

Learning-by-Doing, Organizational Forgetting, and Industry Dynamics – Online Appendix –

David Besanko* Ulrich Doraszelski† Yaroslav Kryukov‡

Mark Satterthwaite§

October 8, 2009

*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.

A1 Model: Organizational forgetting

Below we show that the expected stock of know-how in the absence of further learning is a decreasing convex function of time provided that $\Delta(e_n)$ is increasing in e_n .

Omitting firm subscripts to simplify the notation, let $\varphi(t) = E(e_t|e_0)$ be the expected stock of know-how in period t assuming that the initial stock of know-how is e_0 and that there is no further learning.

Proposition A1 *If $\Delta(e_t)$ is constant in e_t , then $\varphi(t)$ is a decreasing linear function of t . If $\Delta(e_t)$ is increasing in e_t , then $\varphi(t)$ is a decreasing convex function of t .*

Proof. In the absence of further learning, $q_t = 0$ and the evolution of the stock of know-how is governed by the law of motion

$$e_{t+1} = e_t - f_t.$$

Taking expectations (conditional on e_t) gives us

$$E(e_{t+1}|e_t) = e_t - E(f_t|e_t) = e_t - \Delta(e_t).$$

Since for any two random variables X and Y , $E_Y(E_X(X|Y)) = E_X(X)$, we can take expectations (conditional on e_0) on both sides of the above equation to obtain

$$E(e_{t+1}|e_0) = E(e_t|e_0) - E(\Delta(e_t)|e_0).$$

This implies

$$\begin{aligned} \varphi(1) &= e_0 - \Delta(e_0), \\ \varphi(t+1) - \varphi(t) &= -E(\Delta(e_t)|e_0), \quad t \geq 1. \end{aligned}$$

Since $-E(\Delta(e_t)|e_0) < 0$, $\varphi(t)$ is a decreasing function of t . Let $\Delta\varphi(t) = \varphi(t) - \varphi(t-1)$ be its slope, so that

$$\Delta\varphi(t+1) - \Delta\varphi(t) = E(\Delta(e_{t-1})|e_0) - E(\Delta(e_t)|e_0).$$

If $\Delta(e_t)$ is constant in e_t , then $\Delta\varphi(t+1) - \Delta\varphi(t) = 0$, and $\varphi(t)$ is a linear function of t . If, by contrast, $\Delta(e_t)$ is increasing in e_t , then $\Delta\varphi(t+1) - \Delta\varphi(t) > 0$ because the distribution of e_{t-1} stochastically dominates the distribution of e_t in the absence of further learning. Thus, $\varphi(t)$ is a convex function of t . ■

A2 Model: Parameterization

Below we show how to map the empirical estimates of rates of depreciation into in our specification.

Empirical work on organizational forgetting employs a capital-stock model. This model is defined by the deterministic law of motion

$$e'_n = (1 - \xi)e_n + y_n,$$

where ξ is the rate of depreciation and y_n is the flow of orders. If the flow of orders is equal to a constant y , the steady-state stock of know-how is

$$\frac{y}{\xi}.$$

Recall our stochastic law of motion:

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

Taking expectations yields

$$E(e'_n|e_n) = e_n + \gamma - \Delta(e_n),$$

where $\gamma = \Pr(q_n = 1)$ is the probability that the firm makes a sale and $\Delta(e_n) = \Pr(f_n = 1) = 1 - (1 - \delta)^{e_n}$ is the probability that it loses a unit of know-how through organizational forgetting. The steady-state stock of know-how is

$$\frac{\ln(1 - \gamma)}{\ln(1 - \delta)}.$$

We now ask what is the value of the forgetting rate δ so that the two specifications generate the same steady-state stock of know-how? The answer is given by

$$\frac{y}{\xi} = \frac{\ln(1 - \gamma)}{\ln(1 - \delta)}$$

or, equivalently,

$$\delta = 1 - (1 - \gamma)^{\frac{\xi}{y}}.$$

To illustrate, consider Benkard's (2000) empirical analysis of organizational forgetting in the production of wide-bodied airframes. There were 250 L-1011 aircraft produced over a 14 year period. Assuming a smooth flow of orders, this implies $y = 1.5$ units per month. Benkard (2000) estimates a rate of depreciation of $\xi = 4$ percent per month. This implies that the steady-state stock of know-how is equal to $\frac{1.5}{0.04} = 37.50$ units. Matching steady states implies that δ falls in the range between 0.0014 and 0.077 as γ ranges between 0.05 and 0.95, with $\delta = 0.018$ when $\gamma = 0.5$.

A3 Computation: Pakes & McGuire (1994) algorithm

In this section we first relate the Pakes & McGuire (1994) (P-M) algorithm to our homotopy algorithm. Then we discuss in more detail the extent and source of the difficulties the P-M algorithm suffers.

A3.1 Relationship of Jacobians

To relate the P-M algorithm to our homotopy algorithm, recall that $\mathbf{x} = (\mathbf{V}^*, \mathbf{p}^*)$.

Proposition A2 *Let $(\mathbf{x}(s), \delta(s)) \in \mathbf{F}^{-1}(\rho)$ be a parametric path of equilibria. We have*

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

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

Homotopy algorithm. Before proving the proposition, we provide some notation. Recall that the homotopy algorithm searches for a zero of $\mathbf{F}(\cdot)$, the collection of equations (6) in the main paper that defines an equilibrium. Further recall that the maximand in the Bellman equation (1) in the main paper is

$$h_n(\mathbf{e}, p_n, p_{-n}(\mathbf{e}), \mathbf{V}_n) = D_n(p_n, p_{-n}(\mathbf{e})) (p_n - c(e_n)) + \beta \sum_{k=1}^2 D_k(p_n, p_{-n}(\mathbf{e})) \bar{V}_{nk}(\mathbf{e}).$$

It is convenient to reformulate $F_{\mathbf{e}}^1(\mathbf{x}; \delta, \rho)$ and $F_{\mathbf{e}}^2(\mathbf{x}; \delta, \rho)$ in equations (4) and (5) in the main paper as

$$\begin{aligned} F_{\mathbf{e}}^1(\mathbf{x}; \delta, \rho) &= -V^*(\mathbf{e}) + h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*), \\ F_{\mathbf{e}}^2(\mathbf{x}; \delta, \rho) &= \frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_1} \bigg/ \frac{1}{\sigma} D_1(p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]})) \\ &= q_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*), \end{aligned}$$

where

$$\begin{aligned} & q_n(\mathbf{e}, p_n, p_{-n}(\mathbf{e}), \mathbf{V}_n) \\ &= \sigma - (1 - D_n(p_n, p_{-n}(\mathbf{e}))) (p_n - c(e_n)) - \beta \bar{V}_{nn}(\mathbf{e}) + \beta \sum_{k=1}^2 D_k(p_n, p_{-n}(\mathbf{e})) \bar{V}_{nk}(\mathbf{e}). \quad (\text{A1}) \end{aligned}$$

P-M algorithm. Recall that the P-M algorithm searches for a fixed point $\mathbf{x} = (\mathbf{V}^*, \mathbf{p}^*)$ of $\mathbf{G}(\cdot)$, the collection of equations (16) in the main paper that maps old guesses for the value and policy functions of firm 1 into new guesses. Again it is convenient to reformulate

$G_e^2(\mathbf{x})$ and $G_e^1(\mathbf{x})$ in equations (14) and (15) in the main paper as

$$\begin{aligned} G_e^2(\mathbf{x}) &= \arg \max_{p_1} h_1(\mathbf{e}, p_1, p^*(\mathbf{e}^{[2]}), \mathbf{V}^*) = \left\{ p_1 \left| \frac{\partial h_1(\mathbf{e}, p_1, p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_1} = 0 \right. \right\} \\ &= \left\{ p_1 \left| q_1(\mathbf{e}, p_1, p^*(\mathbf{e}^{[2]}), \mathbf{V}^*) = 0 \right. \right\}, \\ G_e^1(\mathbf{x}) &= \max_{p_1} h_1(\mathbf{e}, p_1, p^*(\mathbf{e}^{[2]}), \mathbf{V}^*) = h_1(\mathbf{e}, G_e^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*), \end{aligned}$$

where, recall, $G_e^2(\mathbf{x})$ is uniquely determined because $h_1(\cdot)$ is strictly quasi-concave in p_1 .

Proof of Proposition A2. Let $(\mathbf{x}(s), \delta(s)) \in \mathbf{F}^{-1}(\rho)$ be a parametric path of equilibria. In what follows we suppress the dependence of $(\mathbf{x}(s), \delta(s))$ on s to simplify the notation. Then

$$F_e^1(\mathbf{x}; \delta, \rho) = 0, \tag{A2}$$

$$F_e^2(\mathbf{x}; \delta, \rho) = 0 \tag{A3}$$

because the equilibrium is a zero of $\mathbf{F}(\cdot)$ and also

$$V^*(\mathbf{e}) = G_e^1(\mathbf{x}), \tag{A4}$$

$$p^*(\mathbf{e}) = G_e^2(\mathbf{x}) \tag{A5}$$

because the equilibrium is a fixed point of $\mathbf{G}(\cdot)$. From hereon we assume that $\mathbf{G}(\cdot)$ and its derivatives are evaluated at δ and ρ .

Letting x_i denote the i th element of \mathbf{x} , we have to show that

$$\begin{aligned} \frac{\partial G_e^1(\mathbf{x})}{\partial x_i} &= \frac{\partial F_e^1(\mathbf{x}; \delta, \rho)}{\partial x_i} + 1(x_i = V^*(\mathbf{e})), \quad i = 1, \dots, 2M^2, \\ \frac{\partial G_e^2(\mathbf{x})}{\partial x_i} &= \frac{\partial F_e^2(\mathbf{x}; \delta, \rho)}{\partial x_i} + 1(x_i = p^*(\mathbf{e})), \quad i = 1, \dots, 2M^2 \end{aligned}$$

for all states $\mathbf{e} \in \{1, \dots, M\}^2$.

Case (i): Consider first $F_e^1(\mathbf{x}; \delta, \rho)$ and $G_e^1(\mathbf{x})$ for an arbitrary state $\mathbf{e} \in \{1, \dots, M\}^2$. In what follows we repeatedly use the fact that equation (A3) implies

$$\frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_1} = 0. \tag{A6}$$

Firm's price: If $e_1 \neq e_2$, then we have

$$\frac{\partial F_e^1(\mathbf{x}; \delta, \rho)}{\partial p^*(\mathbf{e})} = \frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_1} = 0$$

because of equation (A6) and

$$\frac{\partial G_{\mathbf{e}}^1(\mathbf{x})}{\partial p^*(\mathbf{e})} = \frac{\partial h_1(\mathbf{e}, G_{\mathbf{e}}^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_1} \frac{\partial G_{\mathbf{e}}^2(\mathbf{x})}{\partial p^*(\mathbf{e})} = 0$$

because of equations (A5) and (A6).

Both prices: If $e_1 = e_2$, then $p^*(\mathbf{e}) = p^*(\mathbf{e}^{[2]})$ and we have

$$\begin{aligned} \frac{\partial F_{\mathbf{e}}^1(\mathbf{x}; \delta, \rho)}{\partial p^*(\mathbf{e})} &= \frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_1} + \frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_2} \\ &= \frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_2} \end{aligned}$$

because of equation (A6) and

$$\begin{aligned} \frac{\partial G_{\mathbf{e}}^1(\mathbf{x})}{\partial p^*(\mathbf{e})} &= \frac{\partial h_1(\mathbf{e}, G_{\mathbf{e}}^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_1} \frac{\partial G_{\mathbf{e}}^2(\mathbf{x})}{\partial p^*(\mathbf{e})} + \frac{\partial h_1(\mathbf{e}, G_{\mathbf{e}}^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_2} \\ &= \frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_2} \end{aligned}$$

because of equations (A5) and (A6).

Other: If $x_i \neq p^*(\mathbf{e})$, then we have

$$\frac{\partial F_{\mathbf{e}}^1(\mathbf{x}; \delta, \rho)}{\partial x_i} = -1(x_i = V^*(\mathbf{e})) + \frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial x_i},$$

where $-1(x_i = V^*(\mathbf{e}))$ is the derivative of $-V^*(\mathbf{e})$ with respect to x_i (with $1(\cdot)$ being the indicator function), and

$$\begin{aligned} \frac{\partial G_{\mathbf{e}}^1(\mathbf{x})}{\partial x_i} &= \frac{\partial h_1(\mathbf{e}, G_{\mathbf{e}}^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_1} \frac{\partial G_{\mathbf{e}}^2(\mathbf{x})}{\partial x_i} + \frac{\partial h_1(\mathbf{e}, G_{\mathbf{e}}^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial x_i} \\ &= \frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial x_i} \end{aligned}$$

because of equations (A5) and (A6).

Case (ii): Next consider $F_{\mathbf{e}}^2(\mathbf{x}; \delta, \rho)$ and $G_{\mathbf{e}}^2(\mathbf{x})$ for an arbitrary state $\mathbf{e} \in \{1, \dots, M\}^2$. In what follows we repeatedly use the fact that equation (A3) implies

$$\frac{\partial q_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), V^*)}{\partial p_1} = -1 + \frac{\partial h_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), V^*)}{\partial p_1} = -1. \quad (\text{A7})$$

Firm's price: If $e_1 \neq e_2$, then we have

$$\frac{\partial F_{\mathbf{e}}^2(\mathbf{x}; \delta, \rho)}{\partial p^*(\mathbf{e})} = \frac{\partial q_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), V^*)}{\partial p_1} = -1$$

because of equation (A7) and

$$\frac{\partial G_{\mathbf{e}}^2(\mathbf{x})}{\partial p^*(\mathbf{e})} = 0$$

because the construction of $G_{\mathbf{e}}^2(\mathbf{x})$ does not depend on $p^*(\mathbf{e})$.

Both prices: If $e_1 = e_2$, then $p^*(\mathbf{e}) = p^*(\mathbf{e}^{[2]})$ and we have

$$\begin{aligned} \frac{\partial F_{\mathbf{e}}^2(\mathbf{x}; \delta, \rho)}{\partial p^*(\mathbf{e})} &= \frac{\partial q_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_1} + \frac{\partial q_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_2} \\ &= -1 + \frac{\partial q_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_2} \end{aligned}$$

because of equation (A7) and

$$\begin{aligned} \frac{\partial G_{\mathbf{e}}^2(\mathbf{x})}{\partial p^*(\mathbf{e})} &= -\frac{\partial q_1(\mathbf{e}, G_{\mathbf{e}}^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)/\partial p_2}{\partial q_1(\mathbf{e}, G_{\mathbf{e}}^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)/\partial p_1} \\ &= \frac{\partial q_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial p_2} \end{aligned}$$

because the construction of $G_{\mathbf{e}}^2(\mathbf{x})$ does not depend on $p^*(\mathbf{e})$, the implicit function theorem, and equations (A5) and (A7).

Other: If $x_i \neq p^*(\mathbf{e})$, then we have

$$\frac{\partial F_{\mathbf{e}}^2(\mathbf{x}; \delta, \rho)}{\partial x_i} = \frac{\partial q_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial x_i}$$

and

$$\begin{aligned} \frac{\partial G_{\mathbf{e}}^2(\mathbf{x})}{\partial x_i} &= -\frac{\partial q_1(\mathbf{e}, G_{\mathbf{e}}^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)/\partial x_i}{\partial q_1(\mathbf{e}, G_{\mathbf{e}}^2(\mathbf{x}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)/\partial p_1} \\ &= \frac{\partial q_1(\mathbf{e}, p^*(\mathbf{e}), p^*(\mathbf{e}^{[2]}), \mathbf{V}^*)}{\partial x_i} \end{aligned}$$

because of the implicit function theorem and equations (A5) and (A7). ■

A3.2 Extent of difficulties

Next we illustrate the equilibria of our model that the P-M algorithm cannot compute. Figure A1 summarizes Propositions 1 and 2 in the main paper by marking equilibria with $\varrho \left(\frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}} \Big|_{(\delta(s), \rho)} \right) \geq 1$ using a dotted line and equilibria with $\varrho \left(\frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}} \Big|_{(\delta(s), \rho)} \right) < 1$ using a solid line. The former are unstable under the P-M algorithm. Note that the P-M algorithm cannot compute any equilibrium on the backward bending part of the path and that it cannot compute some equilibria on the forward bending part.

A3.3 Source of difficulties

Finally we explore the source of the difficulties of the P-M algorithm. 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. In addition, we show that the value function iteration also is contractive holding fixed the policy function.

Best reply dynamics. Defining

$$\mathbf{G}^2(\mathbf{p}; \mathbf{V}) = \begin{pmatrix} G_{(1,1)}^2(\mathbf{V}, \mathbf{p}) \\ G_{(2,1)}^2(\mathbf{V}, \mathbf{p}) \\ \vdots \\ G_{(M,M)}^2(\mathbf{V}, \mathbf{p}) \end{pmatrix}$$

we write the P-M algorithm with the value function held fixed as

$$\mathbf{p}^{k+1} = \mathbf{G}^2(\mathbf{p}^k; \mathbf{V}), \quad k = 0, 1, 2, \dots$$

The following proposition establishes that \mathbf{G}^2 is a contraction. This implies that the best reply dynamics converge to a unique fixed point irrespective of the initial guess.

Proposition A3 *Holding fixed $\mathbf{V} \in [\check{V}, \hat{V}]^{M^2}$, where $-\infty < \check{V} \leq \hat{V} < \infty$, \mathbf{G}^2 is a contraction.*

Proof. Recall that $G_{\mathbf{e}}^2(\mathbf{V}, \mathbf{p})$ is the solution to the equation

$$q_1(\mathbf{e}, p_1, p(\mathbf{e}^{[2]}), \mathbf{V}) = 0, \tag{A8}$$

where $q_1(\cdot)$ is defined in equation (A1). To avoid having to deal with corner solutions, pick $-\infty < \check{p} \leq \hat{p} < \infty$ large enough so that \mathbf{G}^2 maps $[\check{p}, \hat{p}]^{M^2}$ into itself. Note that $[\check{p}, \hat{p}]^{M^2}$ is convex and that \mathbf{G}^2 is continuously differentiable. Moreover, since $G_{\mathbf{e}}^2(\mathbf{V}, \mathbf{p})$ is the solution to equation (A8), it is straightforward to show using the implicit function theorem that the entries of the Jacobian $\frac{\partial \mathbf{G}^2(\mathbf{p}; \mathbf{V})}{\partial \mathbf{p}}$ are generated by

$$\begin{aligned} & \frac{\partial G_{\mathbf{e}}^2(\mathbf{V}, \mathbf{p})}{\partial p(\mathbf{e}^{[2]})} = \frac{D_2(G_{\mathbf{e}}^2(\mathbf{V}, \mathbf{p}), p(\mathbf{e}^{[2]}))}{\sigma} \\ & \times \left(D_1(G_{\mathbf{e}}^2(\mathbf{V}, \mathbf{p}), p(\mathbf{e}^{[2]})) (G_{\mathbf{e}}^2(\mathbf{V}, \mathbf{p}) - c(e_1)) - \beta \bar{V}_2(\mathbf{e}) + \beta \sum_{k=1}^2 D_k(G_{\mathbf{e}}^2(\mathbf{V}, \mathbf{p}), p(\mathbf{e}^{[2]})) \bar{V}_k(\mathbf{e}) \right). \end{aligned} \tag{A9}$$

It is helpful to re-write equation (A9): Since $G_e^2(\mathbf{V}, \mathbf{p})$ is the solution to equation (A8), we have

$$G_e^2(\mathbf{V}, \mathbf{p}) - c(e_1) = \frac{1}{1 - D_1(G_e^2(\mathbf{V}, \mathbf{p}), p(\mathbf{e}^{[2]}))} \left(\sigma - \beta \bar{V}_1(\mathbf{e}) + \beta \sum_{k=1}^2 D_k(G_e^2(\mathbf{V}, \mathbf{p}), p(\mathbf{e}^{[2]})) \bar{V}_k(\mathbf{e}) \right).$$

Substituting into equation (A9) and simplifying yields

$$\frac{\partial G_e^2(\mathbf{V}, \mathbf{p})}{\partial p(\mathbf{e}^{[2]})} = D_1(G_e^2(\mathbf{V}, \mathbf{p}), p(\mathbf{e}^{[2]})) \in [D_1(\hat{p}, \check{p}), D_1(\check{p}, \hat{p})] \subseteq (0, 1).$$

The rest of the proof is a minor modification of the proof of Proposition 1.10 in Section 3.1 of Bertsekas & Tsitsiklis (1997). (In their notation set $m = 1$ and $g(t) = \mathbf{G}^2(t\mathbf{p}^\dagger + (1-t)\mathbf{p}; \mathbf{V})$ to show that $\|\mathbf{G}^2(\mathbf{p}^\dagger; \mathbf{V}) - \mathbf{G}^2(\mathbf{p}; \mathbf{V})\|_\infty = \|g(1) - g(0)\|_\infty \leq \alpha \|\mathbf{p}^\dagger - \mathbf{p}\|_\infty$ with $\alpha = D_1(\check{p}, \hat{p}) < 1$.) ■

Value function iteration. Defining

$$\mathbf{G}^1(\mathbf{V}; \mathbf{p}) = \begin{pmatrix} G_{(1,1)}^1(\mathbf{V}, \mathbf{p}) \\ G_{(2,1)}^1(\mathbf{V}, \mathbf{p}) \\ \vdots \\ G_{(M,M)}^1(\mathbf{V}, \mathbf{p}) \end{pmatrix}$$

we write the P-M algorithm with the policy function held fixed as

$$\mathbf{V}^{k+1} = \mathbf{G}^1(\mathbf{V}^k; \mathbf{p}), \quad k = 0, 1, 2, \dots$$

The following proposition establishes that \mathbf{G}^1 is a contraction, so that the value function iteration converges.

