with the equilibrium in state (M, M) and show that this implies that the equilibrium in state (M, e_2) does the same. In step 3, we assume that the equilibrium in states $(e_1 + 1, e_2)$ and $(e_1, e_2 + 1)$, where $e_1 \in \{m, \ldots, M-1\}$ and $e_2 \in \{m, \ldots, M-1\}$, coincides with the equilibrium in state (M, M) and show that this implies that the equilibrium in state (e_1, e_2) does the same. Also similar to the proof of Proposition 2, we continue to use V_n as shorthand for $V_n(\mathbf{e})$, $\overline{V}_{nk}$ for $\overline{V}_{nk}(\mathbf{e})$, p_n for $p_n(\mathbf{e})$, etc.

Step 1: Consider state $\mathbf{e} = (M, M)$. From the proof of Proposition 2, equilibrium prices are determined by the system of equations (25) and (26). Since equations (25) and (26) are equivalent to equations (31) and (32), equilibrium prices are $p_1 = p_1^\dagger$ and $p_2 = p_2^\dagger$. Substituting equation (25) into (23) and equation (26) into (24) yields equilibrium values

$$V_1 = \frac{\sigma D_1(p_1, p_2)}{(1-\beta) D_2(p_1, p_2)}, \tag{33}$$

$$V_2 = \frac{\sigma D_2(p_1, p_2)}{(1-\beta) D_1(p_1, p_2)}. \tag{34}$$

Step 2a: Consider state $\mathbf{e} = (e_1, M)$, where $e_1 \in \{m, \ldots, M-1\}$. Equilibrium prices are determined by the system of equations (29) and (30). Given $\overline{V}_{11} = V_1(e_1 + 1, M) = V_1(M, M)$ and $\overline{V}_{21} = V_2(e_1 + 1, M) = V_2(M, M)$, it is easy to see that, in state $\mathbf{e} = (e_1, M)$, $p_1 = p_1(M, M)$ and $p_2 = p_2(M, M)$ are a solution. Substituting equation (29) into (27) and equation (30) into (28) yields equilibrium values $V_1 = V_1(M, M)$ and $V_2 = V_2(M, M)$ as given by equations (33) and (34).

Step 2b: Consider state $\mathbf{e} = (M, e_2)$, where $e_2 \in \{m, \ldots, M-1\}$. The argument is completely symmetric to the argument in step 2a and therefore omitted.

Step 3: Consider state $\mathbf{e} = (e_1, e_2)$, where $e_1 \in \{m, \ldots, M-1\}$ and $e_2 \in \{m, \ldots, M-1\}$. Equilibrium prices are determined by the system of equations (21) and (22). Given $\overline{V}_{11} = V_1(e_1 + 1, e_2) = V_1(M, M)$, $\overline{V}_{12} = V_1(e_1, e_2 + 1) = V_1(M, M)$, $\overline{V}_{21} = V_2(e_1 + 1, e_2) = V_2(M, M)$, and $\overline{V}_{22} = V_2(e_1, e_2 + 1) = V_2(M, M)$, it is easy to see that, in state $\mathbf{e} = (e_1, e_2)$, $p_1 = p_1(M, M)$ and $p_2 = p_2(M, M)$ are a solution. Substituting equation (21) into (19) and equation (22) into (20) yields equilibrium values $V_1 = V_1(M, M)$ and $V_2 = V_2(M, M)$ as given by equations (33) and (34).

Part (ii): We show that $p_2(\mathbf{e}) > c(m)$ for all $e_1 \in \{1, \ldots, m-1\}$ and $e_2 \in \{m, \ldots, M\}$. The claim follows because $p^*(\mathbf{e}) = p_2(\mathbf{e}^{[2]})$.

The proof of part (ii) proceeds in two steps. In step 1, we establish that the equilibrium price of firm 2 in state (e_1, M), where $e_1 \in \{1, \ldots, m-1\}$, exceeds $c(m)$. In step 2, we assume that the equilibrium in state $(e_1, e_2 + 1)$, where $e_1 \in \{1, \ldots, m-1\}$ and $e_2 \in \{m, \ldots, M-1\}$, coincides with the equilibrium in state (e_1, M) and show that this implies that the equilibrium in state (e_1, e_2) does the same.

Step 1: Consider state $\mathbf{e} = (e_1, M)$, where $e_1 \in \{1, \ldots, m-1\}$. From the proof of Proposition 2, equilibrium prices are determined by the system of equations (29) and (30). Using the fact that in equilibrium $\overline{V}_{21} \geq 0$, equation (30) implies $p_2 > c(m)$. Substituting

equation (29) into (27) and equation (30) into (28) yields equilibrium values

$$V_1 = \frac{\sigma D_1(p_1, p_2)}{(1-\beta)D_2(p_1, p_2)}, \tag{35}$$

$$V_2 = \frac{\sigma D_2(p_1, p_2) + \beta D_1(p_1, p_2)\overline{V}_{21}}{D_1(p_1, p_2)}. \tag{36}$$

Step 2: Consider state $\mathbf{e} = (e_1, e_2)$, where $e_1 \in \{1, \ldots, m-1\}$ and $e_2 \in \{m, \ldots, M-1\}$. Equilibrium prices are determined by the system of equations (21) and (22). Given $\overline{V}_{12} = V_1(e_1, e_2+1) = V_1(e_1, M)$ and $\overline{V}_{22} = V_2(e_1, e_2+1) = V_2(e_1, M)$, it is easy to see that, in state $\mathbf{e} = (e_1, e_2)$, $p_1 = p_1(e_1, M)$ and $p_2 = p_2(e_1, M)$ are a solution. Substituting equation (21) into (19) and equation (22) into (20) yields equilibrium values $V_1 = V_1(e_1, M)$ and $V_2 = V_2(e_1, M)$ as given by equations (35) and (36). ■

Proof of Proposition 4. We rewrite the FOCs in state $\mathbf{e}$ as

$$0 = \frac{\sigma}{D_2(p_1, p_2)} - \left(p_1 - c(e_1) + \beta\left(\overline{V}_{11} - \overline{V}_{12}\right)\right), \tag{37}$$

$$0 = \frac{\sigma}{D_1(p_1, p_2)} - \left(p_2 - c(e_2) + \beta\left(\overline{V}_{22} - \overline{V}_{21}\right)\right), \tag{38}$$

where, to simplify the notation, $\overline{V}_{nk}$ is shorthand for $\overline{V}_{nk}(\mathbf{e})$, p_n for $p_n(\mathbf{e})$, etc. and we use the fact that $D_1(p_1, p_2) + D_2(p_1, p_2) = 1$. The system of equations (37) and (38) determines equilibrium prices. We have to establish that there is a unique solution for p_1 and p_2 irrespective of $\overline{V}_{11}$, $\overline{V}_{12}$, $\overline{V}_{21}$, and $\overline{V}_{22}$.