Proposition A4 *Holding fixed $\mathbf{p} \in [\check{p}, \hat{p}]^{M^2}$, where $-\infty < \check{p} \leq \hat{p} < \infty$, \mathbf{G}^1 is a contraction.*

Proof. Recall that

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

Pick $-\infty < \check{V} \leq \hat{V} < \infty$ large enough so that \mathbf{G}^1 maps $[\check{V}, \hat{V}]^{M^2}$ into itself.¹ The proof is completed by applying Blackwell's sufficient conditions (monotonicity and discounting, see e.g. p. 54 of Stokey & Lucas (1989)) to show that \mathbf{G}^1 is a contraction. ■

¹For example, if \check{V} and \hat{V} solve $\check{V} = \min_{p_1 \in [\check{p}, \hat{p}], p_2 \in [\check{p}, \hat{p}]} D_1(p_1, p_2)(p_1 - c(M)) + \beta \check{V}$ and $\hat{V} = \max_{p_1 \in [\check{p}, \hat{p}], p_2 \in [\check{p}, \hat{p}]} D_1(p_1, p_2)(p_1 - c(1)) + \beta \hat{V}$, then $\check{V} \leq G_e^1(\mathbf{V}, \mathbf{p}) \leq \hat{V}$.

A4 Equilibrium correspondence

The value functions in Figure A2 correspond to the policy functions in Figure 4 in the main paper. The smooth value functions in the upper panels are typical for flat equilibria. While its value function is increasing in a firm’s state, it is not decreasing by too much in its rival’s state. Turning to the trenchy and extra-trenchy equilibria, the value functions in the lower panels are much less smooth. Both the leader and the follower experience a rise in value as the industry moves from a state on the diagonal of the state space with extremely intense price competition to an asymmetric state. In other words, the diagonal trench in the policy function is mirrored by a diagonal trench in the value function. Further, in an extra-trenchy equilibrium, the value of being a clear leader is very high while the value of being a distant follower is very low.

Tables A1–A8 give the value and policy functions for our four typical cases.

A5 Industry dynamics: Investment stifling

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 A9, then firms cannot expect to make it down their learning curves. More formally, we take $\bar{\delta}(\rho)$ to be the smallest forgetting rate such that state $(1, 1)$ is the mode of the limiting distribution.

There is clearly a limit to the price-cutting incentives of organizational forgetting. If $\delta > \bar{\delta}(\rho)$, then 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 in a war of attrition—as each firm seeks to force the other to be the first to slide back up its learning curve.

A6 Industry dynamics: Leadership reversals

Figure A3 shows the expected time to a leadership reversal τ^∞ . As can be seen, τ^∞ is largest for intermediate degrees of organizational forgetting. Moreover, τ^∞ is of substantial magnitude, easily reaching and exceeding 1,000 periods. Comparing Figure A3 to Figure 3 in the main paper shows that the limiting expected Herfindahl index H^∞ is also largest for intermediate degrees of organizational forgetting. Hence, asymmetries are persistent in our model because the expected time until the leader and the follower switch roles is (perhaps very) long.

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 A9: Critical value $\bar{\delta}(\rho)$ for investment stifling.

period	modal state	leader					follower				
		cost	prize	price	prob.	value	cost	prize	price	prob.	value
0	(1,1)	10.00	4.42	7.79	0.50	9.40	10.00	4.42	7.79	0.50	9.40
8	(5,5)	6.86	1.33	7.59	0.50	16.68	6.86	1.33	7.59	0.50	16.68
16	(9,9)	5.97	0.60	7.40	0.50	19.44	5.97	0.60	7.40	0.50	19.44
32	(17,17)	5.30	0.00	7.30	0.50	21.00	5.30	0.00	7.30	0.50	21.00
64	(30,30)	5.30	0.00	7.30	0.50	21.00	5.30	0.00	7.30	0.50	21.00
∞	(30,30)	5.30	0.00	7.30	0.50	21.00	5.30	0.00	7.30	0.50	21.00

Table A10: Cost, prize, price, probability of making a sale, and value. Flat equilibrium without well ($\rho = 0.85$, $\delta = 0$).

A7 Industry dynamics: Size of state space

Figure A4 displays the limiting expected Herfindahl index for the cases of $M \in \{20, 30, 40\}$. As can be seen, increasing M , whilst holding fixed δ , facilitates persistent asymmetries ($H^\infty > 0.5$). With a larger state space the industry becomes more likely to remain in the interior of the state space. The resulting bidirectional movements break the backward induction logic that underlies uniqueness of equilibrium for the extreme case of $\delta = 0$ (see Proposition 3 in the main paper).

A8 Pricing behavior: Wells and trenches

As a point of comparison, Table A10 provides details on firms' competitive positions at various points in time for our leading example of a flat equilibrium without well ($\rho = 0.85$, $\delta = 0$). Recall that in a flat equilibrium without well the price that a firm charges is fairly insensitive to its rival's stock of know-how. As time passes both firms thus move down their learning curves in tandem. Initially each firm has an incentive to charge a price below its marginal cost in order to move down its learning curve more quickly. This is reflected in the prize from winning the sale: The prize in state (1, 1) is 4.42 and justifies charging a price of 7.79 that is well below the marginal cost of 10. Eventually, however, the prize vanishes completely as no further improvements in competitive position are possible (see modal state (17, 17) in period 32). Once both firms have reached the bottom of their learning curves equilibrium prices coincide with prices in the static Nash equilibrium in line with part (i) of Proposition 5 in the main paper.

period	flat without well	eqbm. with well	flat with well	eqbm.	trenchy eqbm.	extra-trenchy eqbm.
8	(5, 5)	(2, 8), (8, 2)	(2, 7), (7, 2)	(1, 7), (7, 1)		
16	(9, 9)	(5, 10), (10, 5)	(4, 11), (11, 4)	(1, 10), (10, 1)		
32	(17, 17)	(11, 12), (12, 11)	(8, 16), (16, 8)	(1, 15), (15, 1)		
64	(30, 30)	(17, 17)	(13, 21), (21, 13)	(1, 20), (20, 1)		
∞	(30, 30)	(25, 25)	(20, 28), (28, 20)	(1, 30), (30, 1)		

Table A11: Discount factor of $\beta = 0.995$. Modal states of transient and limiting distributions.

A9 Robustness checks: Discount factor

As we increase the discount factor β we continue to obtain wells and trenches. In fact, the wells and trenches deepen: More patient firms have a stronger incentive to cut prices in the present in order to seize the leadership position in the future. Figure A6 exemplifies the policy functions of the typical equilibria with $\beta = 0.995$; it may be compared to our baseline parameterization with $\beta = \frac{1}{1.05}$ in Figure 4 of the main paper. As can be seen, the basic shapes of the policy functions remain unchanged. Figures A7 and A8 display the transient distribution in period 8 and 32, respectively, and Figure A9 displays the limiting distribution. Table A11 lists the modal states of the transient and limiting distributions. As can be seen, industry structure and dynamics are comparable to those in Figures 5–7 in the main paper.

Very high discount factors exacerbate the multiplicity problem as may be expected in light of the folk theorems for repeated games (e.g., Friedman 1971, Rubinstein 1979, Fudenberg & Maskin 1986). We have used our homotopy algorithm to trace out the equilibrium correspondence by varying the discount factor β while fixing the progress ratio ρ and the forgetting rate δ . Figure A10 displays the limiting expected Herfindahl index H^∞ for $\rho = 0.85$ and $\delta \in \{0, 0.0275, 0.08\}$. We use circles to indicate our leading examples with $\beta = \frac{1}{1.05}$ in Figure 4 in the main paper and our leading examples with $\beta = 0.995$ in Figure A6. Note that the horizontal axis is the discount factor β . For $\delta = 0.0275$ (upper right panel), the flat equilibrium with well for $\beta = \frac{1}{1.05}$ and the trenchy equilibria for $\beta = \frac{1}{1.05}$ and for $\beta = 0.995$ are part of the same path and thus deform smoothly into each other. There are other paths. Our leading example of a flat equilibrium with well for $\beta = 0.995$ in Figure A6 lies on one of these paths. But there are still other paths and equilibria for $\beta = 0.995$ and $\delta = 0.0275$ (we have found a total of 21 equilibria). Similarly, there are many paths and equilibria for $\beta = 0.995$ and $\delta = 0.08$ (we have found a total of 21 equilibria).

Figures A11–A13 exemplify the policy functions of some of these additional equilibria (upper left panel), the transient distributions in period 8 and 32 (upper right and lower left panels), and the limiting distributions (lower right panels). The parameter values are

$\rho = 0.85$, $\delta = 0.0275$, and $\beta = 0.995$. As can be seen, these additional equilibria feature variations of wells and trenches: The equilibria in Figures A11 and A12 exhibit long wells. As usual, a well serves to build, but not to defend, a competitive advantage. The equilibria in Figures A11 and A12 imply $H^\wedge = 0.9171$ and $H^\infty = 0.5003$ and $H^\wedge = 0.9296$ and $H^\infty = 0.5003$, respectively, so that asymmetries are transitory. The equilibrium in Figure A13 exhibits a well around state $(1, 1)$ and extra trenches around states $(21, 27)$ and $(28, 22)$ that may lead to persistent asymmetries ($H^\wedge = 0.7109$ and $H^\infty = 0.5696$).

A10 Robustness checks: Frequency of sales

In our model the discount factor β plays two roles. First, it captures firms' time preference. This is the interpretation that we take in Section A9: A higher β means more patient firms. Second, β determines the length of a period. Following Cabral & Riordan (1994), we use a period just long enough for a firm to make a sale where a typical sale consists of an order for z units of the good. Thus, if buyers cut the size of an order in half while keeping the total volume demanded per unit of time constant, then the period length is cut in half and the frequency of sales is doubled. Modeling this requires more than increasing the discount factor; it also requires expanding the state space and changing the specification of learning and forgetting. In order to pin down the the nature of these changes, assume that organizational forgetting operates through labor turnover. This implies that the probability of forgetting a given amount of know-how is proportional to the length of a period. Assume further that the reduction in marginal cost from learning-by-doing is proportional to the size of an order.

To capture this increased frequency of sales we divide a period into $K > 1$ subperiods, each with one order of $\frac{z}{K}$ units. Thus, if r is the discount rate per period, then $\frac{r}{K}$ is the discount rate per subperiod and $\beta = \frac{1}{1+\frac{r}{K}}$ is the discount factor. We further "fill in" the state space to ensure that the reduction in marginal cost that is achievable by a period's worth of sales in the original specification is comparable to the reduction that is achievable by K subperiods' worth of sales in the alternative specification. We similarly adjust the probability of forgetting to make the loss from organizational forgetting that can occur in one period of the original specification comparable to the expected loss that can occur in K subperiods of the alternative specification. 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 $\Delta\left(\frac{e_n-1}{K} + 1\right)$, requiring us to expand the state space to $\{1, \dots, K(M-1) + 1\}^2$. 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 $K = 2$. Figure A14 exemplifies the policy functions of the typical equilibria. As can be seen, 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

period	subperiod	flat without well	eqbm. with well	flat	eqbm.	trenchy eqbm.	extra-trenchy eqbm.
8	16	(5, 5)	(3, 6.5), (6.5, 3)	(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)	(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)	(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)	(18, 21.5), (21.5, 18)	(1, 18), (18, 1)	
∞	∞	(30, 30)	(29, 30), (30, 29)	(26, 30), (30, 26)	(26, 30), (30, 26)	(1, 24), (24, 1)	

Table A12: Frequency of sales with $K = 2$. Modal states of transient and limiting distributions.

equilibrium ($\delta = 0.09$), similar to the four typical cases in Figure 4. Figures A15 and A16 display the transient distribution in period 8 and 32 (subperiod 16 and 64), respectively, and Figure A17 displays the limiting distribution for the four typical cases with $K = 2$. Table A12 lists the modal states of the transient and limiting distributions. As can be seen, industry structure and dynamics are comparable to those in Figures 5–7 in the main paper.² Overall, it appears that our results are not sensitive to the frequency of sales.

Reducing the size of an order while keeping the total volume demand per unit of time constant—as we have just done—is one way to shorten the length of a period and increase the frequency of sales. Another way is to increase total volume while keeping order size constant. This restores the original specification of $c(e_n)$ and m . Hence, as the frequency of sales K increases the discount factor β approaches 1, the probability of forgetting approaches 0, and the expected time it takes a firm to reach the bottom of its learning curve approaches zero.

It has been computationally infeasible for us to explore this latter case for some moderate value of K such as 10 because the size of the state space explodes. This is regrettable because, building on earlier work by Spence (1981), Cabral & Riordan (1994) show in their Theorem 3.2 that, without forgetting, as $\beta \rightarrow 1$ firms price *as if* at they had already reached the bottom of their learning curves. Intuitively, the relevant marginal cost for an infinitely patient firm is always $c(m)$ because the time before the firm reaches the bottom of its learning curve pales in comparison to the time after. Thus, in the absence of organizational forgetting, an extreme example of a flat equilibrium is obtained in which $p^*(\mathbf{e}) = p^\dagger(m, m)$ for all states $\mathbf{e} \in \{1, \dots, M\}^2$. Our conjecture is that, in the presence of organizational forgetting, this same equilibrium is approached as K becomes large, though the existence of sunspots suggests that there may also be other equilibria.

²With 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 $K = 1$.

A11 Robustness checks: 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 we decrease σ from 1 to 0.2, the respective elasticities become -102.00 , -0.00 , 0.00 , and 55.0 . As we increase σ from 1 over 2 to 10, the respective 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 $\sigma = 10$.

Figure A18 displays the limiting and maximum expected Herfindahl indices for the case of weaker product differentiation with $\sigma = 0.2$. Figures A19 and A20 do the same for the case of stronger product differentiation with $\sigma \in \{2, 10\}$. These figures may be compared to our baseline parameterization with $\sigma = 1$ in Figure 3 in the main paper. As can be seen, 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.

A12 Robustness checks: 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

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

where we assume that the outside good is supplied under conditions of perfect competition with price equal to marginal cost, $p_0 = c_0$. We compute the price elasticity of aggregate demand as the percentage change in aggregate demand $\sum_{n=1}^2 D_n(\mathbf{p})$ that results from a one-percent change in both prices p_1 and p_2 :

$$\eta(\mathbf{p}) = \frac{-D_1(\mathbf{p})(1 - D_1(\mathbf{p}))p_1 + D_1(\mathbf{p})D_2(\mathbf{p})(p_1 + p_2) - D_2(\mathbf{p})(1 - D_2(\mathbf{p}))p_2}{\sigma(D_1(\mathbf{p}) + D_2(\mathbf{p}))}.$$

As $v_0 - c_0 \rightarrow -\infty$, $D_0(\mathbf{p}) = 1 - \sum_{n=1}^2 D_n(\mathbf{p}) \rightarrow 0$ and $\eta(\mathbf{p}) \rightarrow 0$, and we revert to the Cabral & Riordan (1994) setting in which the buyer always purchases from one of the two firms in the industry and the price elasticity of aggregate demand is zero.

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 $\rho = 0.85$. The price elasticity of aggregate demand is no longer zero and ranges from -7.08 in state (1, 1) over -3.33 in state (2, 2) and -1.35 in state (4, 4) to -0.22 in state (15, 15). To further increase the attractiveness of the outside good we set $v_0 - c_0 \in \{3, 5, 10\}$. If $v_0 - c_0 = 3$, then the share of the outside good is quite large in general and declines from 0.97 in state (1, 1) over 0.87 in state (2, 2) and 0.67 in state (4, 4) to 0.30 in state (15, 15). The price elasticity of aggregate demand ranges from -10.63 in state (1, 1) over -8.03 in state (2, 2) and -5.68 in state (4, 4) to -2.04 in state (15, 15).

Figure A21 illustrates the extent of multiplicity for the case of an outside good with $v_0 - c_0 = 0$. It shows the number of equilibria for each combination of forgetting rate δ and progress ratio ρ . Darker shades indicate more equilibria. This figure may be compared to our baseline parameterization with $v_0 - c_0 = -\infty$ in Figure 2 in the main paper. As can be seen, multiple equilibria continue to arise in the presence of an outside good, although less frequently as the outside good becomes more attractive. We have found up to nine equilibria for some values of ρ and δ . Because the outside good sufficiently constrains firms' pricing behavior we no longer have sunspots for a progress ratio of $\rho = 1$.

Figure A22 displays the limiting and maximum expected Herfindahl indices for the case an outside good with $v_0 - c_0 = 0$. Figures A23–A25 do the same for the case of a more attractive outside good with $v_0 - c_0 \in \{3, 5, 10\}$. These figures may be compared to our baseline parameterization with $v_0 - c_0 = -\infty$ in Figure 3 in the main paper. As can be seen from Figure A25, 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. Despite the reduced extent of multiplicity, the basic shape of the equilibrium correspondence remains largely the same: For intermediate degrees of organizational forgetting, asymmetries arise and persist.

The types of equilibria that arise (and the dynamics that they imply) are also the same: For the case of an outside good with $v_0 - c_0 = 0$, Figure A26 shows the analogs to our leading examples of a flat equilibrium without well, a flat equilibrium with well, a trenchy equilibrium, and an extra-trenchy equilibrium; it may be compared to Figure 4 in the main paper. To obtain the analogs to our leading examples for the case of a more attractive outside good with $v_0 - c_0 = 3$ in Figure A27, we increase the discount factor β to $\frac{1}{1.01}$ from its baseline value of $\frac{1}{1.05}$. We emphasize that in these equilibria the price elasticity of aggregate demand is economically significant. The reason that we have to increase the discount factor is this: Since the price elasticity of aggregate demand in state (1, 1) is very high (-10.63 in the static Nash equilibrium), the probability of making a sale and moving down the learning curve is very low (the share of the outside good is 0.97). At the same

time, however, there are rewards to be had from fighting one's way down the learning curve: The price elasticity of aggregate demand in state (15, 15) is much lower (-2.04 in the static Nash equilibrium) than in state (1, 1), indicating that the outside good is a much less formidable competitor to a firm at the bottom of its learning curve than to a firm at the top. But in combination with a nonnegligible forgetting rate ($\delta \in \{0.035, 0.045\}$), the rewards of pricing aggressively are too far off into the future to justify the cost of doing so unless firms are more patient than in our baseline parameterization.

A13 Robustness checks: 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, in the absence of organizational forgetting, must therefore eventually reach the bottom of its learning curve. This precludes long-run market dominance to occur in the absence of organizational forgetting.

Suppose instead that the probability that firm n makes a sale is given by the linear specification

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

Due to the choke price in the linear specification, a firm is able to surely deny its rival a sale by pricing sufficiently aggressively. Note that we choose the slope parameter in the linear specification so that in the static Nash equilibrium the own-price elasticity of demand in state (1, 1) is the same as in the logit specification in order to allow for a fair comparison between linear and logit demand.

Figure A28 displays the limiting and maximum expected Herfindahl indices for logit demand (right panels) and linear demand (left panel) and various degrees of product differentiation. Note that the horizontal axis is the progress ratio ρ . As can be seen, for linear demand with $\sigma = 1$ the industry evolves towards a symmetric structure in the absence of organizational forgetting ($\delta = 0$). This is no longer the case with $\sigma = 0.2$: Asymmetries can arise and persist in the model with a linear demand even in the absence of organizational forgetting.

Figure A29 exemplifies the policy function of a flat equilibrium with well (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 $\rho = 0.85$, $\delta = 0$, and $\sigma = 0.2$. As can be seen, 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. In other words, the choke price is a shut-out model element.

Yet, in the absence of organizational forgetting, we never found a trenchy or extra-

period	asymmetric		symmetric	
	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 A13: Bottomless learning. Most-likely asymmetric and symmetric states of transient and limiting distributions. Plateau equilibrium 1 ($\rho = 0.9$, $\delta = 0.04$).

trenchy equilibrium with linear demand. 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.

A14 Robustness checks: 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 $m = M$. Figure A30 displays the limiting and maximum expected Herfindahl indices for this *bottomless learning* specification. This figure may be compared to our baseline parameterization with $m = 15 < M$ in Figure 3 in the main paper.

We obtain other types of equilibria in the bottomless learning specification in addition to the four typical cases in Figure 4. Figure A31 exemplifies the policy function of a *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 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 A13 summarizes the dynamics of the industry by providing the most-likely asymmetric and symmetric states of the transient and limiting distributions.

period	asymmetric		symmetric	
	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 A14: Bottomless learning. Most-likely asymmetric and symmetric states of transient and limiting distributions. Plateau equilibrium 2 ($\rho = 0.9$, $\delta = 0.04$).

As can be seen, the likelihood of cooperation goes up with time.

Figure A32 provides another example of a plateau equilibrium and Table A14 summarizes the dynamics of the industry. The parameter values are the same as before ($\rho = 0.9$ 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. That is, the likelihood of cooperation goes down with time.

A15 Robustness checks: 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.

Figure A33 illustrates the extent of multiplicity for this *constant forgetting* specification with $\Delta(e_n) = \delta$.³ It shows the number of equilibria for each combination of forgetting rate δ and progress ratio ρ . Darker shades indicate more equilibria. Note that the horizontal axis is on a linear scale. As can be seen, we have found up to eleven equilibria for some values of δ and ρ . Multiplicity is especially pervasive for forgetting rates δ between 0.4 and 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.

Our results regarding market dominance also carry over to the *constant forgetting* specification. Figure A34 displays the limiting and maximum expected Herfindahl indices for the constant forgetting specification. Note that the horizontal axis is on a linear scale. As can be seen, if organizational forgetting is sufficiently weak, then asymmetries may arise

³Figure A33 looks somewhat rough because we use a grid of 20 rather than 100 values of $\rho \in (0, 1]$.

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.

A16 Robustness checks: Entry and exit

Below we describe the N -firm version of our model with entry and exit. 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.

To ensure the existence of an equilibrium, we use the approach in Doraszelski & Satterthwaite (2008) and assume that salvage values and set-up costs are privately observed. Since the analysis of entry and exit requires a well-posed monopoly problem, we include an outside good.

Order of moves. In each period the sequence of events is as follows:

1. Each of the N^* incumbent firms learns its own salvage value and makes an exit decision. Each of the $N - N^*$ potential entrants learns its own set-up cost and makes an entry decision. Entry and exit decisions are made simultaneously. In this process, the industry transits from state \mathbf{e} to state \mathbf{e}' . Specifically, incumbent firm n transits from state $e_n \neq 0$ to state $e'_n = 0$ upon exiting and potential entrant n transits from state $e_n = 0$ to state $e'_n = e^0 \neq 0$ upon entering the industry.
2. Price competition takes place among active firms, where firm n is active if and only if $e'_n \neq 0$. Learning-by-doing and organizational forgetting occur. In this process, the industry transits from state \mathbf{e}' to state \mathbf{e}'' .

Before making their entry and exit decisions, all firms observe state \mathbf{e} , and all firms observe state \mathbf{e}' prior to making their pricing decisions.

Entry and exit. Before price competition takes place, incumbent firms can choose to exit the industry and potential entrants can choose to enter it. If an incumbent firm exits the industry, it receives a salvage value and perishes. We assume that at the beginning of each period, incumbent firm n draws a salvage value X_n from a symmetric triangular

distribution $G_X(\cdot)$ with support $[\bar{X} - a, \bar{X} + a]$, where $a > 0$ is a parameter. That is, letting $Z_n = \frac{X_n - \bar{X}}{a}$, the density and distribution of X_n are given by

$$g_X(X_n) = \begin{cases} 0 & \text{if } Z_n < -1, \\ \frac{1}{a}(1 - |Z_n|) & \text{if } -1 \leq Z_n < 1, \\ 0 & \text{if } Z_n \geq 1, \end{cases}$$

$$G_X(X_n) = \begin{cases} 0 & \text{if } Z_n < -1, \\ \frac{1}{2}(1 + 2Z_n + Z_n^2) & \text{if } -1 \leq Z_n < 0, \\ \frac{1}{2}(1 + 2Z_n - Z_n^2) & \text{if } 0 \leq Z_n < 1, \\ 1 & \text{if } Z_n \geq 1. \end{cases}$$

Salvage values are independently and identically distributed across firms and periods, and firm n 's realization is observed only by itself but not by its rivals. Let $\tau_n(\mathbf{e}, X_n) = 1$ denote the decision of incumbent firm n to remain in the industry in state \mathbf{e} when it has drawn salvage value X_n , while $\tau_n(\mathbf{e}, X_n) = 0$ denotes the decision to exit.

Simultaneous with the exit decisions of incumbent firms, potential entrants make entry decisions. If a potential entrant decides not to enter, it receives nothing and perishes; if it enters, it incurs a set-up cost. At the beginning of each period, potential entrant n draws a set-up cost S_n from a symmetric triangular distribution $G_S(\cdot)$ with support $[\bar{S} - b, \bar{S} + b]$, where $b > 0$ is a parameter. Set-up costs are independently and identically distributed across firms and periods, and its realization is private to a firm. Let $\tau_n(\mathbf{e}, S_n) = 1$ denote the decision of potential entrant n to enter the industry in state \mathbf{e} when it has drawn set-up cost S_n , while $\tau_n(\mathbf{e}, S_n) = 0$ denotes the decision to stay out.

Combining the firms' entry and exit decisions, let $\lambda_n(\mathbf{e})$ denote the probability that firm n operates in the industry in state \mathbf{e} . If $e_n \neq 0$ so that firm n is an incumbent, then $\lambda_n(\mathbf{e}) = \int \tau_n(\mathbf{e}, X_n) dG_X(X_n)$. If $e_n = 0$ so that firm n is an entrant, then $\lambda_n(\mathbf{e}) = \int \tau_n(\mathbf{e}, S_n) dG_S(S_n)$.

Bellman equation. To develop the Bellman equation, we first consider firms' pricing decisions. We then consider the exit decisions of incumbent firms and the entry decisions of potential entrants. Throughout we use $V_n(\mathbf{e})$ to denote the expected net present value of future cash flows to firm n in state \mathbf{e} *before* entry and exit decisions have been made. In addition, we use $U_n(\mathbf{e}')$ to denote the expected net present value of future cash flows to active firm n in state \mathbf{e}' *after* entry and exit decisions have been made.

Pricing decisions. Consider an industry that, via a process of entry and exit, has transitioned from state \mathbf{e} to state \mathbf{e}' . The expected net present value of future cash flows to active firm n is given by

$$U_n(\mathbf{e}') = \max_{p_n} D_n(p_n, \mathbf{p}_{-n}(\mathbf{e}'))(p_n - c(e'_n)) + \beta \sum_{k=0}^N D_k(p_n, \mathbf{p}_{-n}(\mathbf{e}')) \bar{V}_{nk}(\mathbf{e}'), \quad (\text{A10})$$

where $p_{-n}(\mathbf{e}')$ denotes the prices charged by the other firms in state \mathbf{e}' and $\bar{V}_{nk}(\mathbf{e})$ is the expectation of firm n 's value function conditional on the buyer purchasing good $k \in \{0, 1, \dots, N\}$ as given by

$$\begin{aligned}\bar{V}_{n0}(\mathbf{e}') &= \sum_{e'_1=e'_1-1}^{e'_1} \cdots \sum_{e'_N=e'_N-1}^{e'_N} V_n(\mathbf{e}'') \prod_{i=1}^N \Pr(e''_i | e'_i, 0), \\ \bar{V}_{nk}(\mathbf{e}') &= \sum_{e'_1=e_1-1}^{e'_1} \cdots \sum_{e'_{k-1}=e'_{k-1}-1}^{e'_{k-1}} \sum_{e''_k=e'_k}^{e'_k+1} \sum_{e''_{k+1}=e'_{k+1}-1}^{e'_{k+1}} \cdots \sum_{e'_N=e'_N-1}^{e'_N} V_n(\mathbf{e}'') \\ &\quad \prod_{j \neq k} \Pr(e''_j | e'_j, 0) \Pr(e''_k | e'_k, 1), \quad k \in \{1, \dots, N\}.\end{aligned}$$

Note that we include an outside good (good 0) in the specification to ensure a well-posed monopoly problem.

Let $h_n(\mathbf{e}', p_n, \mathbf{p}_{-n}(\mathbf{e}'), \mathbf{V}_n)$ denote the maximand in equation (A10). Using the same argument as in Section 2 in the main paper, if the FOC $\frac{\partial h_n(\cdot)}{\partial p_n} = 0$ 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 $\mathbf{p}_{-n}(\mathbf{e}')$). If firm n is inactive, we assign $p_n(\mathbf{e}') = \infty$.

Exit decisions. To develop the Bellman equation determining $V_n(\mathbf{e})$, consider the exit decision $\tau_n(\mathbf{e}, X_n)$ of incumbent firm n who has drawn salvage value X_n . It remains in the industry in state \mathbf{e} if its realized salvage value is less than or equal to the expected value of continuing forward to the price-setting stage:

$$\tau_n(\mathbf{e}, X_n) = \begin{cases} 1 & \text{if } X_n \leq \hat{X}_n(\mathbf{e}), \\ 0 & \text{if } X_n \geq \hat{X}_n(\mathbf{e}), \end{cases}$$

where

$$\hat{X}_n(\mathbf{e}) = E[U_n(\mathbf{e}') | \mathbf{e}, e'_n = e_n, \boldsymbol{\lambda}_{-n}(\mathbf{e})]$$

is the expected value to incumbent firm n of continuing forward to the price-setting stage as an active firm with its current stock of know-how, i.e., $e'_n = e_n$, taking into account the operating probabilities $\boldsymbol{\lambda}_{-n}(\mathbf{e})$ of the other firms. $\hat{X}_n(\mathbf{e})$ is computed as

$$\begin{aligned}\sum_{e'_1 \in \mathcal{E}'_1} \cdots \sum_{e'_{n-1} \in \mathcal{E}'_{n-1}} \sum_{e'_{n+1} \in \mathcal{E}'_{n+1}} \cdots \sum_{e'_N \in \mathcal{E}'_N} U_n(e'_1, \dots, e'_{n-1}, e_n, e'_{n+1}, \dots, e'_N) \\ \prod_{k \neq n, e'_k \neq 0} \lambda_k(\mathbf{e}) \prod_{k \neq n, e'_k = 0} (1 - \lambda_k(\mathbf{e})),\end{aligned}$$

where

$$\mathcal{E}'_n = \begin{cases} \{0, e_n\} & \text{if } e_n \neq 0, \\ \{0, e^0\} & \text{if } e_n = 0. \end{cases}$$

The expected net present value of future cash flows $V_n(\mathbf{e}, X_n)$ to incumbent firm n who has drawn salvage value X_n is given by

$$V_n(\mathbf{e}, X_n) = \max \left\{ \widehat{X}_n(\mathbf{e}), X_n \right\}.$$

Integrating over all possible salvage values yields the value function $V_n(\mathbf{e}) = \int V_n(\mathbf{e}, X_n) dG_X(X_n)$ for incumbent firm n in state \mathbf{e} . Letting $\widehat{Z}_n(\mathbf{e}) = \frac{\widehat{X}_n(\mathbf{e}) - \bar{X}}{a}$ we have

$$V_n(\mathbf{e}) = \begin{cases} \bar{X} & \text{if } \widehat{Z}_n(\mathbf{e}) < -1, \\ \bar{X} + \frac{a}{6} \left(1 + 3\widehat{Z}_n(\mathbf{e}) + 3(\widehat{Z}_n(\mathbf{e}))^2 + (\widehat{Z}_n(\mathbf{e}))^3 \right) & \text{if } -1 \leq \widehat{Z}_n(\mathbf{e}) < 0, \\ \bar{X} + \frac{a}{6} \left(1 + 3\widehat{Z}_n(\mathbf{e}) + 3(\widehat{Z}_n(\mathbf{e}))^2 - (\widehat{Z}_n(\mathbf{e}))^3 \right) & \text{if } 0 \leq \widehat{Z}_n(\mathbf{e}) < 1, \\ \bar{X} + a\widehat{Z}_n(\mathbf{e}) & \text{if } \widehat{Z}_n(\mathbf{e}) \geq 1, \end{cases} \quad (\text{A11})$$

where, recall, the support of salvage values is $[\bar{X} - a, \bar{X} + a]$.

Since salvage values are private, from the point of view of the other firms, the probability that incumbent firm n remains in the industry is

$$\lambda_n(\mathbf{e}) = G_X(\widehat{X}_n(\mathbf{e})) = \begin{cases} 0 & \text{if } \widehat{Z}_n(\mathbf{e}) < -1, \\ \frac{1}{2} \left(1 + 2\widehat{Z}_n(\mathbf{e}) + (\widehat{Z}_n(\mathbf{e}))^2 \right) & \text{if } -1 \leq \widehat{Z}_n(\mathbf{e}) < 0, \\ \frac{1}{2} \left(1 + 2\widehat{Z}_n(\mathbf{e}) - (\widehat{Z}_n(\mathbf{e}))^2 \right) & \text{if } 0 \leq \widehat{Z}_n(\mathbf{e}) < 1, \\ 1 & \text{if } \widehat{Z}_n(\mathbf{e}) \geq 1, \end{cases} \quad (\text{A12})$$

Entry decisions. Consider the entry decision $\tau_n(\mathbf{e}, S_n)$ of potential entrant n who has drawn set-up cost S_n . It enters the industry in state \mathbf{e} if its realized set-up cost is less than or equal to the expected value of continuing forward to the price-setting stage:

$$\tau_n(\mathbf{e}, S_n) = \begin{cases} 1 & \text{if } S_n \leq \widehat{S}_n(\mathbf{e}'), \\ 0 & \text{if } S_n \geq \widehat{S}_n(\mathbf{e}'), \end{cases}$$

where

$$\widehat{S}_n(\mathbf{e}) = E[U_n(\mathbf{e}') | \mathbf{e}, e'_n = e^0, \boldsymbol{\lambda}_{-n}(\mathbf{e})]$$

is the expected value of continuing forward to the price-setting stage as an active firm with the initial stock of know-how, i.e., $e'_n = e^0$, taking into account the operating probabilities $\boldsymbol{\lambda}_{-n}(\mathbf{e})$ of the other firms. $\widehat{S}_n(\mathbf{e})$ is computed analogously to $\widehat{X}_n(\mathbf{e})$.

The expected net present value of future cash flows $V_n(\mathbf{e}, S_n)$ to potential entrant n who has drawn set-up cost S_n is given by

$$V_n(\mathbf{e}, S_n) = \max \left\{ \widehat{S}_n(\mathbf{e}) - S_n, 0 \right\}.$$

Integrating over all possible set-up costs yields the value function $V_n(\mathbf{e}) = \int V_n(\mathbf{e}, S_n) dG_S(S_n)$ for potential entrant n in state \mathbf{e} . Letting $\widehat{Z}_n(\mathbf{e}) = \frac{\widehat{S}_n(\mathbf{e}) - \bar{S}}{b}$ we have

$$V_n(\mathbf{e}) = \begin{cases} 0 & \text{if } \widehat{Z}_n(\mathbf{e}) < -1, \\ \frac{b}{6} \left(1 + 3\widehat{Z}_n(\mathbf{e}) + 3\left(\widehat{Z}_n(\mathbf{e})\right)^2 + \left(\widehat{Z}_n(\mathbf{e})\right)^3 \right) & \text{if } -1 \leq \widehat{Z}_n(\mathbf{e}) < 0, \\ \frac{b}{6} \left(1 + 3\widehat{Z}_n(\mathbf{e}) + 3\left(\widehat{Z}_n(\mathbf{e})\right)^2 - \left(\widehat{Z}_n(\mathbf{e})\right)^3 \right) & \text{if } 0 \leq \widehat{Z}_n(\mathbf{e}) < 1, \\ b\widehat{Z}_n(\mathbf{e}) & \text{if } \widehat{Z}_n(\mathbf{e}) \geq 1, \end{cases} \quad (\text{A13})$$

where, recall, the support of set-up costs is $[\bar{S} - b, \bar{S} + b]$.

Finally, from the point of view of the other firms, the probability that potential entrant n enters the industry is

$$\lambda_n(\mathbf{e}) = G_S(\widehat{S}_n(\mathbf{e})) = \begin{cases} 0 & \text{if } \widehat{Z}_n(\mathbf{e}) < -1, \\ \frac{1}{2} \left(1 + 2\widehat{Z}_n(\mathbf{e}) + \left(\widehat{Z}_n(\mathbf{e})\right)^2 \right) & \text{if } -1 \leq \widehat{Z}_n(\mathbf{e}) < 0, \\ \frac{1}{2} \left(1 + 2\widehat{Z}_n(\mathbf{e}) - \left(\widehat{Z}_n(\mathbf{e})\right)^2 \right) & \text{if } 0 \leq \widehat{Z}_n(\mathbf{e}) < 1, \\ 1 & \text{if } \widehat{Z}_n(\mathbf{e}) \geq 1, \end{cases} \quad (\text{A14})$$

Equilibrium. We restrict ourselves to symmetric and anonymous Markov perfect equilibria. Symmetry allows us to focus on the problem of firm 1 and anonymity (also called exchangeability) says that firm 1 does not care about the identity of its rivals, only about the distribution of their states (see, e.g., Doraszelski & Satterthwaite (2008) for a formal definition). It therefore suffices to determine the value and policy functions of firm 1, and we define $V^*(\mathbf{e}) = V_1(\mathbf{e})$, $p^*(\mathbf{e}) = p_1(\mathbf{e})$, and $\lambda^*(\mathbf{e}) = \lambda_1(\mathbf{e})$ for each state \mathbf{e} . The corresponding value and policy functions for firm n in state \mathbf{e} are recovered as $V_n(\mathbf{e}) = V^*(\mathbf{e}^{[n]})$, $p_n(\mathbf{e}) = p^*(\mathbf{e}^{[n]})$, and $\lambda_n(\mathbf{e}) = \lambda^*(\mathbf{e}^{[n]})$, where $\mathbf{e}^{[n]}$ is constructed from \mathbf{e} by interchanging the stocks of know-how of firms 1 and n .

Parameterization. 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 and set $e^0 = 1$. We set $v = 10$ and $v_0 - c_0 = 0$ to include an outside good. We further set $\bar{X} = 1.5$ and $a = 1.5$ to ensure that salvage values are drawn from a symmetric triangular distribution with support $[0, 3]$ and $\bar{S} = 4.5$ and $b = 1.5$ to ensure that set-up costs are drawn from a symmetric triangular distribution with support $[3, 6]$.⁴

Results. Organizational forgetting remains a source of aggressive pricing behavior, market dominance, and multiple equilibria in the general model with entry and exit. The

⁴This 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.

possibility of 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 (see below for a concrete example). Figure A35 displays the limiting and maximum expected Herfindahl indices for the general model with entry and exit. Because our parameterization includes an outside good with $v_0 - c_0 = 0$, this figure may be compared to Figure A22.

As Theorem 4.1 in Cabral & Riordan (1994) shows, entry and exit may give rise to multiple equilibria even in the absence of organizational forgetting. For a progress ratio of $\rho = 0.75$, for example, we found three equilibria in the absence of organizational forgetting ($\delta = 0$), in contrast to Proposition 3. While these equilibria are flat either without or with well, the implied long-run industry structures range from symmetric (with the modal 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 $\mathbf{e} \in \{1, \dots, M\}^2$). 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). Because entry and exit are shut-out model elements, asymmetries can arise and persist even in the absence of organizational forgetting.

References

- Benkard, L. (2000), ‘Learning and forgetting: The dynamics of aircraft production’, *American Economic Review* **90**(4), 1034–1054.
- Bertsekas, D. & Tsitsiklis, J. (1997), *Parallel and distributed computation: Numerical methods*, Athena Scientific, Belmont.
- Cabral, L. & Riordan, M. (1994), ‘The learning curve, market dominance, and predatory pricing’, *Econometrica* **62**(5), 1115–1140.
- Doraszelski, U. & Satterthwaite, M. (2008), Computable Markov-perfect industry dynamics, Working paper, Harvard University, Cambridge.
- Friedman, J. (1971), ‘A non-cooperative equilibrium for supergames’, *Review of Economic Studies* **38**(1), 1–12.

- Fudenberg, D. & Maskin, E. (1986), 'The folk theorem in repeated games with discounting or with incomplete information', *Econometrica* **54**(3), 533–554.
- 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.
- Rubinstein, A. (1979), 'Equilibrium in supergames with the overtaking criterion', *Journal of Economic Theory* **21**(1), 1–9.
- Spence, M. (1981), 'The learning curve and competition', *Bell Journal of Economics* **12**(1), 49–70.
- Stokey, N. & Lucas, R. with Prescott, E. (1989), *Recursive methods in economic dynamics*, Harvard University Press, Cambridge.