Let $H_1(p_1, p_2)$ and $H_2(p_1, p_2)$ denote the RHS of equation (37) and (38), respectively. Proceeding as in step 3 of the proof of Proposition 2, set $H(p_1) = p_1 - p_1^\natural(p_2^\natural(p_1))$, where $p_1^\natural(p_2)$ and $p_2^\natural(p_1)$ are defined by $H_1(p_1^\natural(p_2), p_2) = 0$ and $H_2(p_1, p_2^\natural(p_1)) = 0$, respectively. We have to show that $H(\cdot)$ is strictly monotone. Straightforward differentiation shows that

$$\frac{\left(-\frac{\partial H_1}{\partial p_2}\right)}{\frac{\partial H_1}{\partial p_1}} = \frac{-\frac{D_1(p_1,p_2)}{D_2(p_1,p_2)}}{-\frac{1}{D_2(p_1,p_2)}} = D_1(p_1, p_2) \in (0,1),$$

$$\frac{\left(-\frac{\partial H_2}{\partial p_1}\right)}{\frac{\partial H_2}{\partial p_2}} = \frac{-\frac{D_2(p_1,p_2)}{D_1(p_1,p_2)}}{-\frac{1}{D_1(p_1,p_2)}} = D_2(p_1, p_2) \in (0,1).$$

It follows that $H'(p_1) > 0$. ■

References

Ackerberg, D., Benkard, L., Berry, S. & Pakes, A. (2005), Econometric tools for analyzing market outcomes, *in* J. Heckman, ed., 'Handbook of Econometrics', Vol. 6, North-Holland, Amsterdam, p. forthcoming.

Alchian, A. (1963), 'Reliability of progress curves in airframe production', *Econometrica* **31**(4), 679–693.

Argote, L., Beckman, S. & Epple, D. (1990), 'The persistence and transfer of learning in an industrial setting', *Management Science* **36**(2), 140–154.

Athey, S. & Schmutzler, A. (2001), 'Investment and market dominance', *Rand Journal of Economics* **32**(1), 1–26.

Bailey, C. (1989), 'Forgetting and the learning curve: A laboratory study', *Management Science* **35**(3), 340–352.

Baloff, N. (1971), 'Extension of the learning curve: Some empirical results', *Operational Reserach Quarterly* **22**(4), 329–340.

Benkard, L. (2000), 'Learning and forgetting: The dynamics of aircraft production', *American Economic Review* **90**(4), 1034–1054.

Benkard, L. (2004), 'A dynamic analysis of the market for wide-bodied commercial aircraft', *Review of Economic Studies* **71**(3), 581–611.

Berry, S. & Pakes, A. (2006), 'The pure characteristics demand model', *International Economic Review* **forthcoming**.

Bertsekas, D. & Tsitsiklis, J. (1997), *Parallel and distributed computation: Numerical methods*, Athena Scientific, Belmont.

Bohn, R. (1995), 'Noise and learning in semiconductor manufacturing', *Management Science* **41**(1), 31–42.

Budd, C., Harris, C. & Vickers, J. (1993), 'A model of the evolution of duopoly: Does the asymmetry between firms tend to increase or decrease?', *Review of Economic Studies* **60**(3), 543–573.

Cabral, L. & Riordan, M. (1994), 'The learning curve, market dominance, and predatory pricing', *Econometrica* **62**(5), 1115–1140.

Cabral, L. & Riordan, M. (1997), 'The learning curve, predation, antitrust, and welfare', *Journal of Industrial Economics* **45**(2), 155–169.

Caplin, A. & Nalebuff, B. (1991), 'Aggregation and imperfect competition: On the existence of equilibrium', *Econometrica* **59**(1), 26–59.

Caves, R. & Porter, M. (1977), 'From entry barriers to mobility barriers: Conjectural decisions and contrived deterrence to new competition', *Quarterly Journal of Economics* **91**(2), 241–261.

Darr, E., Argote, L. & Epple, D. (1995), 'The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises', *Management Science* **41**(11), 1750–1762.

Dasgupta, P. & Stiglitz, J. (1988), 'Learning-by-doing, market structure and industrial and trade policies', *Oxford Economics Papers* **40**(2), 246–268.

DeJong, J. (1957), 'The effects of increasing skills on cycle time and its consequences for time standards', *Ergonomics* **1**(1), 51–60.

Demichelis, S. & Germano, F. (2002), 'On (un)knots and dynamics in games', *Games and Economic Behavior* **41**, 46–60.

Doraszelski, U. & Judd, K. (2004), Avoiding the curse of dimensionality in dynamic stochastic games, Technical working paper no. 304, NBER, Cambridge.

Doraszelski, U. & Satterthwaite, M. (2007), Computable Markov-perfect industry dynamics: Existence, purification, and multiplicity, Working paper, Harvard University, Cambridge.

Dudley, L. (1972), 'Learning and productivity changes in metal products', *American Economic Review* **62**(4), 662–669.

Dutton, J. & Thomas, A. (1984), 'Treating progress functions as a managerial opportunity', *Academy of Management Review* **9**(2), 235–247.

Ericson, R. & Pakes, A. (1995), 'Markov-perfect industry dynamics: A framework for empirical work', *Review of Economic Studies* **62**, 53–82.

Fudenberg, D. & Tirole, J. (1983), 'Learning-by-doing and market performance', *Bell Journal of Economics* **14**(2), 522–530.

Fudenberg, D. & Tirole, J. (1991), *Game theory*, MIT Press, Cambridge.

Garcia, C. & Zangwill, W. (1979), 'An approach to homotopy and degree theory', *Mathematics of Operations Reserach* **4**(4), 390–405.

Ghemawat, P. & Spence, M. (1985), 'Learning curve spillovers and market performance', *Quarterly Journal of Economics* **100**(Supplement), 839–852.

Gruber, H. (1992), 'The learning curve in the production of semiconductor memory chips', *Applied Economics* **24**, 885–895.

Hatch, N. & Mowery, D. (1998), 'Process innovation and learning by doing in semiconductor manufacturing', *Management Science* **44**(1), 1461–1477.

Herings, J. & Peeters, R. (2004), 'Stationary equilibria in stochastic games: Structure, selection, and computation', *Journal of Economic Theory* **118**, 32–60.

Hirsch, W. (1952), 'Progress functions of machine tool manufacturing', *Econometrica* **20**(1), 81–82.