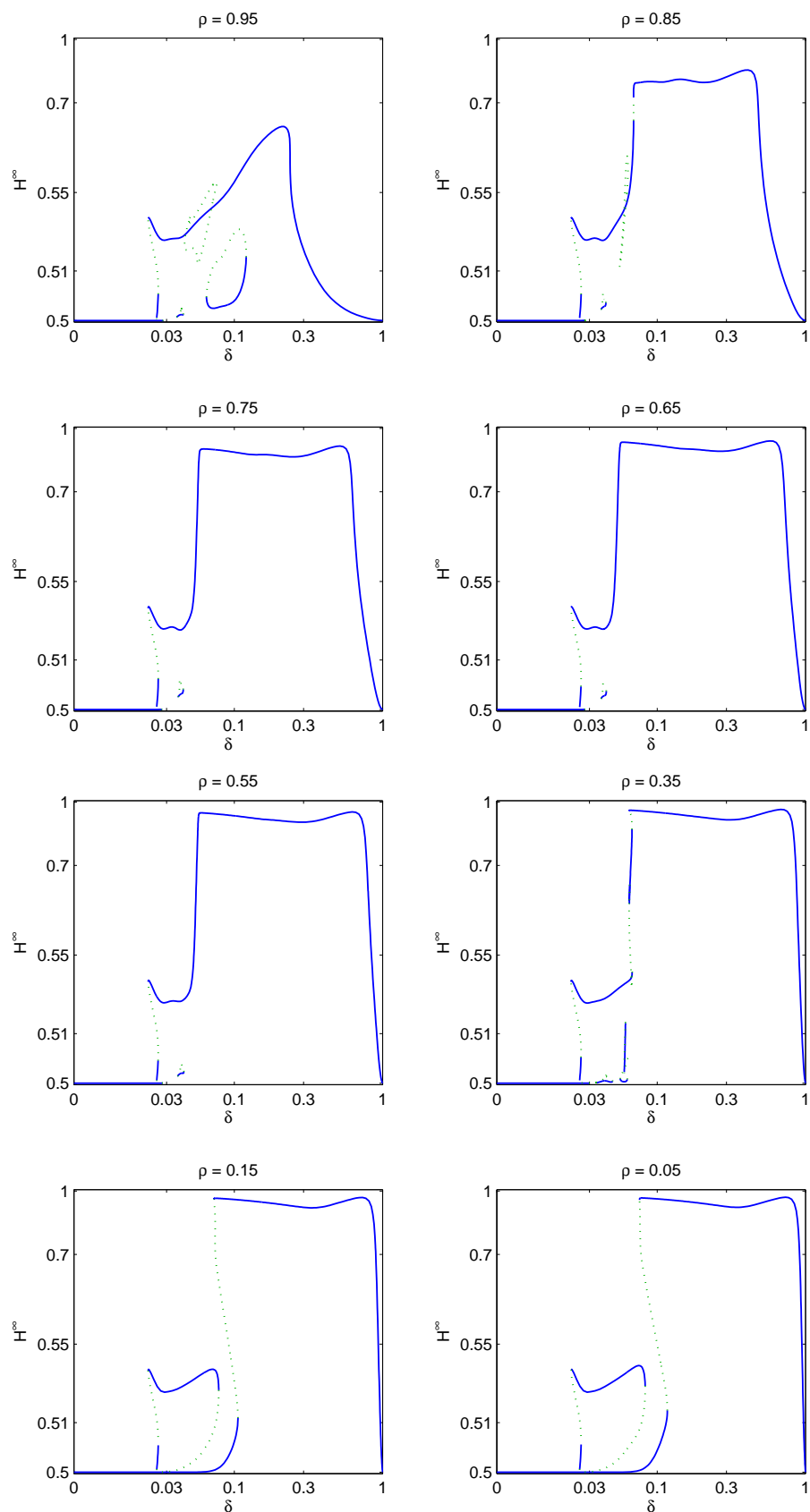


Figure A1: Limiting expected Herfindahl index H^∞ . Equilibria with $\varrho \left(\left. \frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}} \right|_{(\delta(s), \rho)} \right) < 1$ (solid line) and equilibria with $\varrho \left(\left. \frac{\partial \mathbf{G}(\mathbf{x}(s))}{\partial \mathbf{x}} \right|_{(\delta(s), \rho)} \right) \geq 1$ (dotted line).

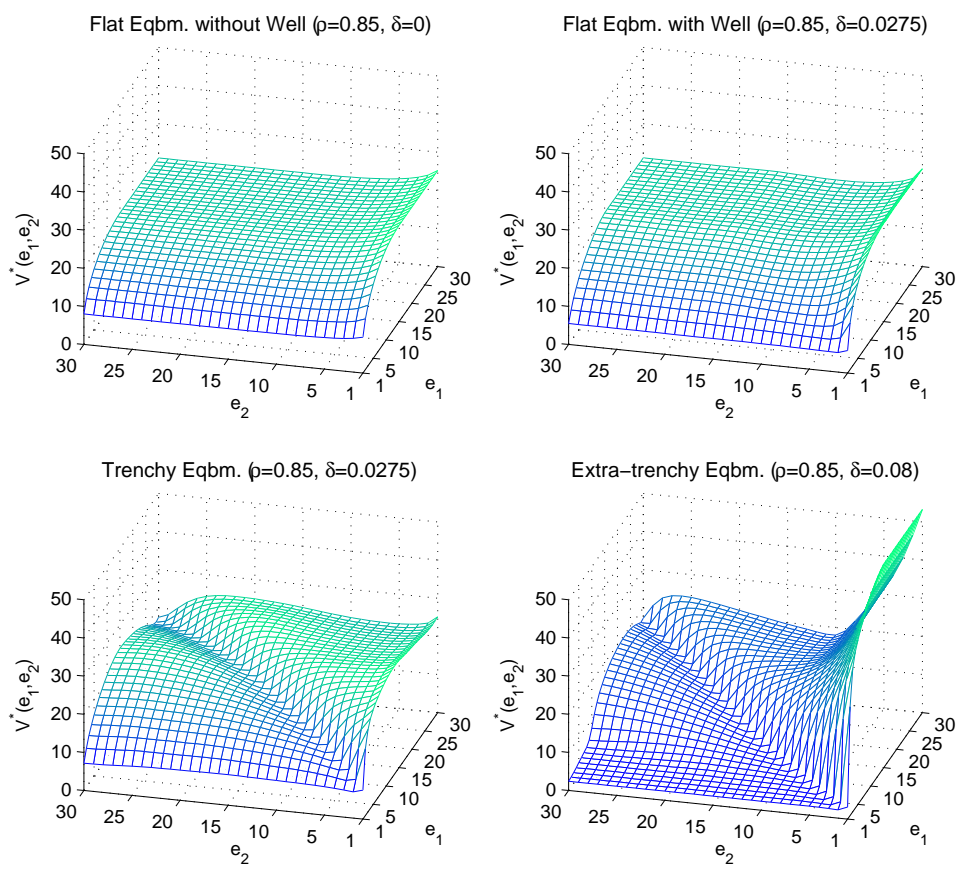


Figure A2: Value function $V^*(e_1, e_2)$.

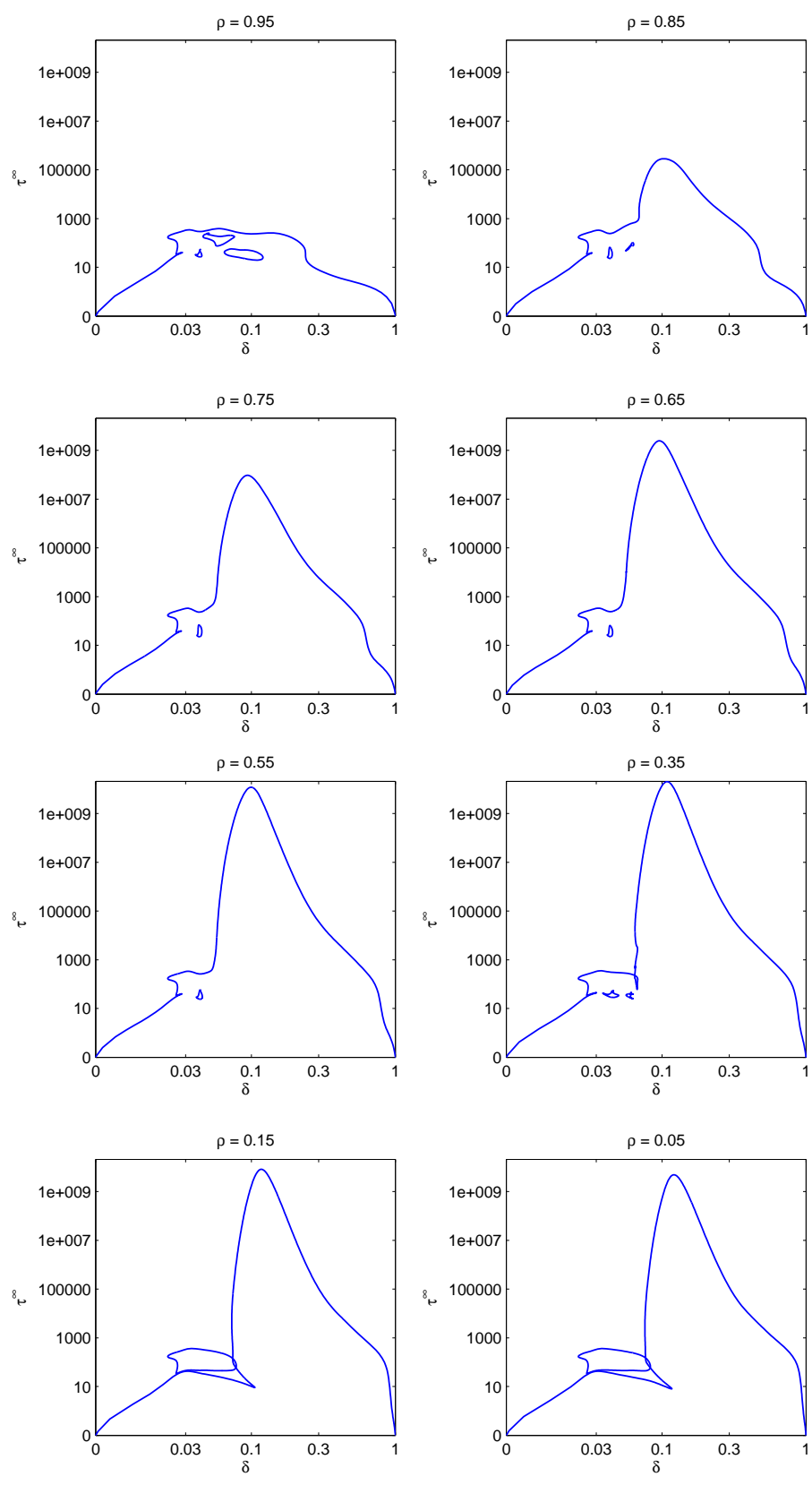


Figure A3: Expected time to leadership reversal τ^∞ .

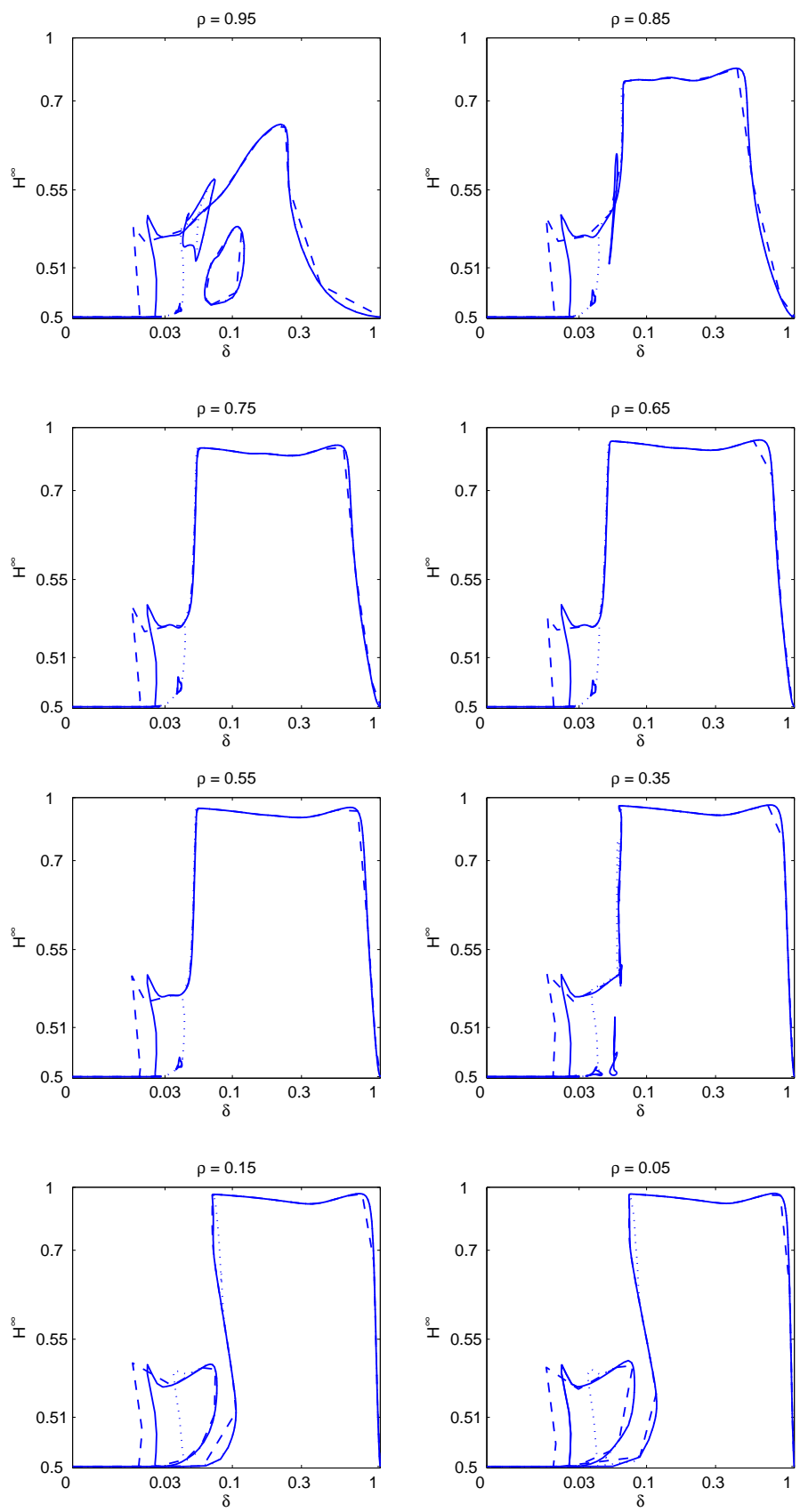


Figure A4: Size of state space with $M \in \{20, 30, 40\}$. Limiting expected Herfindahl index H^∞ for $M = 20$ (dotted line), $M = 30$ (solid line), and $M = 40$ (dashed line).

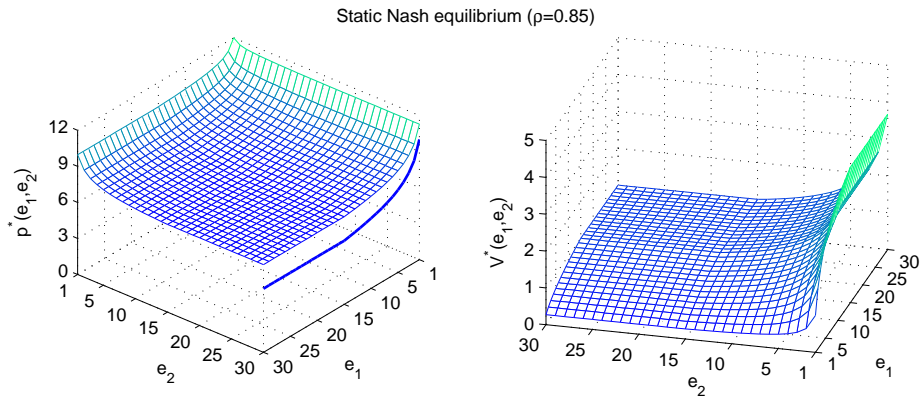


Figure A5: Discount factor of $\beta = 0$ (static Nash equilibrium). Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane) (right panel). Value function $V^*(e_1, e_2)$ (left panel).

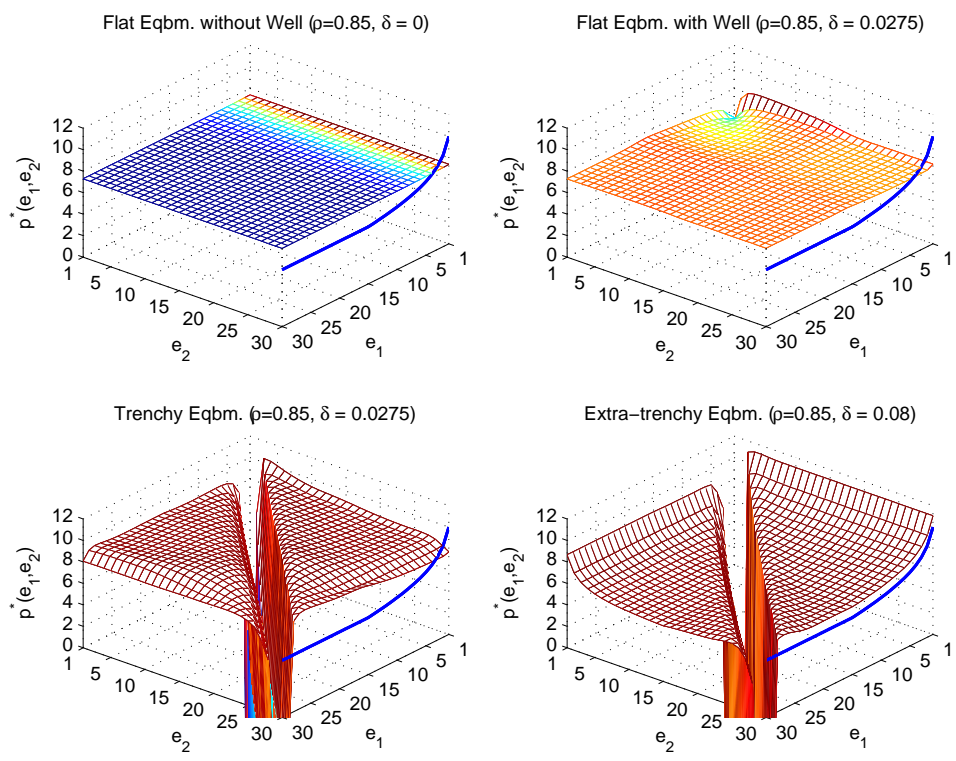


Figure A6: Discount factor of $\beta = 0.995$. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane).

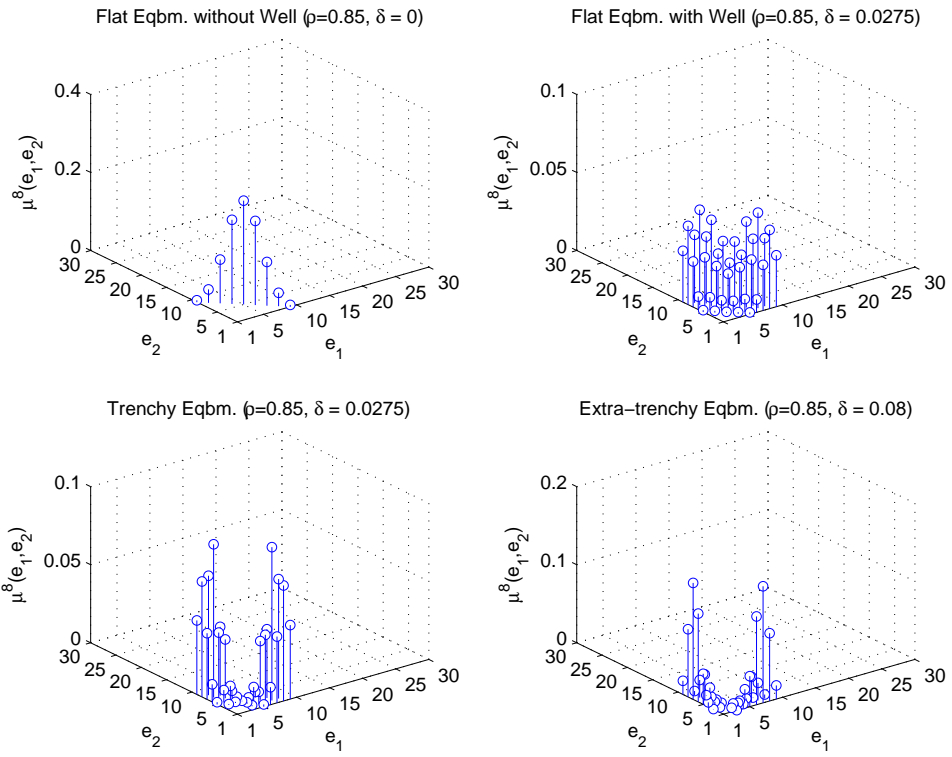


Figure A7: Discount factor of $\beta = 0.995$. Transient distribution over states in period 8 given initial state (1, 1).

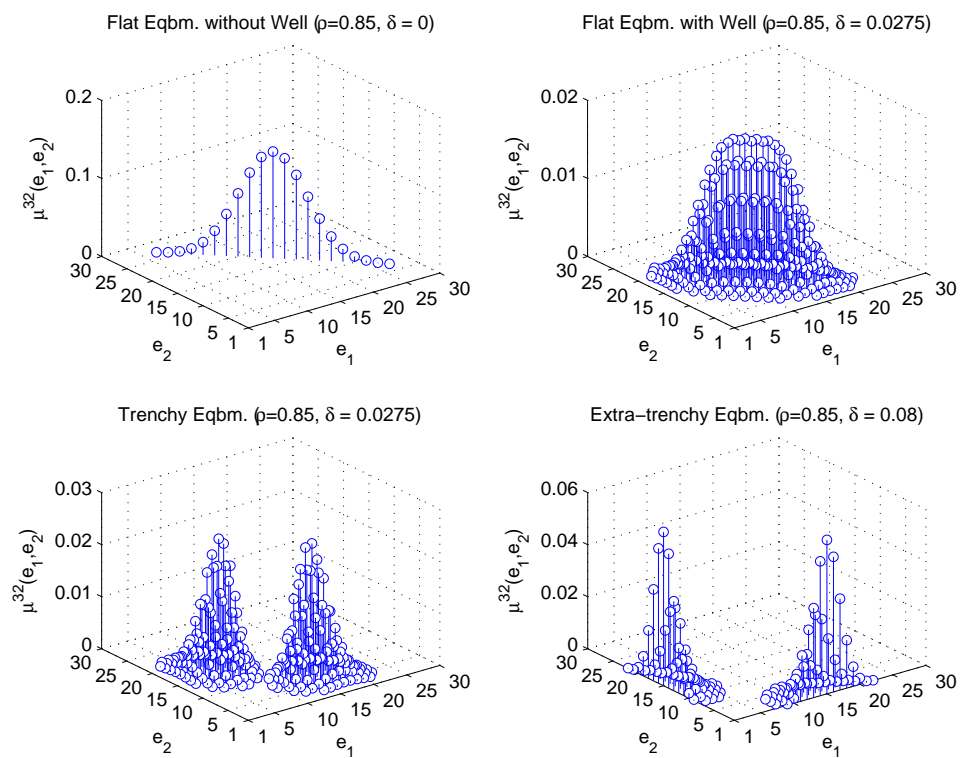


Figure A8: Discount factor of $\beta = 0.995$. Transient distribution over states in period 32 given initial state $(1, 1)$.

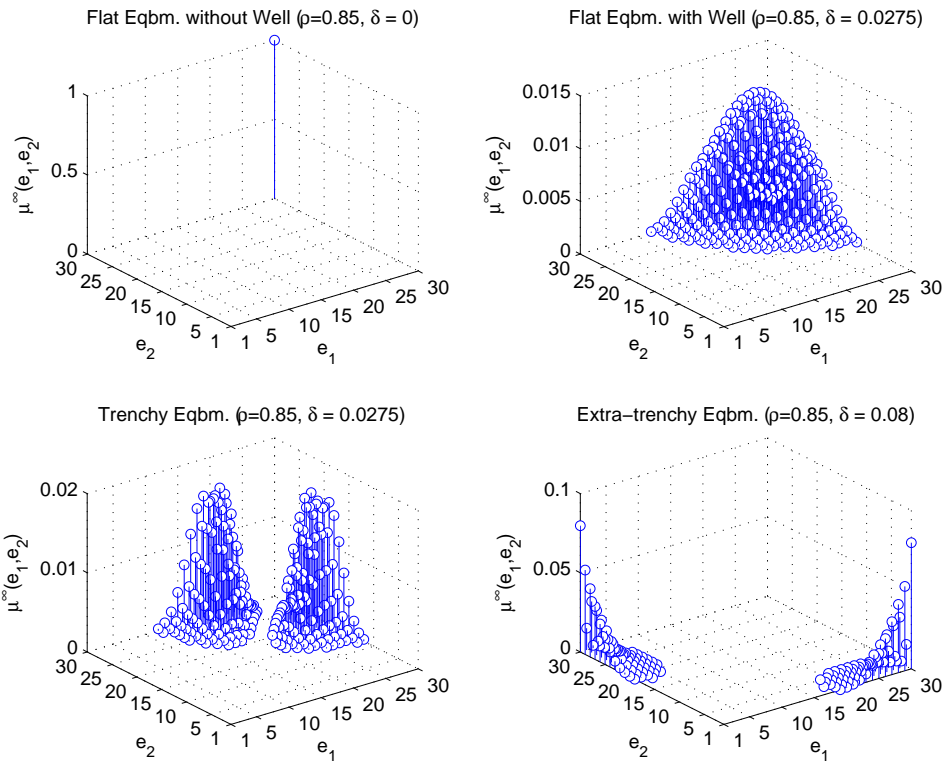


Figure A9: Discount factor of $\beta = 0.995$. Limiting distribution over states.

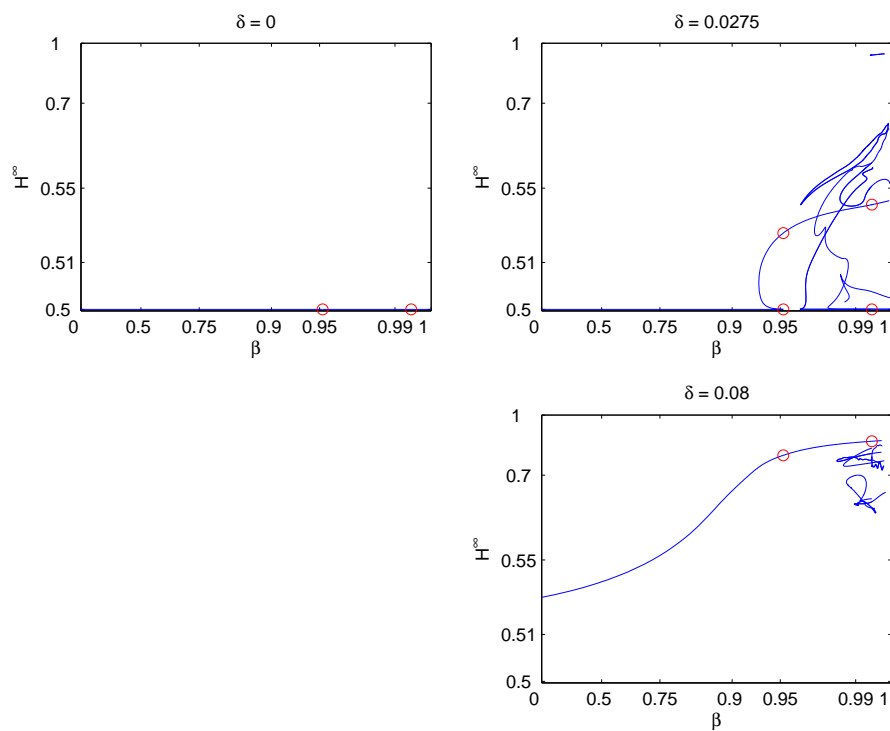


Figure A10: Discount factor. Limiting expected Herfindahl index H^∞ (solid line) and leading examples of equilibria for $\beta = \frac{1}{1.05}$ and $\beta = 0.995$ (circles).

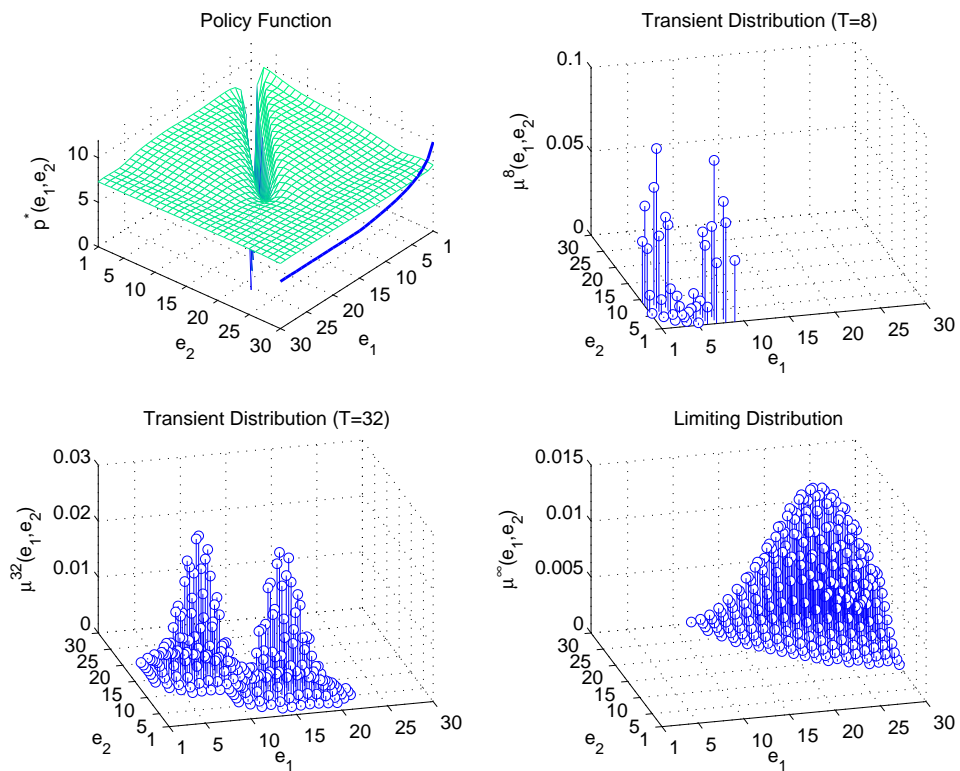


Figure A11: Discount factor. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane) (upper left panel). Transient distribution over states in period 8 and 32 given initial state (1, 1) (upper right and lower left panels). Limiting distribution over states (lower right panel). Additional equilibrium 1 ($\rho = 0.85$, $\delta = 0.0275$, $\beta = 0.995$).

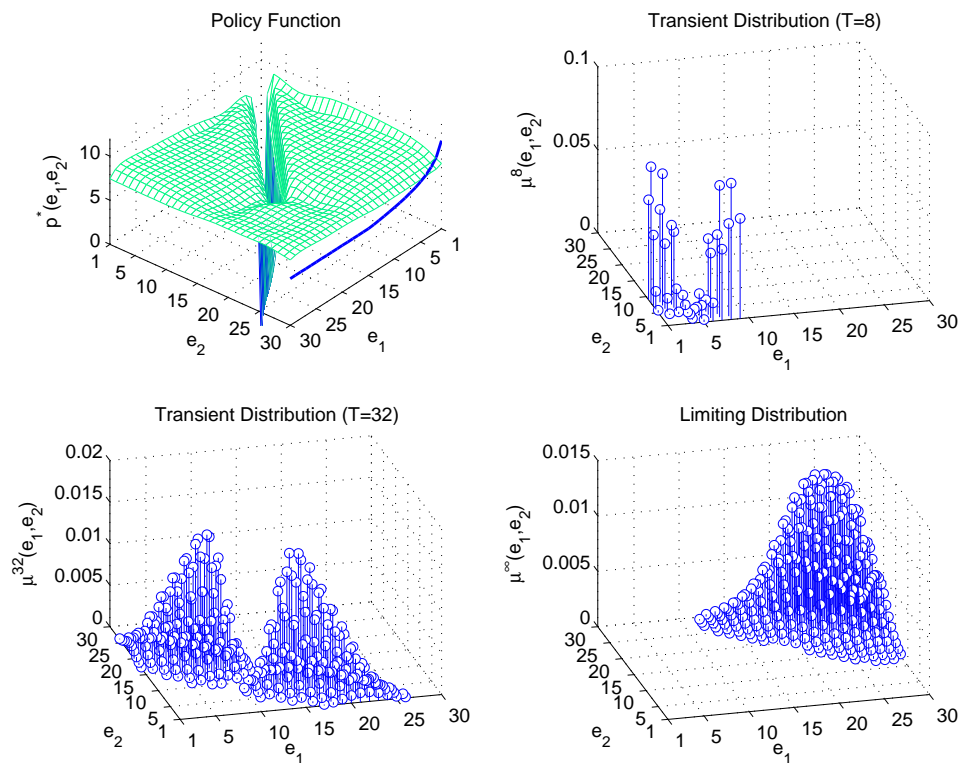


Figure A12: Discount factor. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane) (upper left panel). Transient distribution over states in period 8 and 32 given initial state $(1, 1)$ (upper right and lower left panels). Limiting distribution over states (lower right panel). Additional equilibrium 2 ($\rho = 0.85$, $\delta = 0.0275$, $\beta = 0.995$).

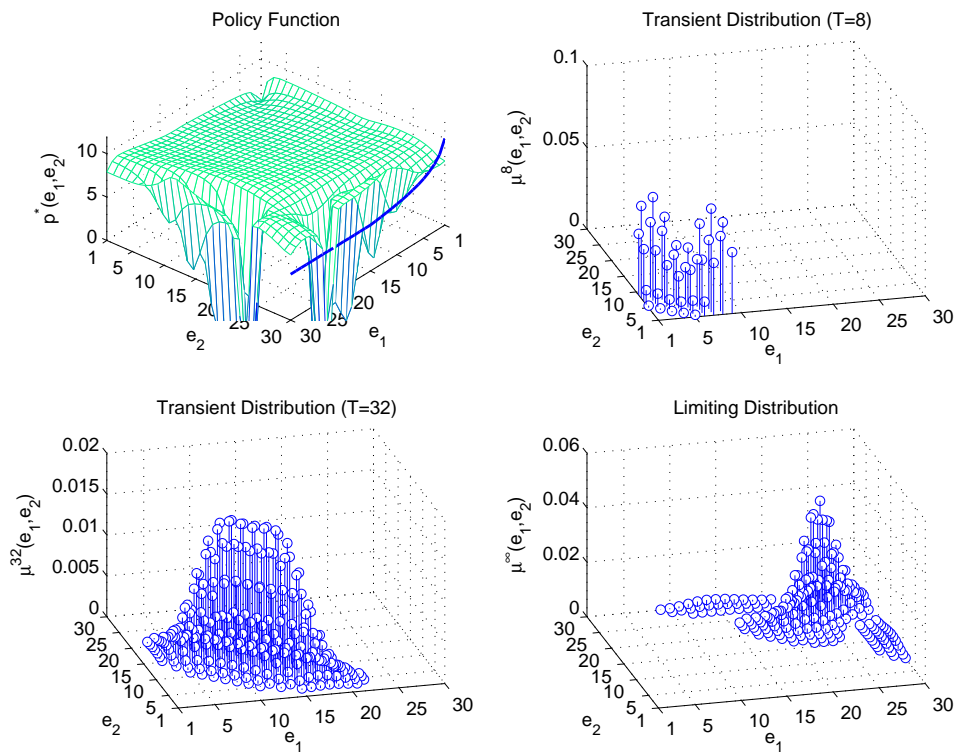


Figure A13: Discount factor. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane) (upper left panel). Transient distribution over states in period 8 and 32 given initial state $(1, 1)$ (upper right and lower left panels). Limiting distribution over states (lower right panel). Additional equilibrium 3 ($\rho = 0.85$, $\delta = 0.0275$, $\beta = 0.995$).

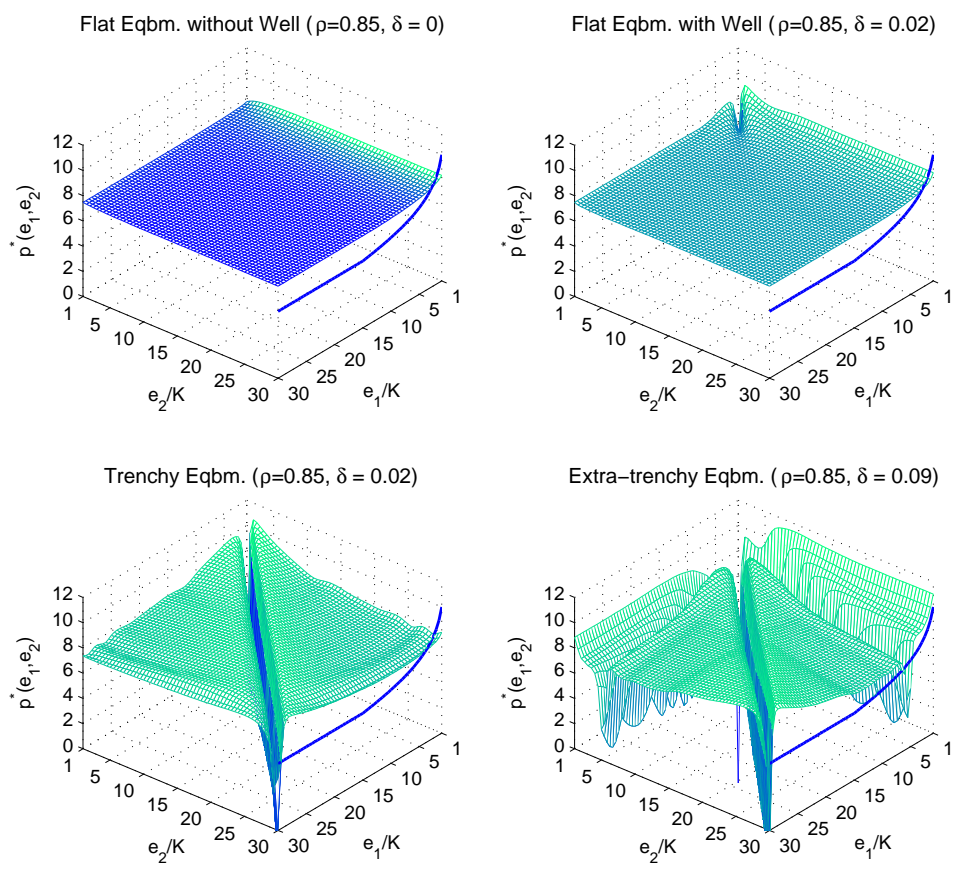


Figure A14: Frequency of sales with $K = 2$. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane).

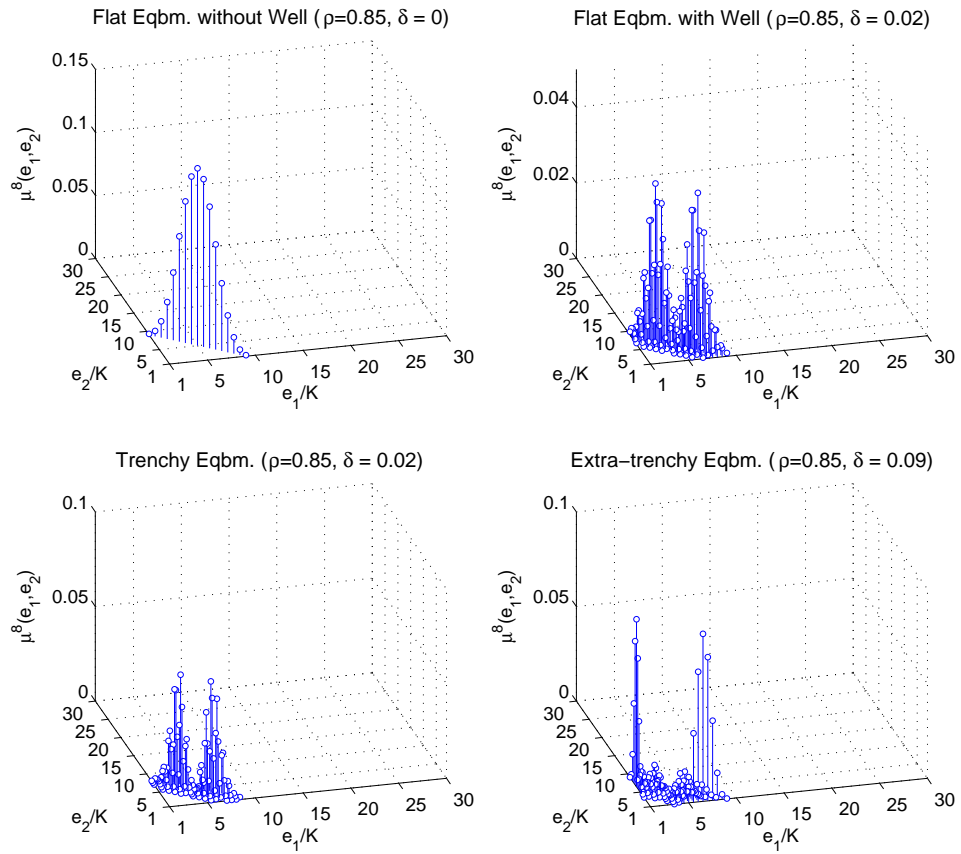


Figure A15: Frequency of sales with $K = 2$. Transient distribution over states in period 8 (subperiod 16) given initial state $(1, 1)$.

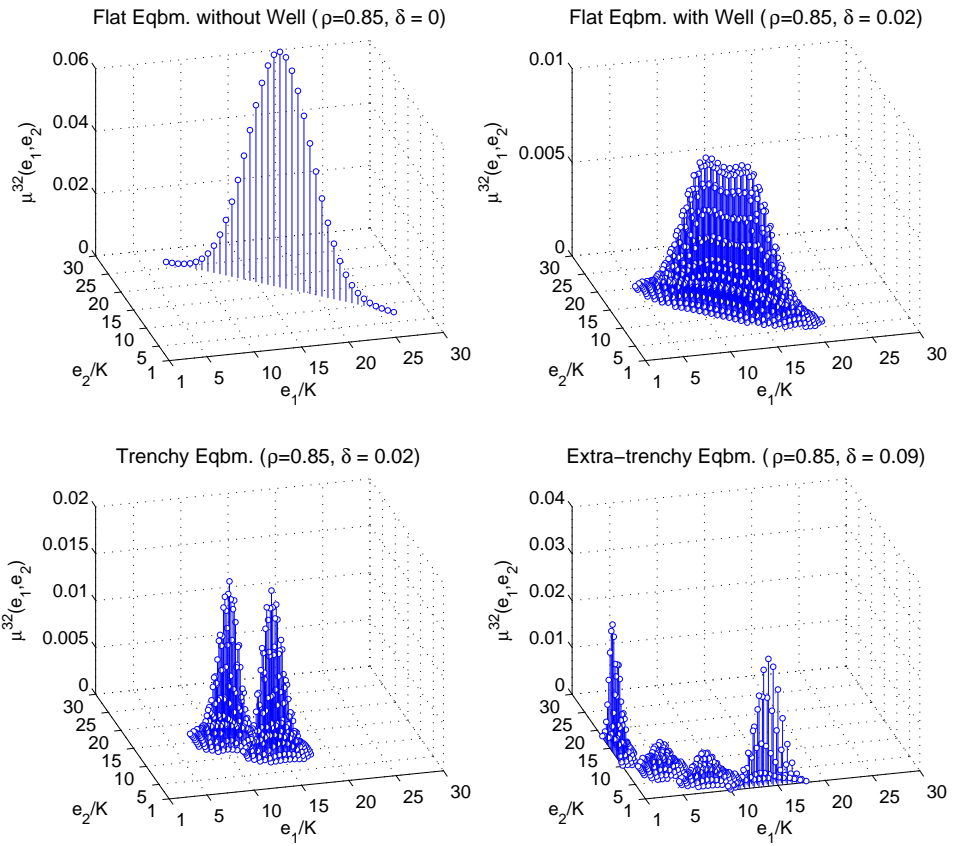


Figure A16: Frequency of sales with $K = 2$. Transient distribution over states in period 32 (subperiod 64) given initial state $(1, 1)$.