Hirschmann, W. (1964), 'Profit from the learning curve', *Harvard Business Review* **42**(1), 125–139.

Irwin, D. & Klenow, P. (1994), 'Learning-by-doing spillovers in the seminconductor industry', *Journal of Political Economy* **102**(6), 1200–1227.

Jarmin, R. (1994), 'Learning by doing and competition in the early rayon industry', *Rand Journal of Economics* **25**(3), 441–454.

Judd, K. (1998), *Numerical methods in economics*, MIT Press, Cambridge.

Judd, K. & Schmedders, K. (2004), A computational approach to proving uniqueness in dynamic games, Working paper, Hoover Institution, Stanford.

Kilbridge, M. (1962), 'A model of industrial learning', *Management Science* **8**(4), 516–527.

Levy, F. (1965), 'Adaptation in the production process', *Management Science* **11**(6), 136–154.

Lewis, T. & Yildirim, H. (2002), 'Managing dynamic competition', *American Economic Review* **92**(4), 779–797.

Lewis, T. & Yildirim, H. (2005), 'Managing switching costs in multiperiod procurements with strategic buyers', *International Economic Review* **46**(4), 1233–1269.

Lieberman, M. (1984), 'The learning curve and pricing in the chemical processing industries', *Rand Journal of Economics* **15**(2), 213–228.

Pakes, A. (2006), Econometrics and theory in empirical IO, Fisher-Schultz lecture, Harvard University, Cambridge.

Pakes, A. & McGuire, P. (1994), 'Computing Markov-perfect Nash equilibria: Numerical implications of a dynamic differentiated product model', *Rand Journal of Economics* **25**(4), 555–589.

Pisano, G. (1994), 'Knowledge, integration, and the locus of learning: An empirical analysis of process development', *Strategic Management Journal* **15**, 85–100.

Preston, L. & Keachie, E. (1964), 'Cost functions and progress functions: An integration', *American Economic Review* **54**(2), 100–107.

Ross, D. (1986), 'Learning to dominate', *Journal of Industrial Economics* **34**(4), 337–353.

Schmedders, K. (1998), 'Computing equilibria in the general equilibrium model with incomplete asset markets', *Journal of Economic Dynamics and Control* **22**, 1375–1401.

Schmedders, K. (1999), 'A homotopy algorithm and an index theorem for the general equilibrium model with incomplete asset markets', *Journal of Mathematical Economics* **32**(2), 225–241.

Shafer, S., Nembhard, D. & Uzumeri, M. (2001), 'The effects of worker learning, forgetting, and heterogeneity on assembly line productivity', *Management Science* **47**(12), 1639–1653.

Spence, M. (1981), 'The learning curve and competition', *Bell Journal of Economics* **12**(1), 49–70.

Thompson, P. (2001), 'How much did the Liberty shipbuilders learn? new evidence for an old case study', *Journal of Political Economy* **109**(1), 103–137.

Thompson, P. (2003), How much did the Liberty shipbuilders forget?, Working paper, Florida International University, Miami.

Thornton, R. & Thompson, P. (2001), 'Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding', *American Economic Review* **91**(5), 1350–1368.

Vickers, J. (1986), 'The evolution of market structure when there is a sequence of innovations', *Journal of Industrial Economics* **35**(1), 1–12.

Vives, X. (1999), *Oligopoly pricing: Old ideas and new tools*, MIT Press, Cambridge.

Watson, L., Billups, S. & Morgan, A. (1987), 'HOMPACK: A suite of codes for globally convergent homotopy algorithms', *ACM Transcations on Mathematical Software* **13**(3), 281–310.

Watson, L., Sosonkina, M., Melville, R. & Morgan, A. (1997), 'HOMPACK90: A suite of Fortran 90 codes for globally convergent homotopy algorithms', *ACM Transcations on Mathematical Software* **23**(4), 514–549.

Wixted, M. & Ebbesen, E. (1991), 'On the form of forgetting', *Psychological Science* **2**(6), 409–415.

Wright, T. (1936), 'Factors affecting the cost of airplanes', *Journal of Aeronautical Sciences* **3**(4), 122–128.

Zangwill, W. & Garcia, C. (1981), *Pathways to solutions, fixed points, and equilibria*, Prentice Hall, Englewood Cliffs.

Zimmerman, M. (1982), 'Learning effects and the commercialization of new energy technologies: The case of nuclear power', *Bell Journal of Economics* **13**(2), 297–310.

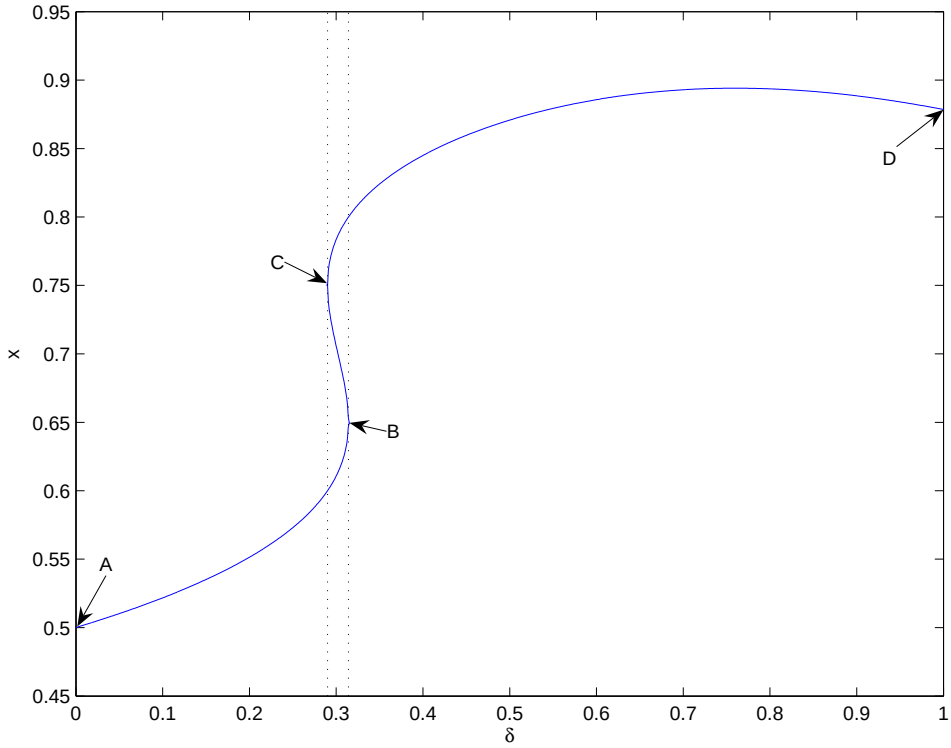

Figure 1: Homotopy example.

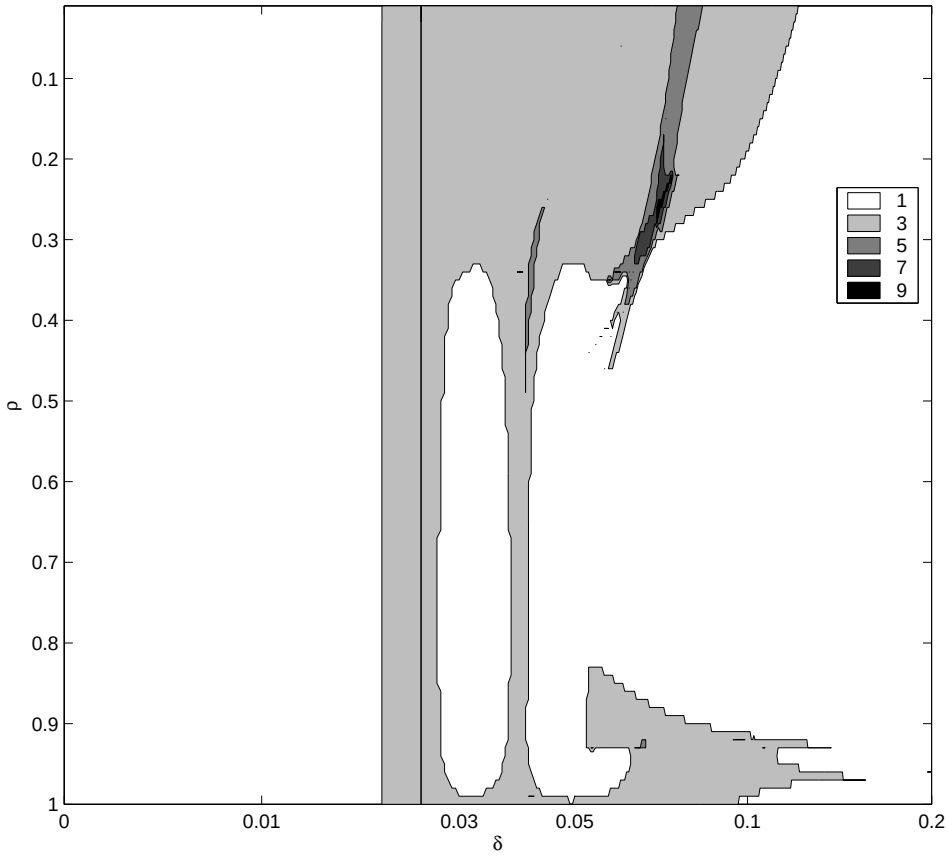

Figure 2: Number of equilibria.

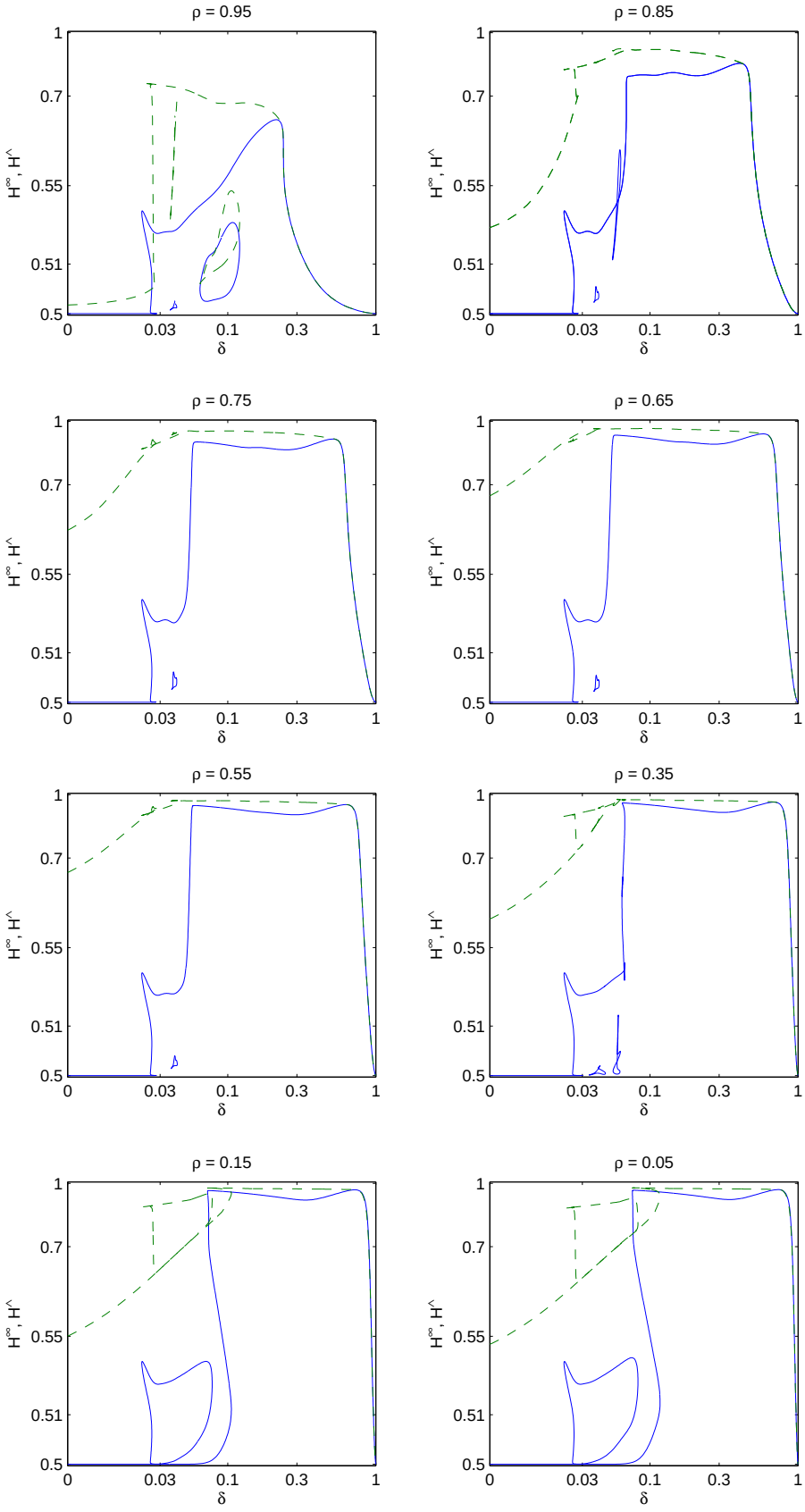

Figure 3: Limiting expected Herfindahl index H^{∞} (solid line) and maximum expected Herfindahl index $H^{\wedge}$ (dashed line).