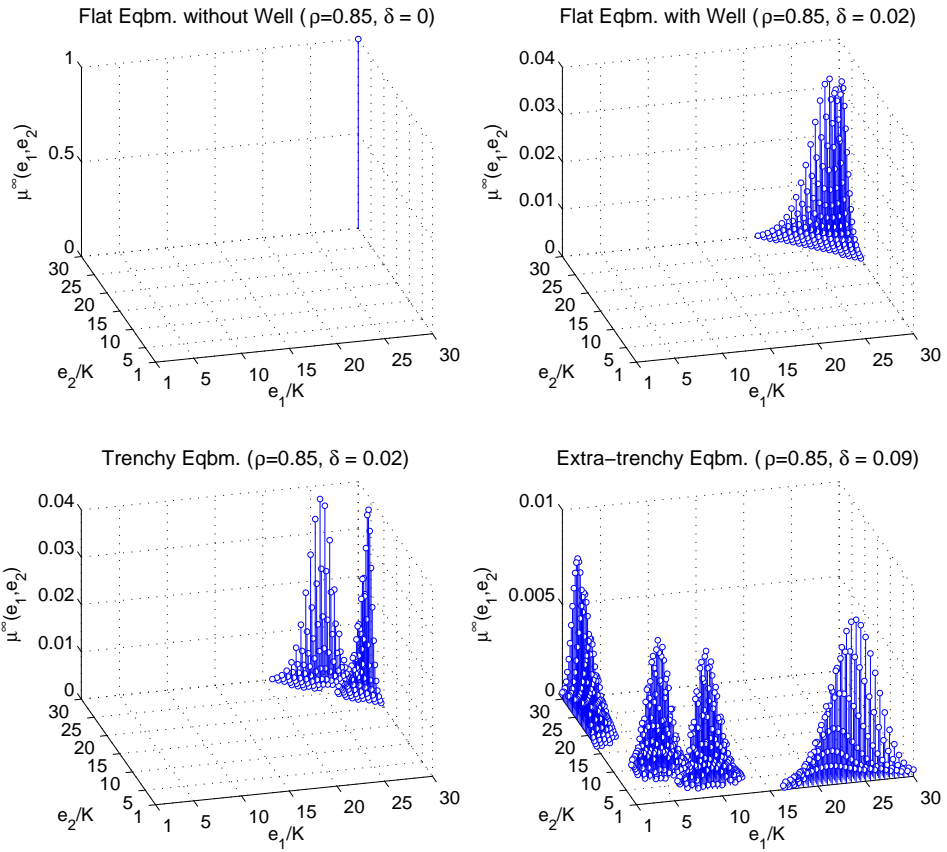


Figure A17: Frequency of sales with $K = 2$. Limiting distribution over states.

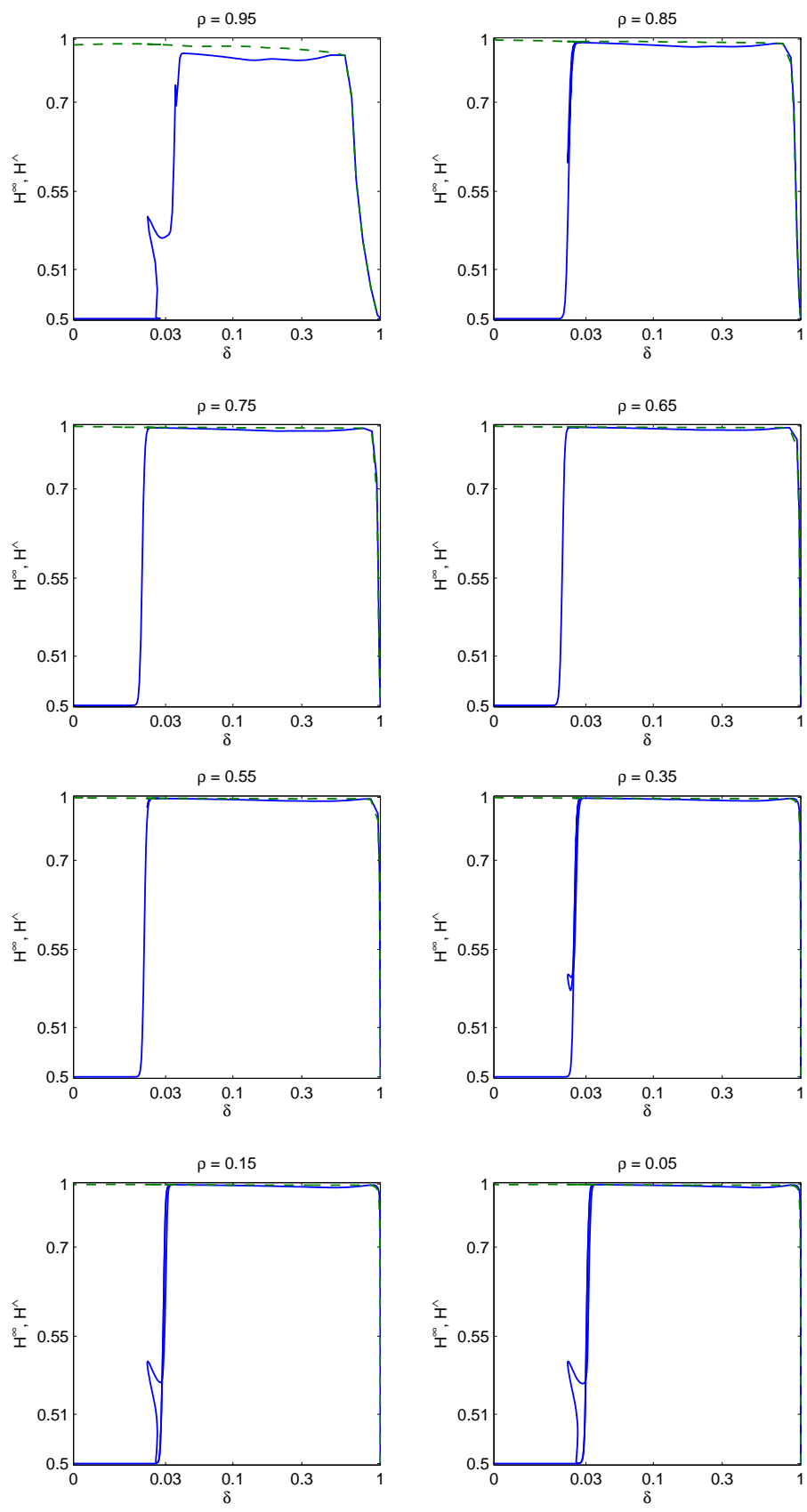


Figure A18: Product differentiation with $\sigma = 0.2$. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

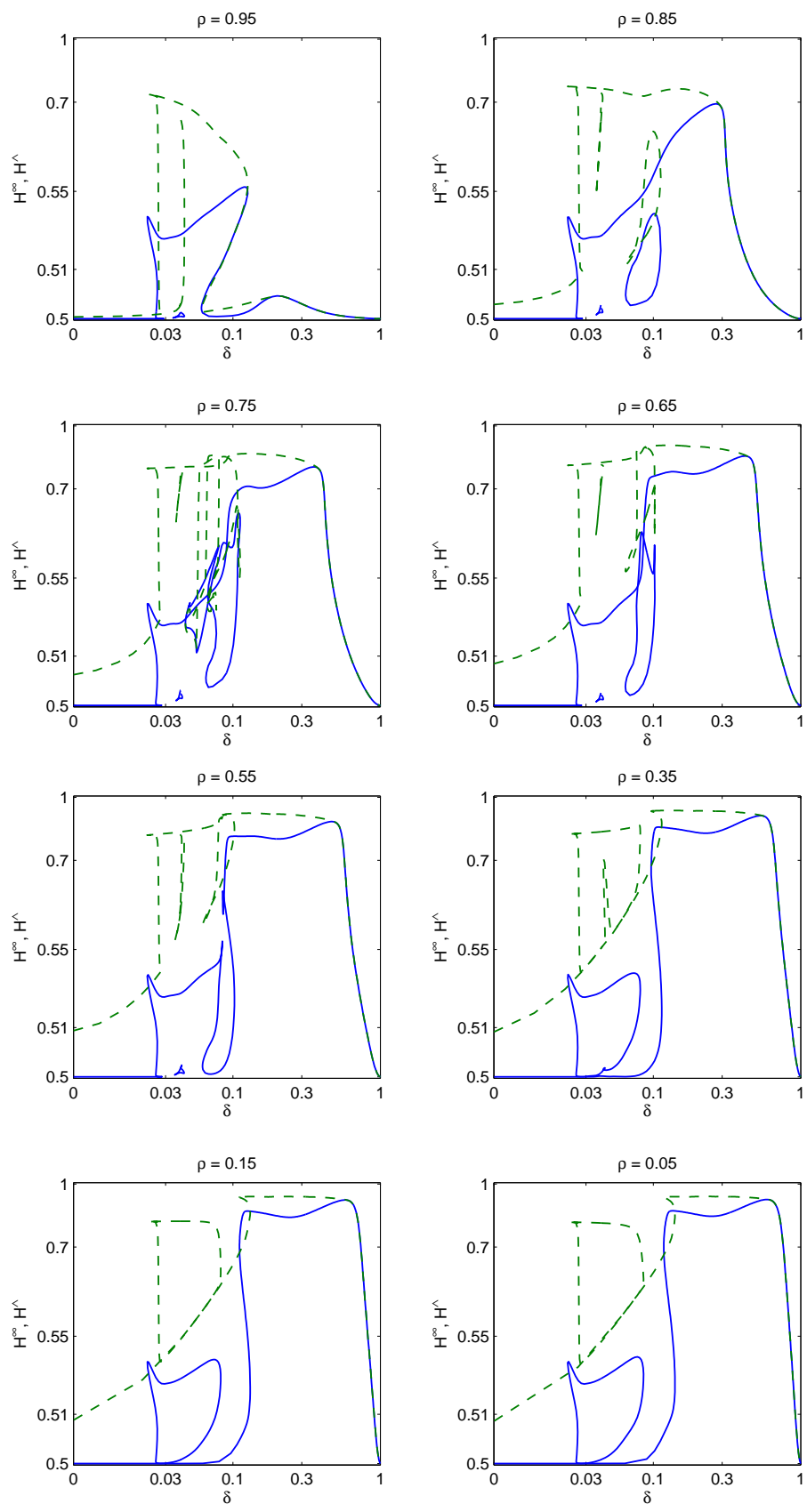


Figure A19: Product differentiation with $\sigma = 2$. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

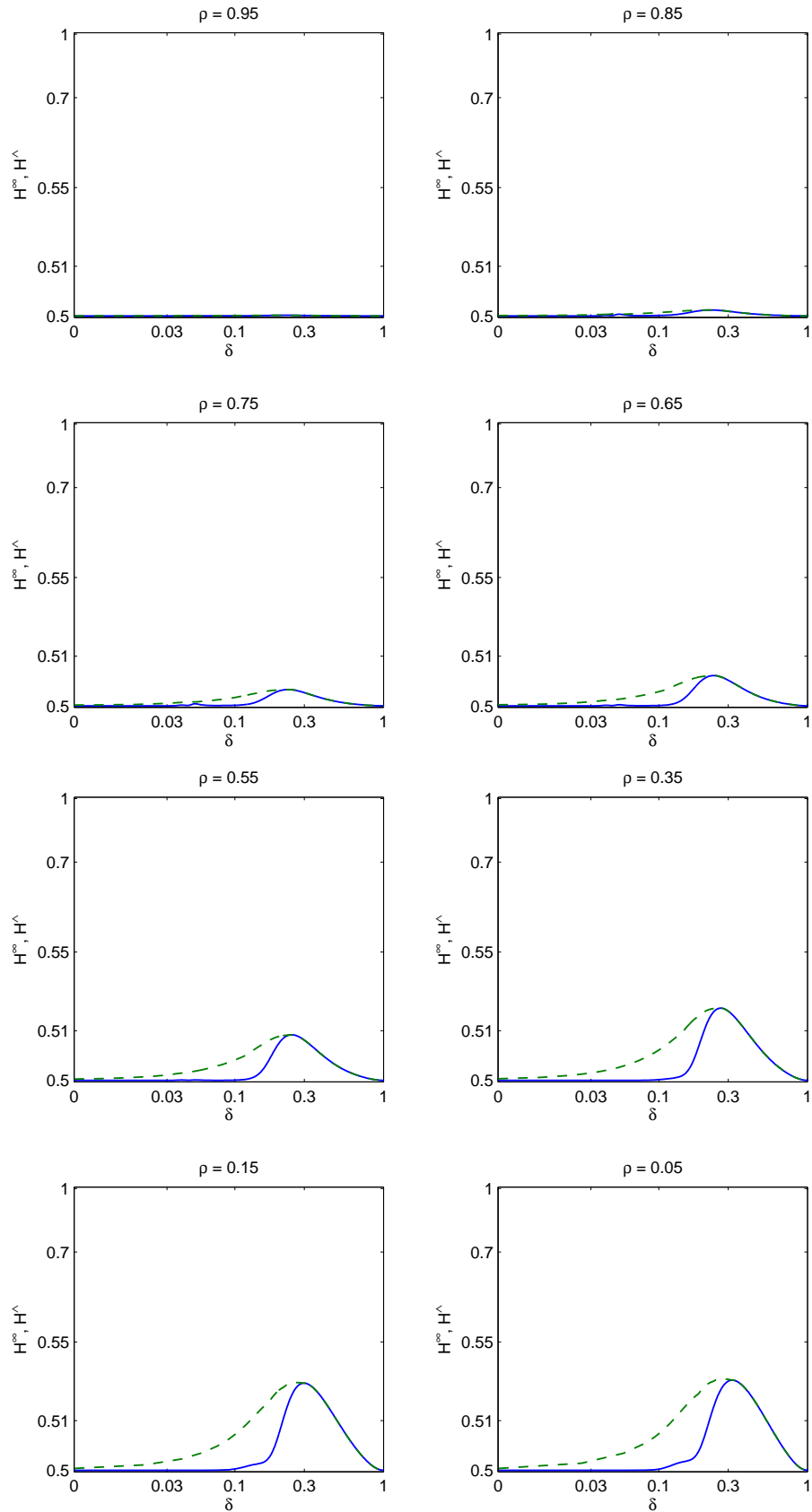


Figure A20: Product differentiation with $\sigma = 10$. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

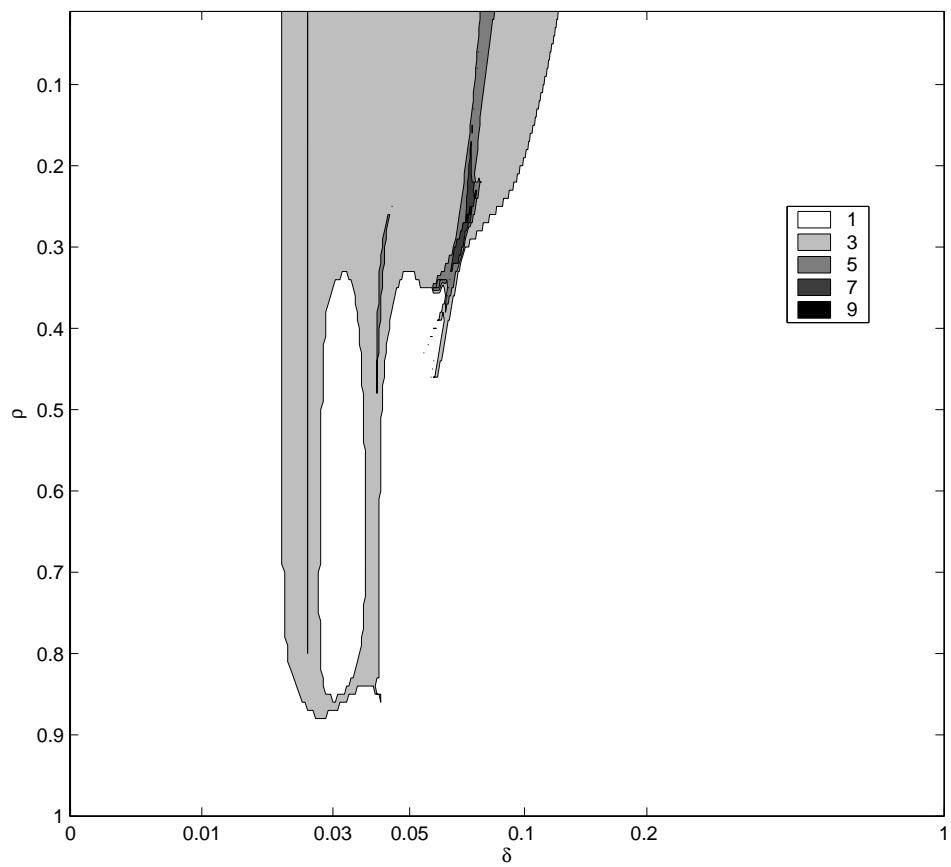


Figure A21: Outside good with $v_0 - c_0 = 0$. Number of equilibria.

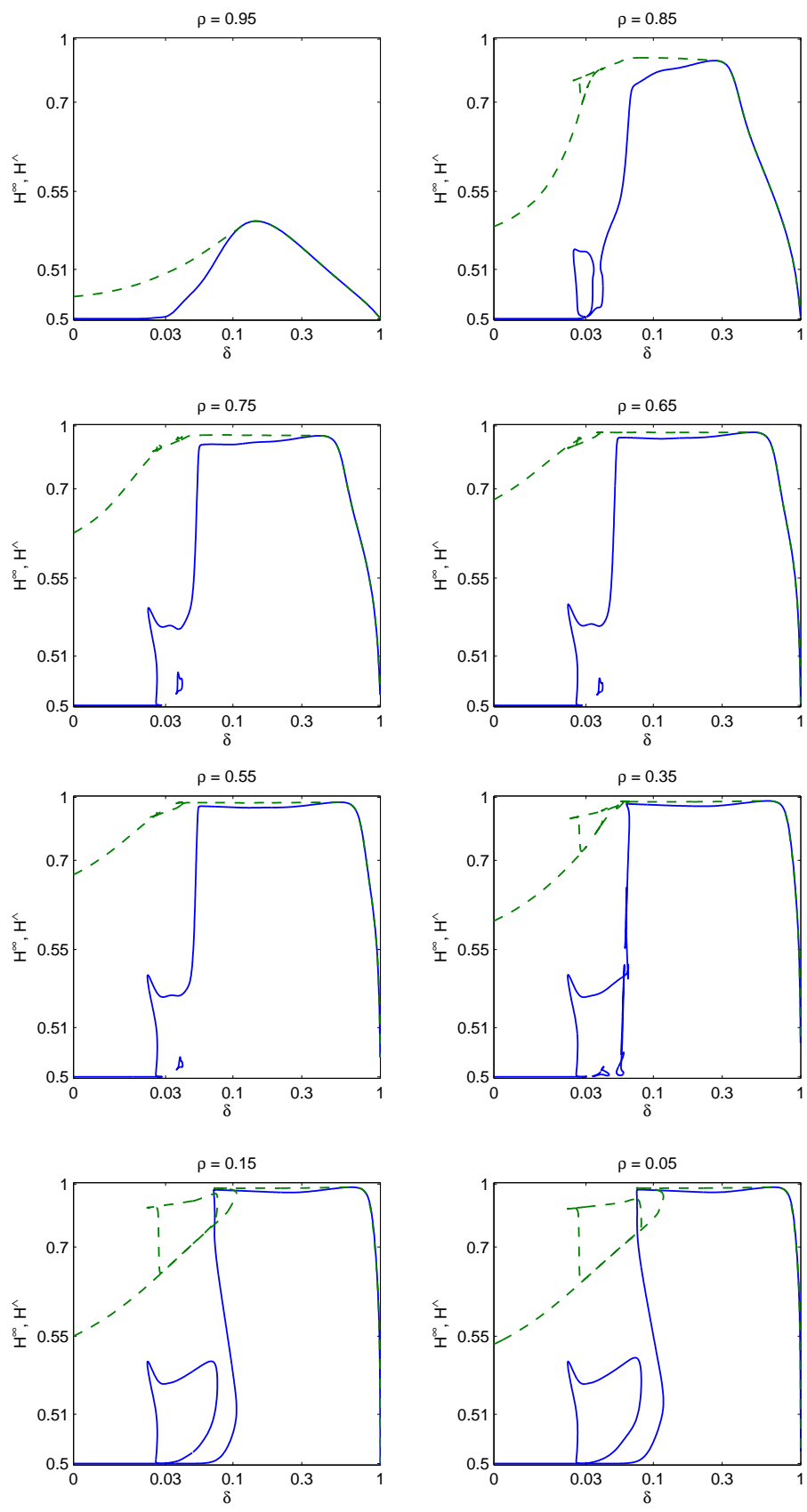


Figure A22: Outside good with $v_0 - c_0 = 0$. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