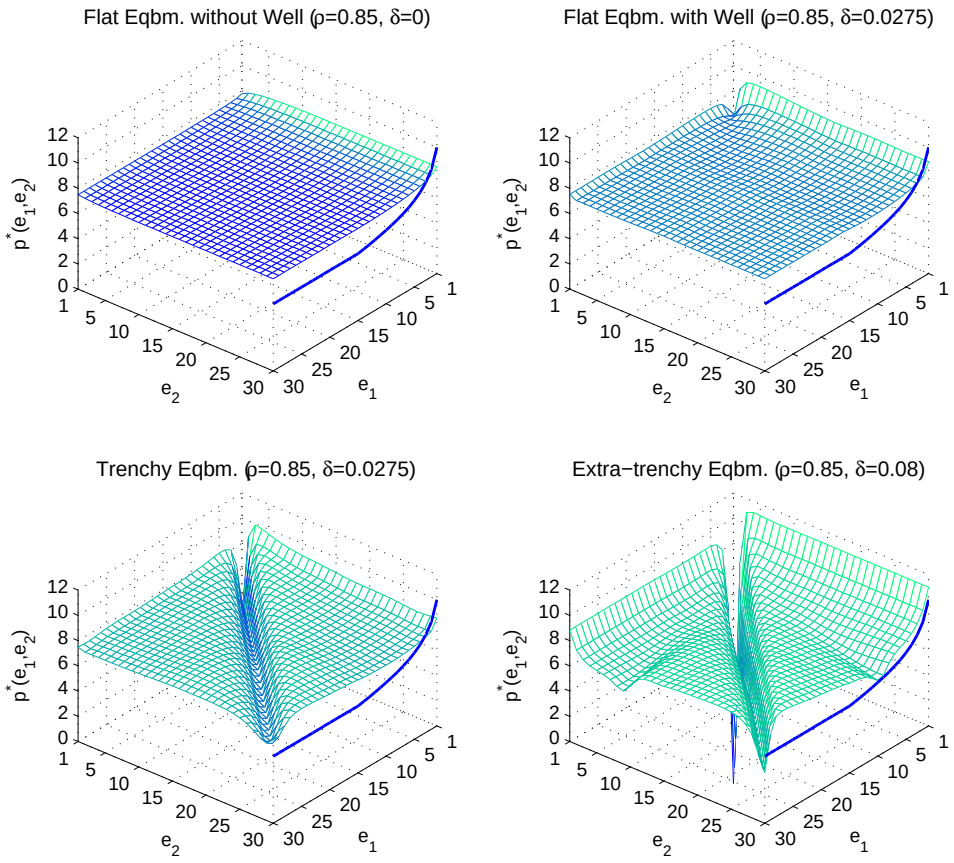

Figure 4: Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$-plane).

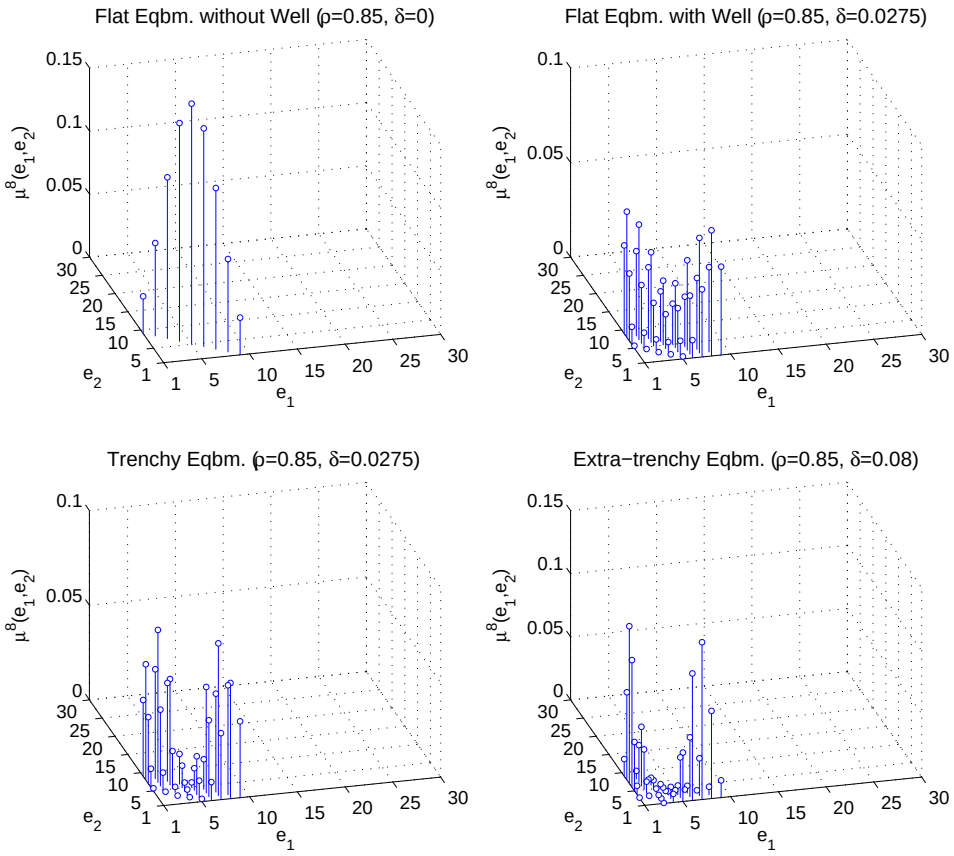

Figure 5: Transient distribution over states in period 8 given initial state $(1, 1)$.

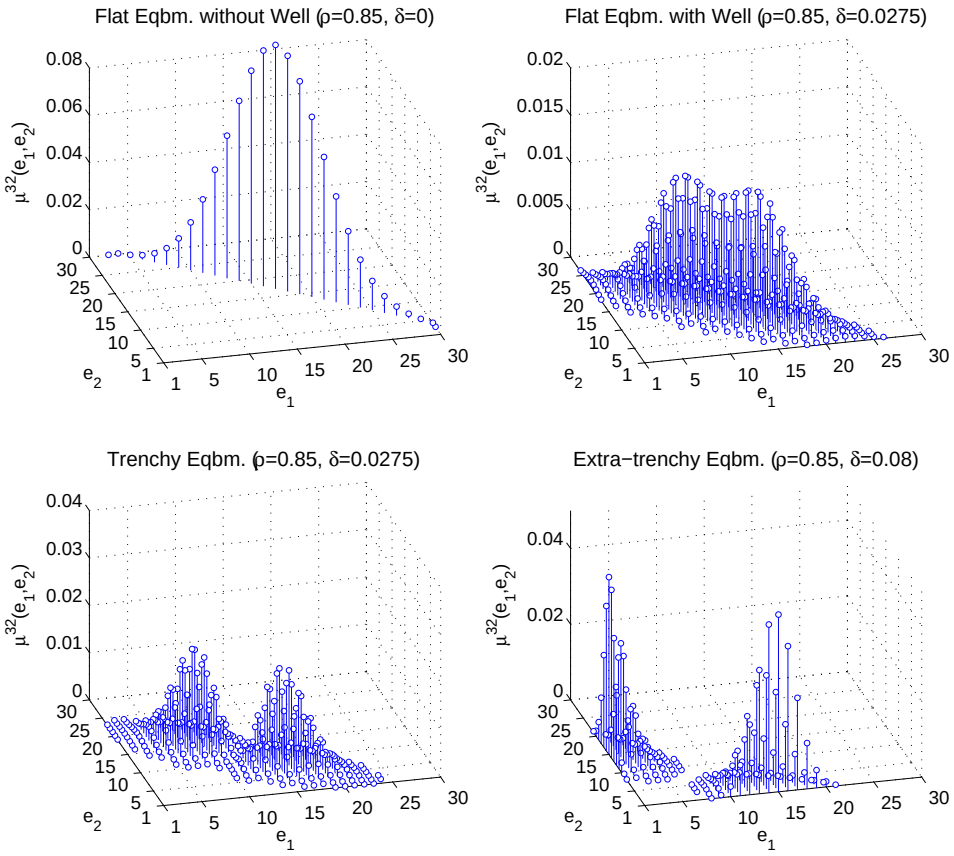

Figure 6: Transient distribution over states in period 32 given initial state $(1, 1)$.

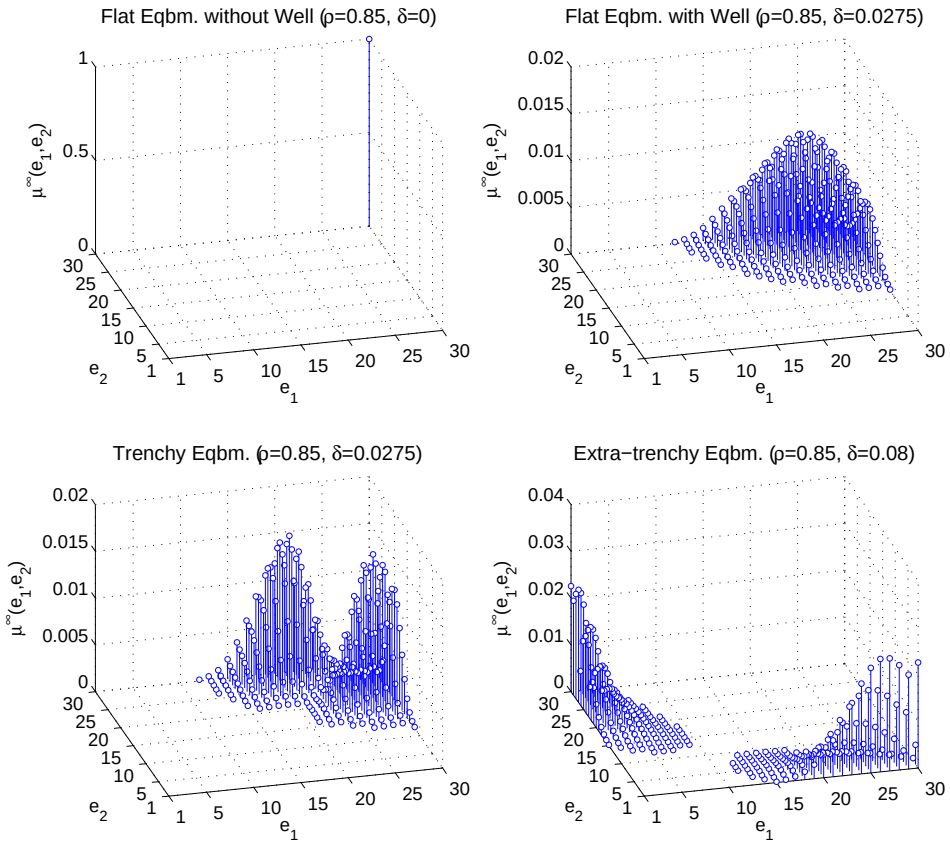

Figure 7: Limiting distribution over states.

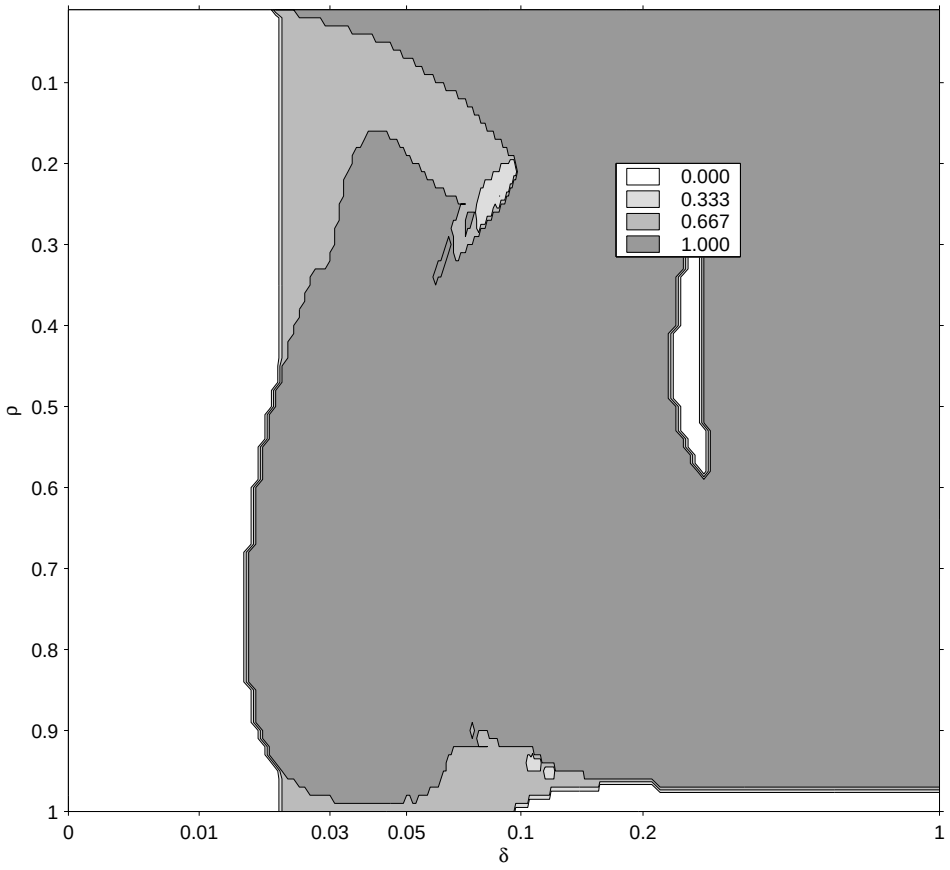

Figure 8: Share of equilibria violating upper bound on equilibrium prices (Result 4).