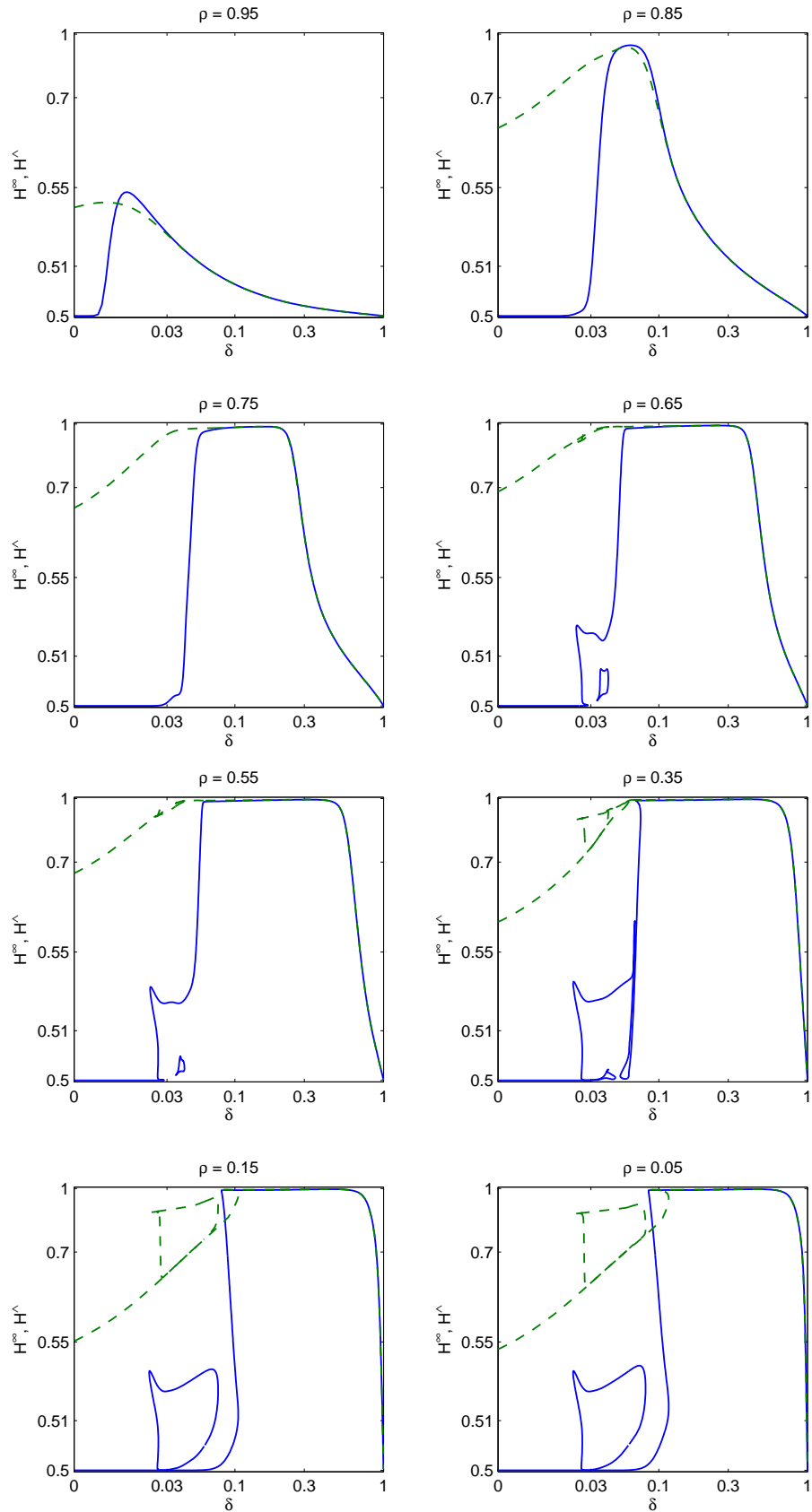


Figure A23: Outside good with $v_0 - c_0 = 3$. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

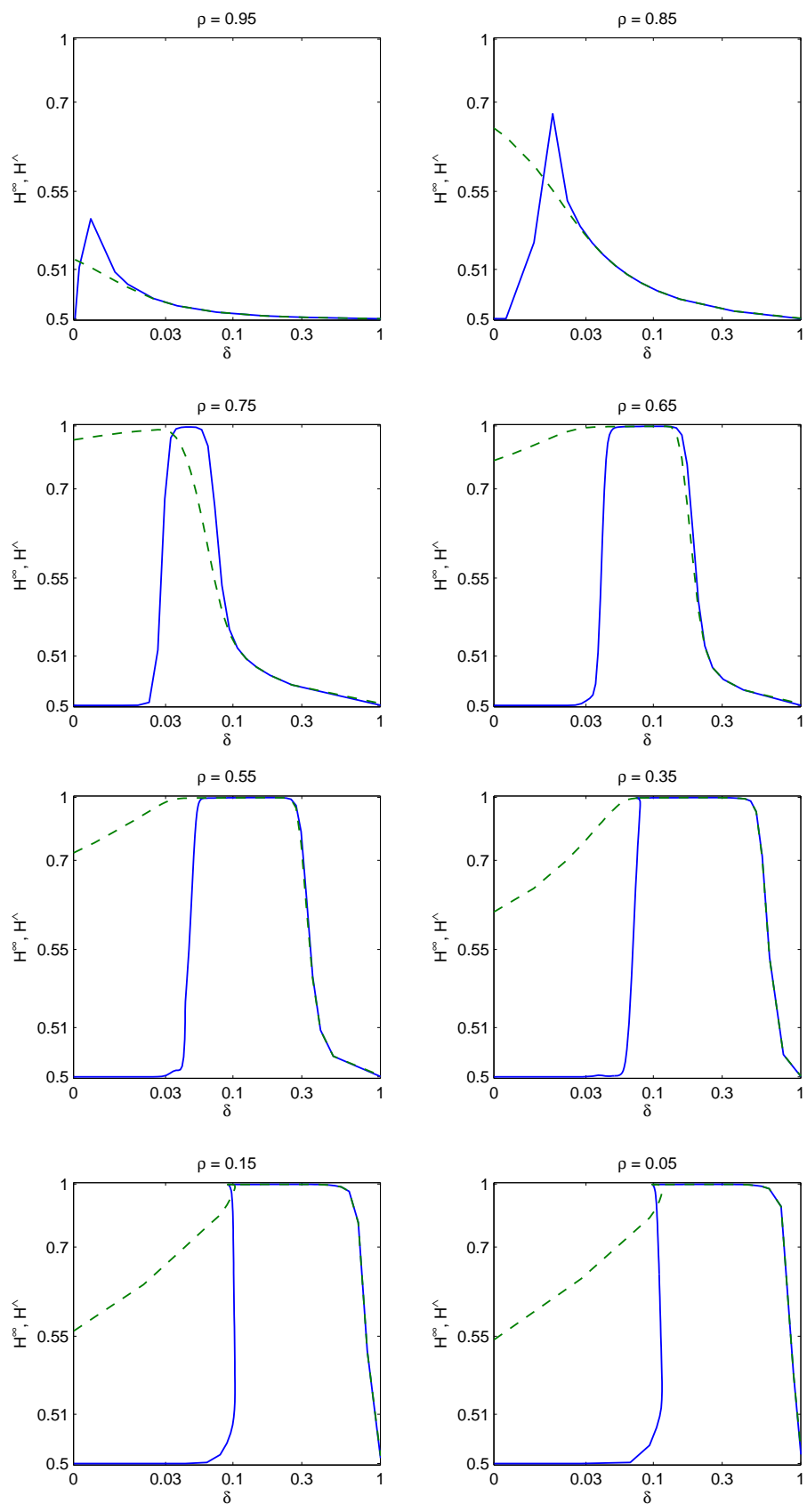


Figure A24: Outside good with $v_0 - c_0 = 5$. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

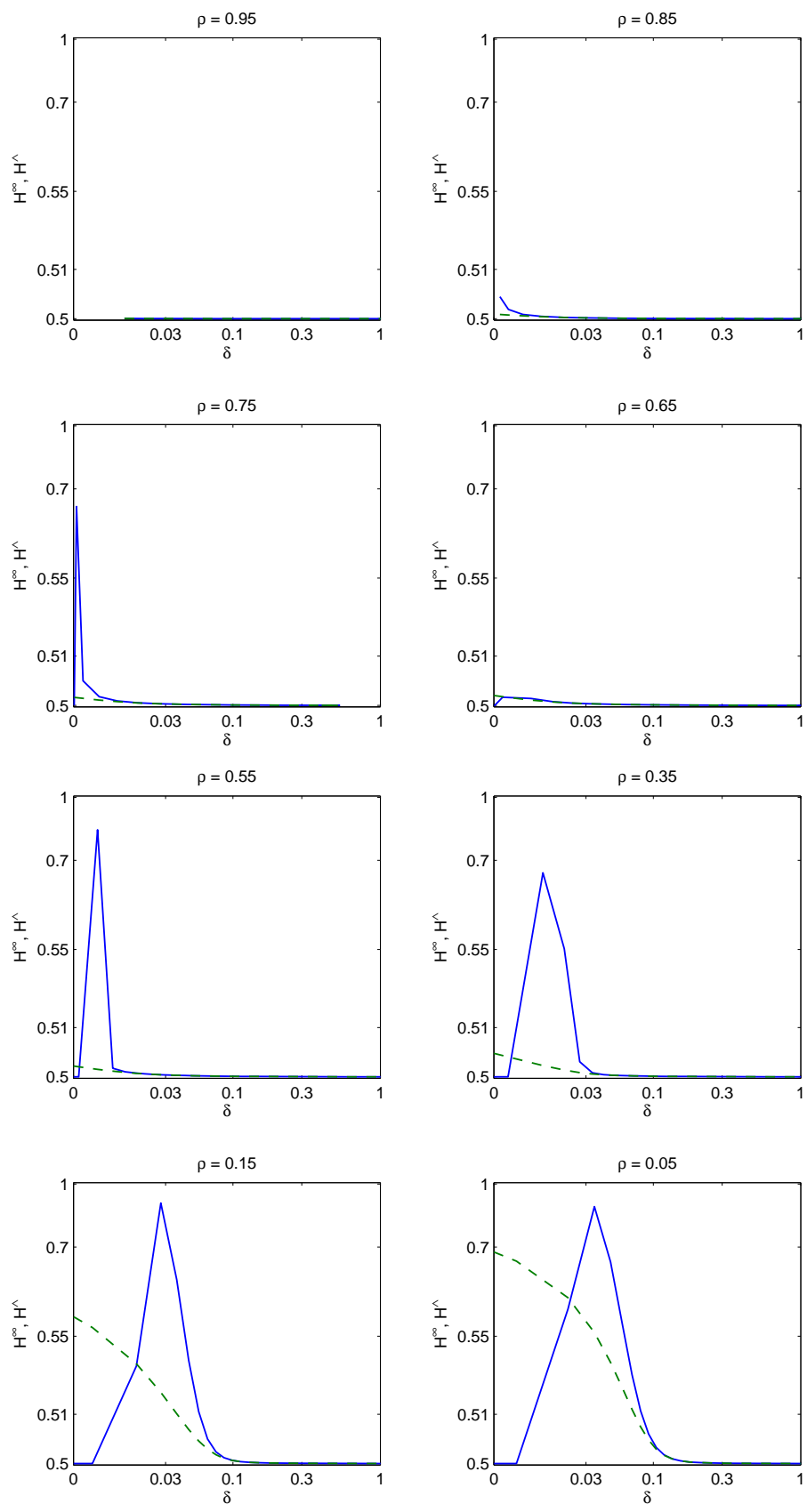


Figure A25: Outside good with $v_0 - c_0 = 10$. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

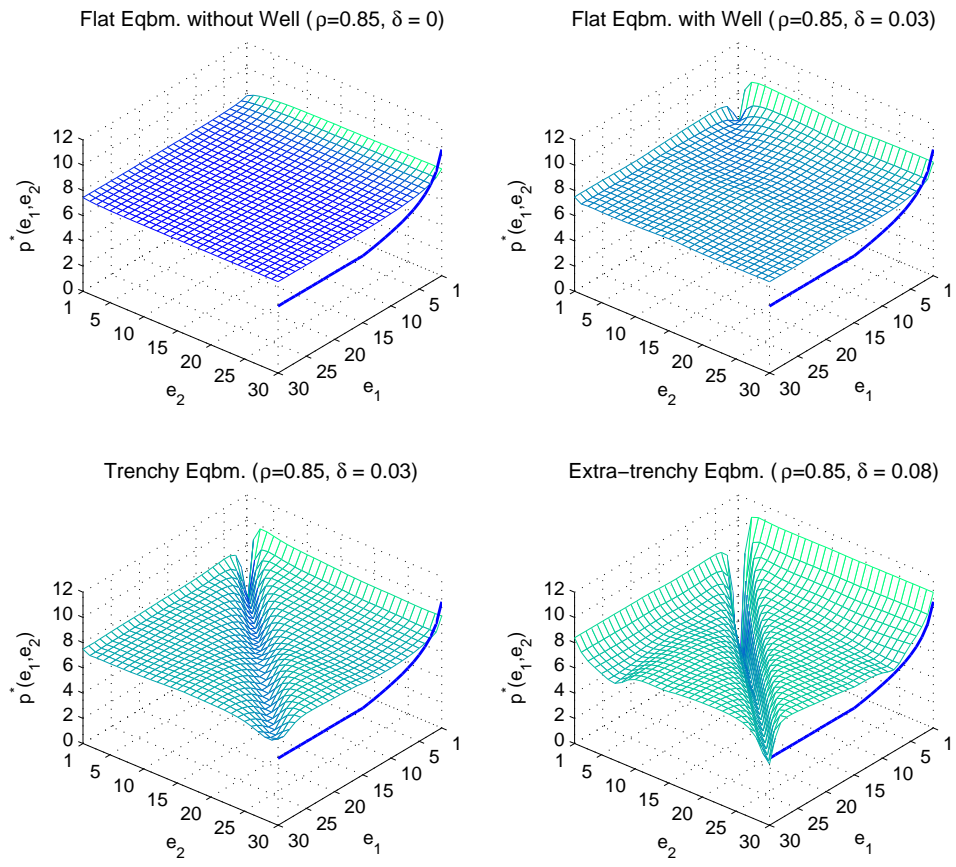


Figure A26: Outside good with $v_0 - c_0 = 0$. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane).

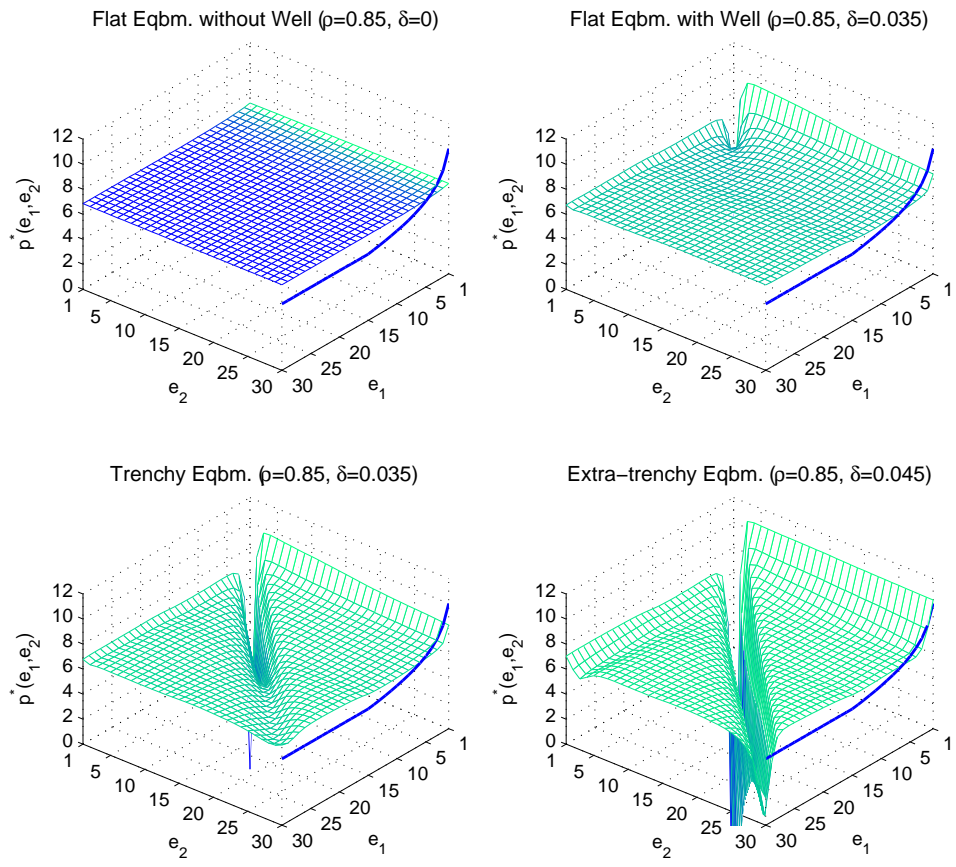


Figure A27: Outside good with $v_0 - c_0 = 3$ and $\beta = \frac{1}{1.01}$. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane).

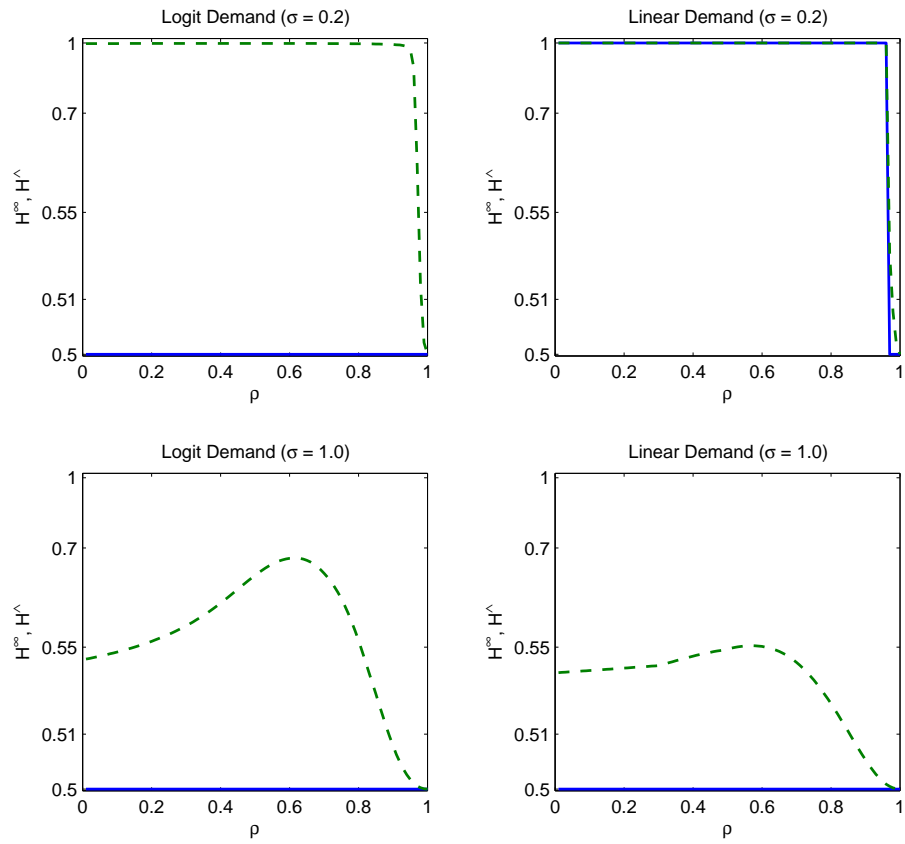


Figure A28: Choke price. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

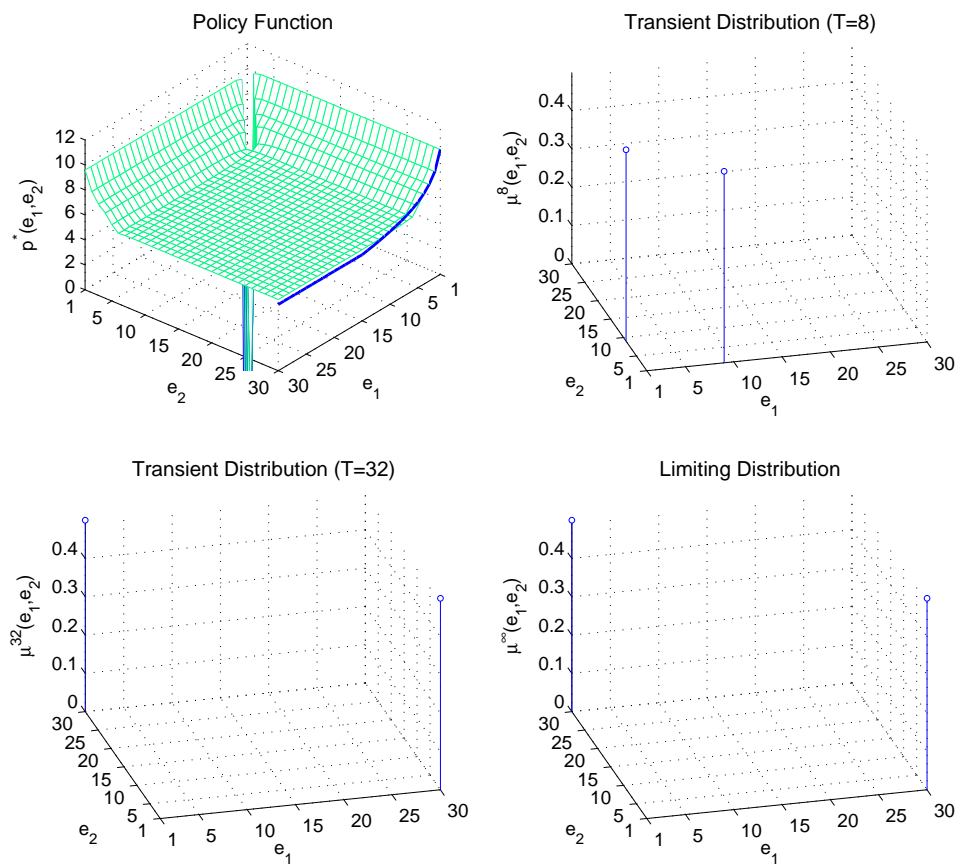


Figure A29: Choke price. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane) (upper left panel). Transient distribution over states in period 8 and 32 given initial state $(1, 1)$ (upper right and lower left panels). Limiting distribution over states (lower right panel). Linear demand ($\rho = 0.85$, $\delta = 0$, $\sigma = 0.2$).