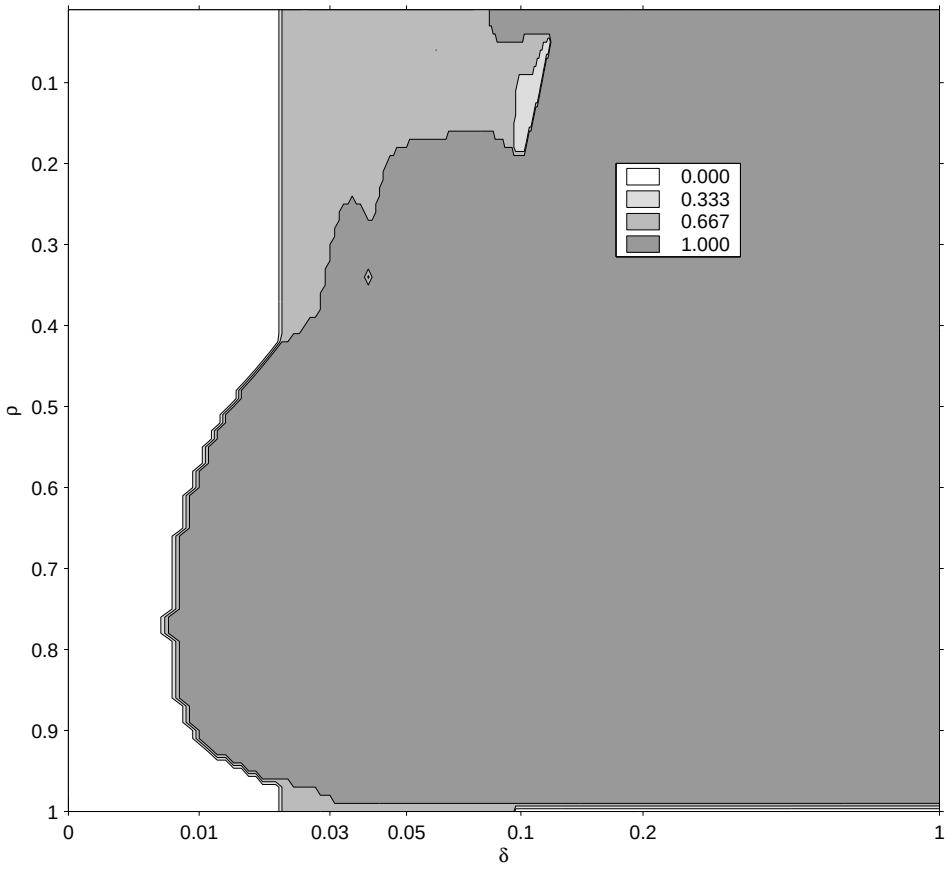

Figure 9: Share of equilibria violating lower bound on equilibrium prices (parts (i) and (ii) of Proposition 3).

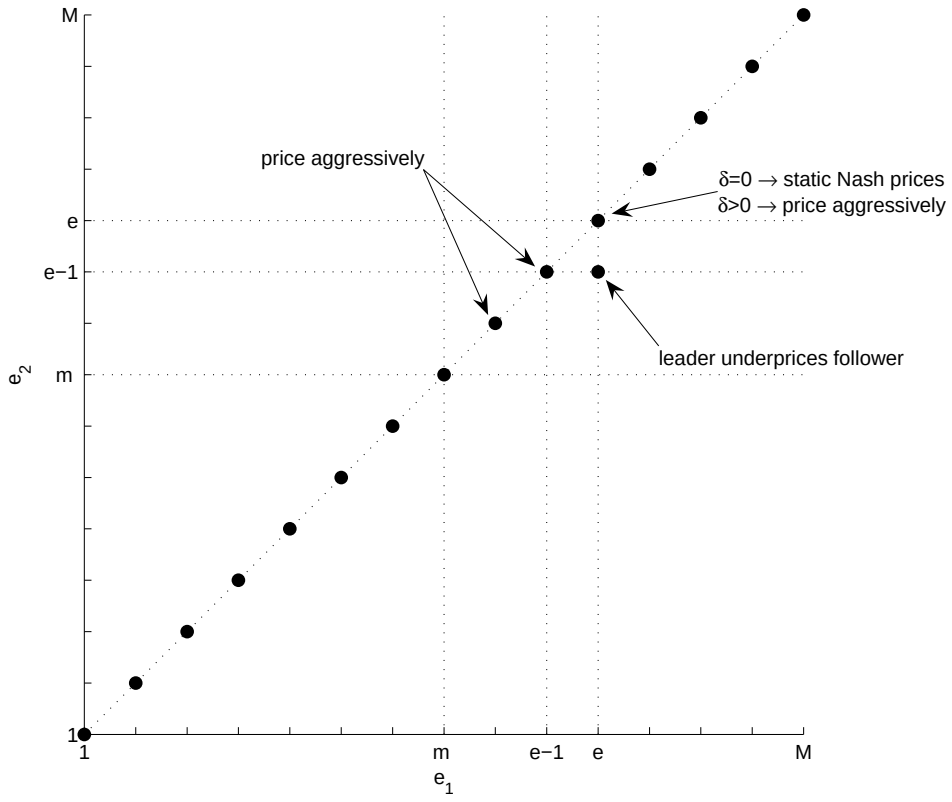

Figure 10: Diagonal trench.