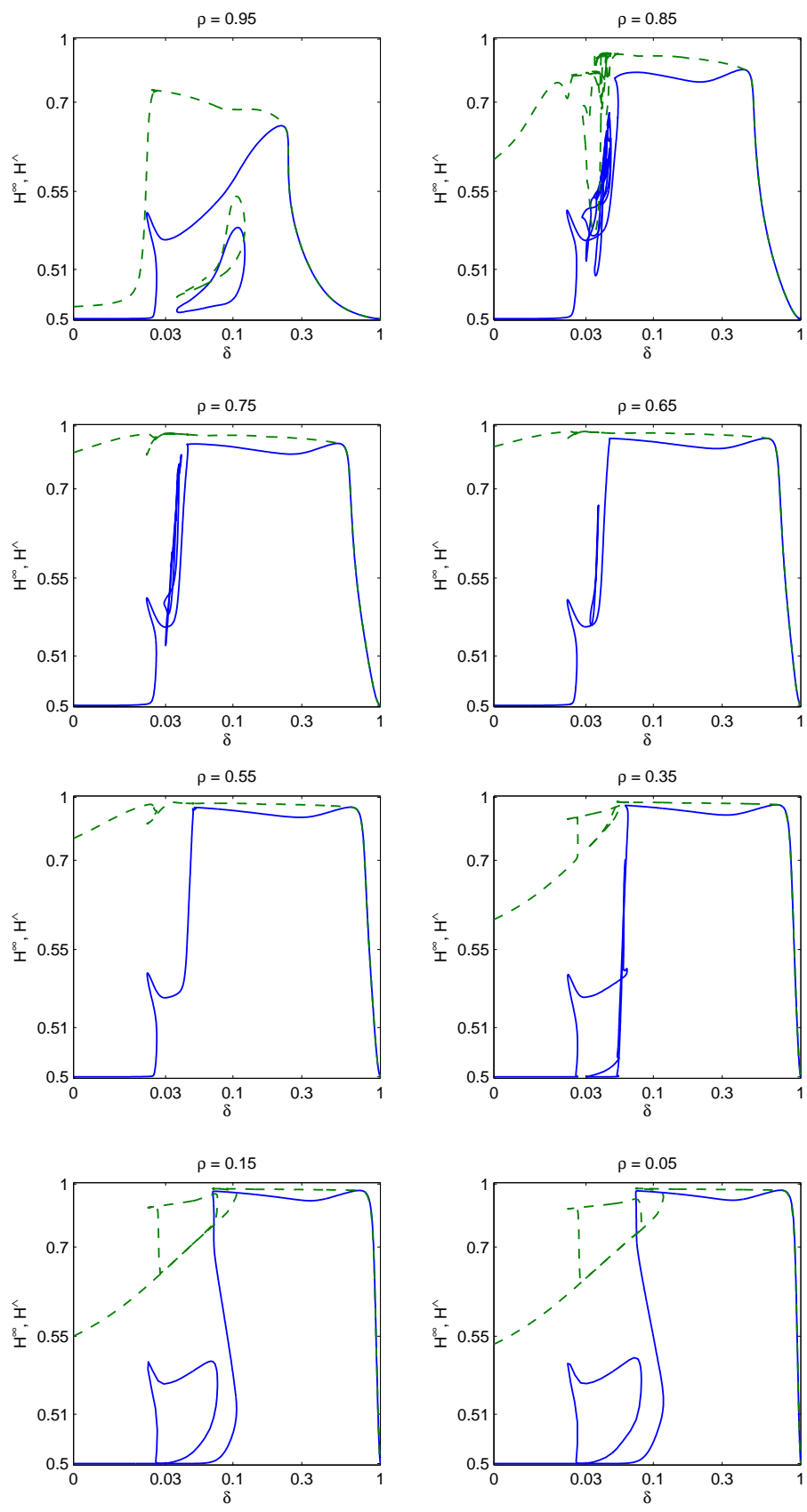


Figure A30: Bottomless learning. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

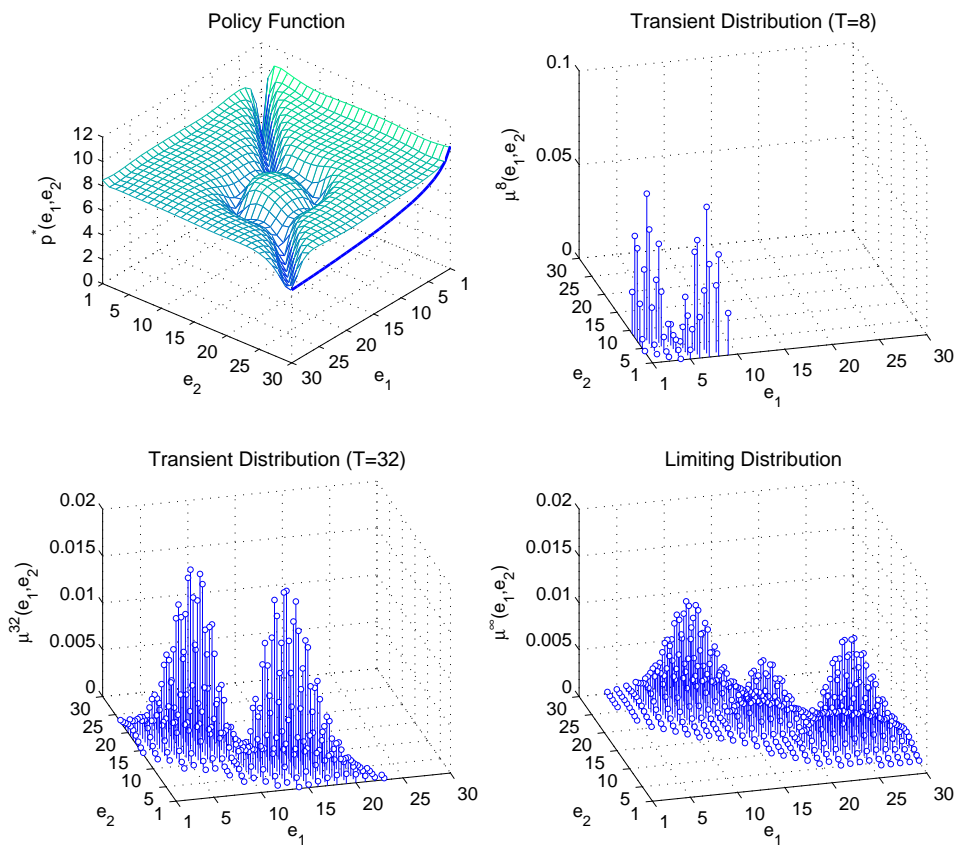


Figure A31: Bottomless learning. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane) (upper left panel). Transient distribution over states in period 8 and 32 given initial state (1, 1) (upper right and lower left panels). Limiting distribution over states (lower right panel). Plateau equilibrium 1 ($\rho = 0.9$, $\delta = 0.04$).

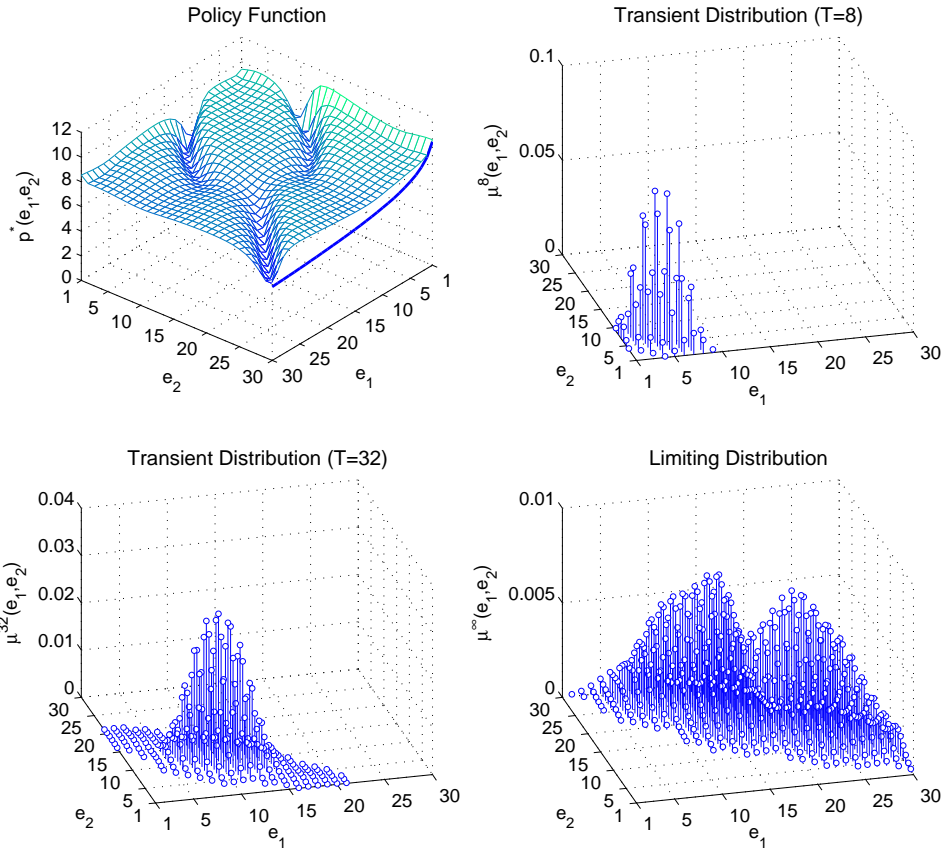


Figure A32: Bottomless learning. Policy function $p^*(e_1, e_2)$. Marginal cost $c(e_1)$ (solid line in $e_2 = 30$ -plane) (upper left panel). Transient distribution over states in period 8 and 32 given initial state $(1, 1)$ (upper right and lower left panels). Limiting distribution over states (lower right panel). Plateau equilibrium 2 ($\rho = 0.9$, $\delta = 0.04$).

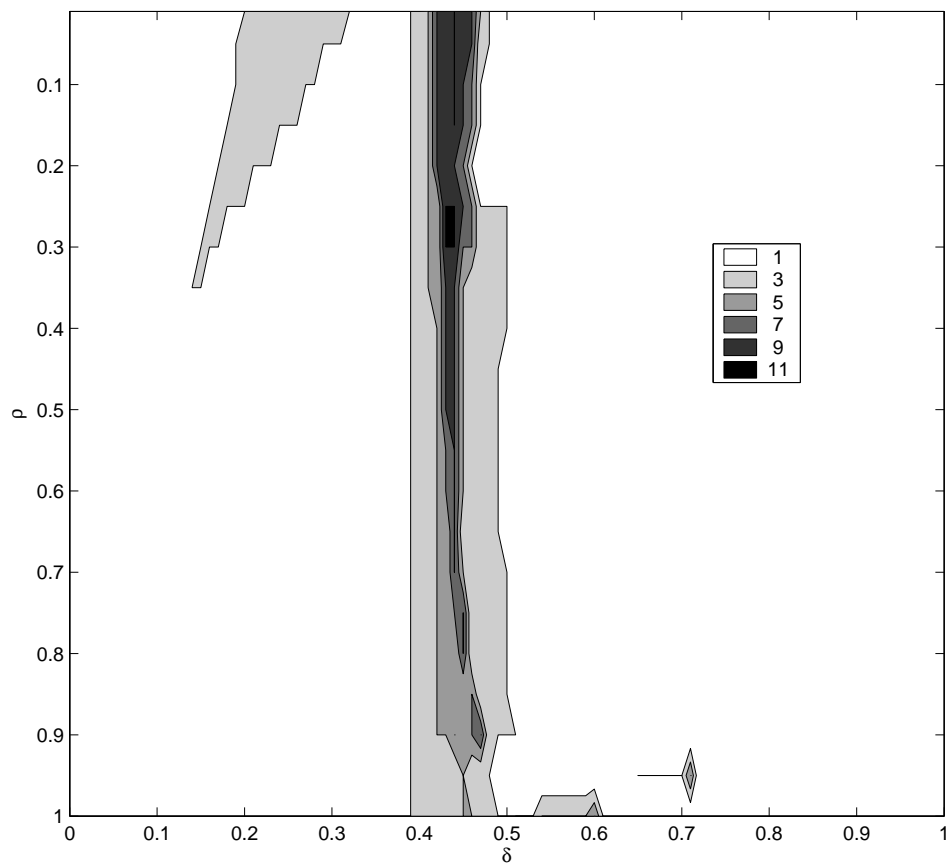


Figure A33: Constant forgetting. Number of equilibria.

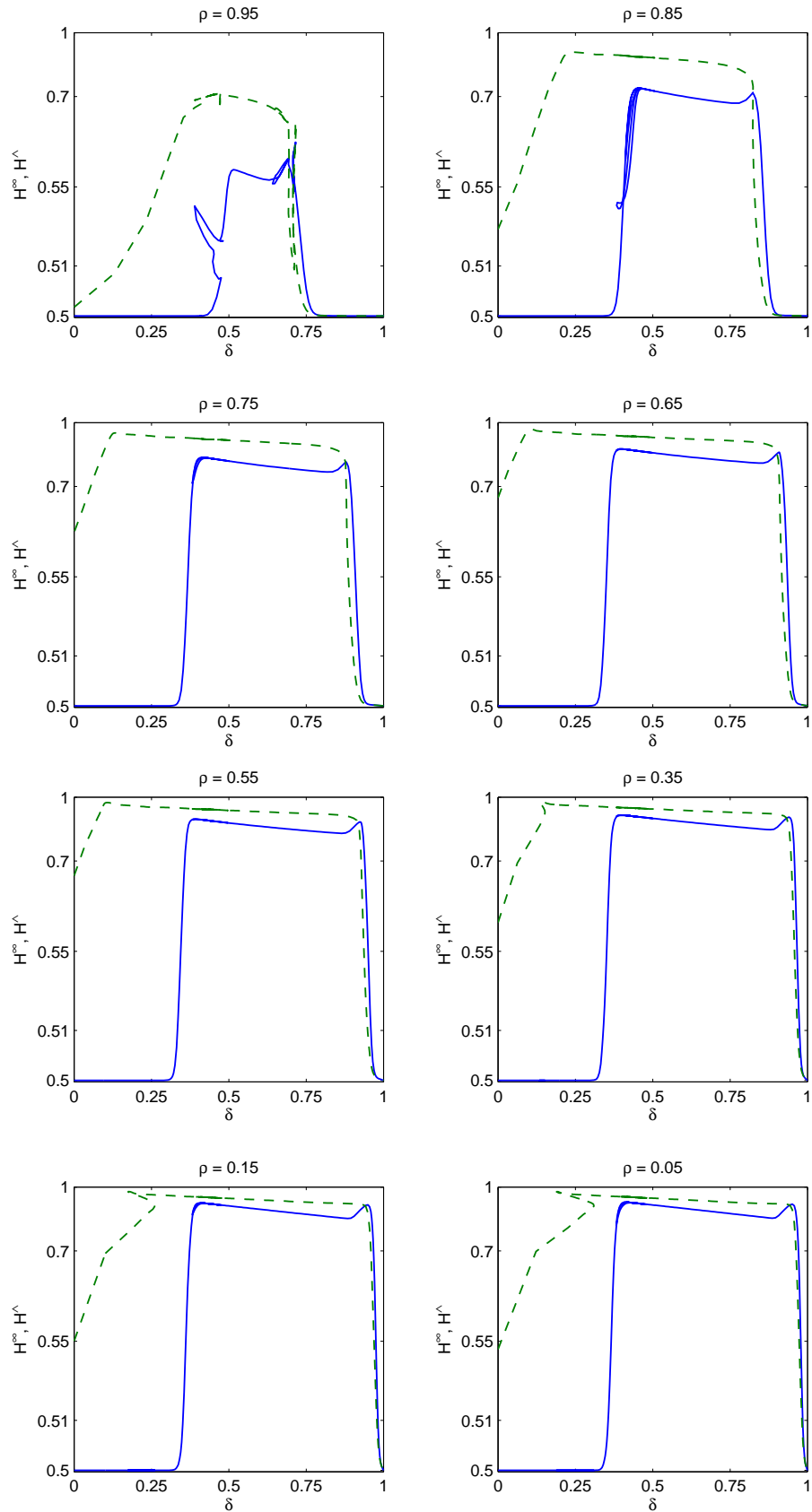


Figure A34: Constant forgetting. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).

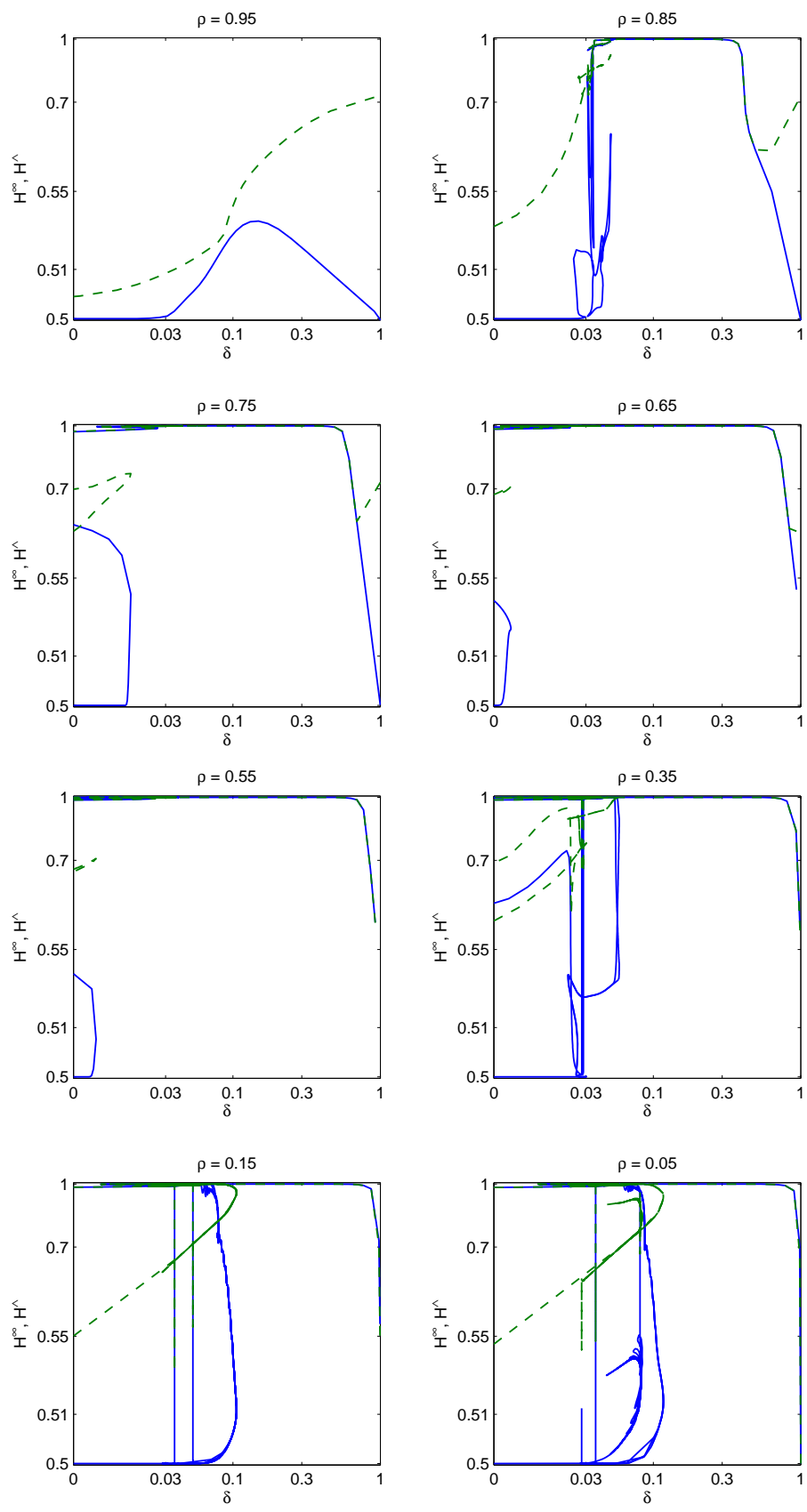


Figure A35: Entry and exit. Limiting expected Herfindahl index H^∞ (solid line) and maximum expected Herfindahl index H^\wedge (dashed line).