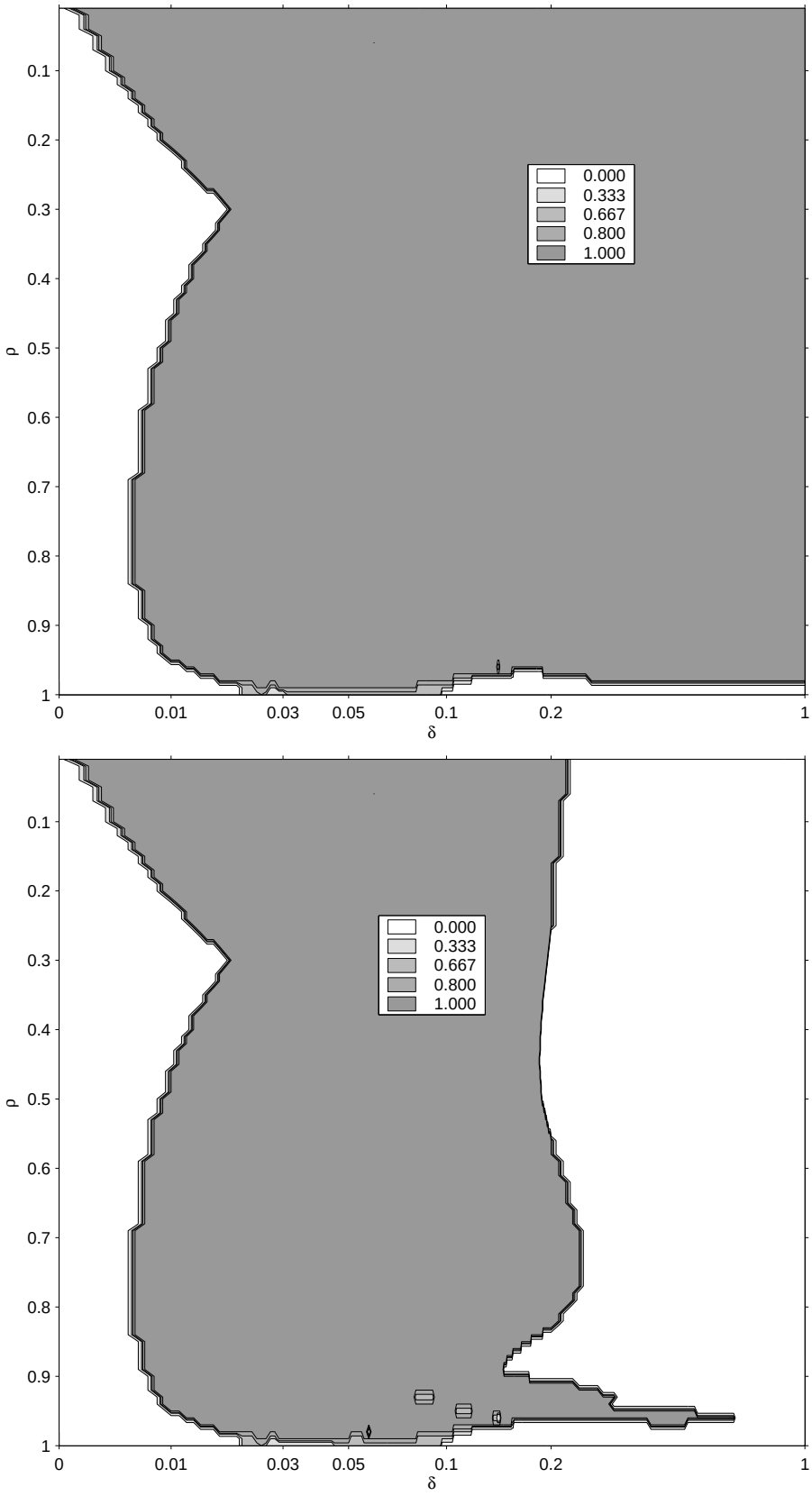

Figure 11: Share of equilibria violating IID (part (i) of Result 7, upper panel) and share of equilibria violating ID (part (ii) of Result 7, lower panel).

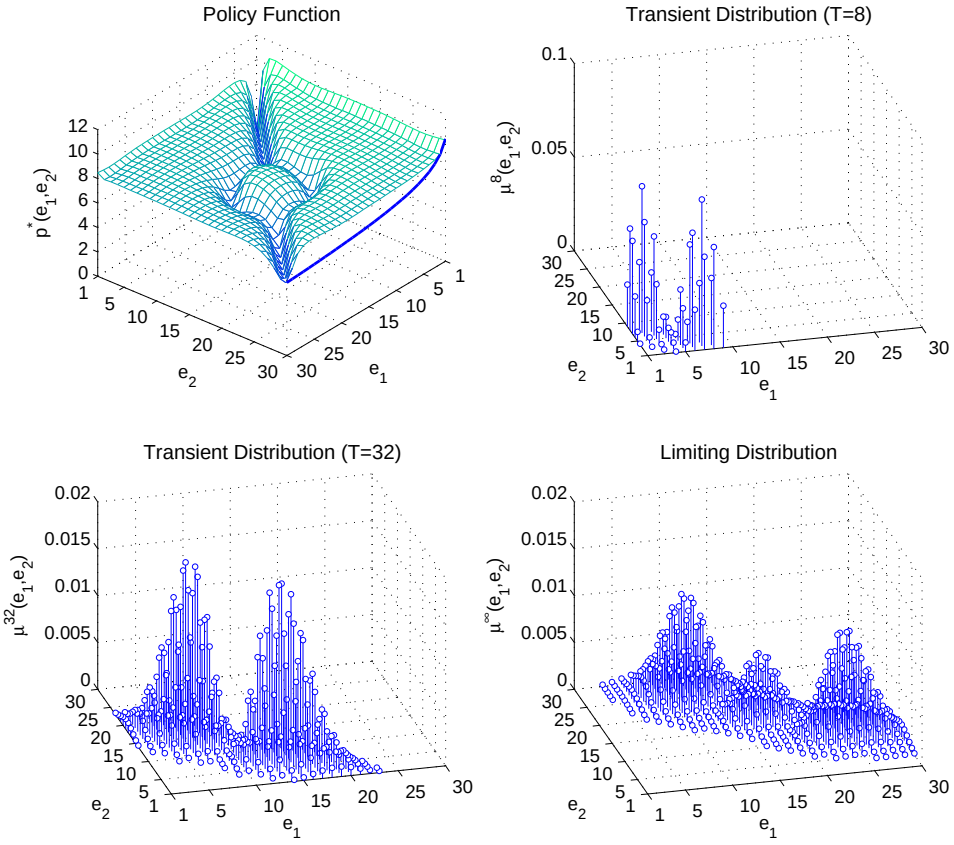

Figure 12: Bottomless learning. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$-plane) (upper left panel). Transient distribution over states in period 8 and 32 given initial state $(1, 1)$ (upper right and lower left panels). Limiting distribution over states (lower right panel). Plateau equilibrium ($\rho = 0.9$, $\delta = 0.04$).