

OPTIMAL EXPERIMENTATION IN A CHANGING ENVIRONMENT*

Godfrey Keller[†] Sven Rady[‡]

January 1995

Revised: October 1995

This version: July 1997

Abstract

This paper studies optimal experimentation by a monopolist who faces an unknown demand curve subject to random changes, and who maximises profits over an infinite horizon in continuous time. We show that there are two qualitatively very different regimes, determined by the discount rate and the intensities of demand curve switching, and the dependence of the optimal policy on these parameters is discontinuous. One regime is characterised by extreme experimentation and good tracking of the prevailing demand curve, the other by moderate experimentation and poor tracking. Moreover, in the latter regime the agent eventually becomes ‘trapped’ into taking actions in a strict subset of the feasible set.

Keywords: Bayesian Learning, Optimal Control, Monopoly Experimentation

JEL classification: D83, C61, D42

*Research Paper No. 1443, Graduate School of Business, Stanford University. We would like to thank Patrick Bolton and Lucien Foldes for their encouragement and guidance, and Margaret Bray, Phil Dybvig, Michael Harrison, John Hardman Moore, John Roberts, Andrew Stuart and Dimitri Vayanos for helpful discussions and comments. Previous versions of this paper were presented in seminars at Columbia University, ECARE (Brussels), Illinois-Chicago, LSE, NYU, Stanford, Tel Aviv, Toulouse, Washington University St. Louis, Wharton School, Wisconsin-Madison, and at conferences in Barcelona, Canterbury, Prague and Tokyo; we would like to thank our audiences for their comments. We both are grateful for the assistance of the Financial Markets Group at LSE and the financial support of the ESRC. The first author acknowledges the financial support of the European Commission through its Human Capital & Mobility Programme. Part of this paper was written while the second author benefited from a travel grant by STICERD at LSE. Finally, we would both like to thank the Studienzentrum Gerzensee where the final revisions to this version were made.

[†]ECARE, Université Libre de Bruxelles, 39 Avenue F.D. Roosevelt, B-1050 Bruxelles, Belgium; & Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE; e-mail: R.G.Keller@lse.ac.uk.

[‡]Graduate School of Business, Stanford University, Stanford, CA 94305-5015; tel.: (+1) 415 723 5512; e-mail: Rady_Sven@gsb.stanford.edu.

Introduction

In this paper, we study a situation in which an economic agent can learn by experimentation, that is, by deviating from the action which maximises current pay-off in order to generate information which will increase future pay-offs. Specifically, the agent is facing an unknown and changing reward function. After taking an action, he observes a noisy signal of the current state, leading to a revision of his beliefs. The agent's objective is to maximise the expected discounted pay-off over an infinite horizon. When choosing an action, he therefore has to weigh the opportunity cost of experimentation against its long-term informational benefits.

We are interested in a number of issues. How does the agent's optimal action differ from what is myopically optimal? Is this difference large or small (in a sense to be made precise)? How well or poorly does the agent track the prevailing state? Is there an analogue to the 'incomplete learning' results which are common in the literature on experimentation in an unchanging environment, mentioned below?

Our main result is the identification of two qualitatively very different regimes. One regime is characterised by large deviations from myopic behavior and by good tracking of the prevailing demand curve, the other by small deviations and poor tracking. Moreover, in the latter regime the agent eventually becomes 'trapped' in a subset of the action space and is bounded away from the myopic optimum for one of the states.

Two approaches to the problem of learning by experimentation in a single-agent setting have emerged in the literature. One approach is to restrict the analysis to a two-period framework, and then to determine in which way the ability to gather information in the first period (which will be useful in the second period) affects the agent's behaviour; this is usually measured as a deviation from the agent's first period myopic action. Examples of this approach can be found in Mirman, Samuelson and Urbano (1993) and other papers referenced in their introduction.

The second approach involves formulating an infinite-horizon model, in which case it is natural to look at the limiting behaviour of the agent. The first such model in the economics literature is due to Rothschild (1974), and has subsequently been extended in a number of different directions; see, for example, McLennan (1984), Easley and Kiefer (1988), Kiefer (1989a), and Aghion, Bolton, Harris and Jullien (1991). A common result of these models is that the agent's beliefs converge, in which case experimentation will cease and no further information will be gathered – in the limit, the agent will learn everything that is *worth* knowing. The question then arises as to whether the beliefs converge to a one-point distribution at the true parameter value. Typically the answer is that with positive probability they do not, i.e. there is incomplete learning.

We take the second approach as our starting point. However, whereas the above papers assume there to be an unknown but fixed reward function, we follow Kiefer (1989b) and Balvers and Cosimano (1990) in allowing this reward function to change randomly. This adds more realism in that new data continues to be pertinent, so beliefs continue to evolve, and the agent is not doomed to take the same action for evermore.

The agent in our model is a monopolist facing an unknown and changing demand function. The time parameter is continuous. There are two states, each characterised

by a linear demand curve, and transitions between these states are governed by a Markov process. The monopolist knows the slope and intercept of each demand curve and the transition probabilities of the Markov process. However, he does not know which demand curve he faces. At each moment of time, he chooses a quantity from a given interval of feasible quantities, and observes a price which is the price for the current demand curve plus noise. Given this noisy signal of the underlying demand curve, the monopolist then updates his belief about the current state in a Bayesian fashion. We call the quantity corresponding to the intersection of the two demand curves the confounding quantity because it leads to the same expected price in either state and thus to a completely uninformative price signal.

We naturally expect experimentation to be in the direction of widening spreads between the two demand curves, thus making price observations more informative. If the confounding quantity lies outside the range of the myopic policy function (a closed interval spanned by the quantities that maximise instantaneous profit for each of the two demand curves), then this direction is unambiguous. The monopolist deviates from the myopic action by moving away from the intersection – experimentation is always in one direction, meaning that the monopolist either produces always more than the myopic quantity, or always less. Experimentation is moderate in the sense that the optimal quantity always lies inside the range of the myopic policy function, and the optimal policy function is continuous and monotonic. Moreover, experimentation is qualitatively the same for all parameter values.

On the other hand, if the confounding quantity lies inside the range of the myopic policy function, then there is a unique belief – the confounding belief – at which a myopic agent would choose the confounding quantity, and the direction of widening spreads is no longer unambiguous. We see experimentation in both directions: for some beliefs, the monopolist produces more than the myopic quantity, for others, less. But we also find that there are two very different *experimentation regimes*. For a given level of noise, when the discount rate and the intensity of state switching are both low (in a fairly precise sense) then experimentation is *extreme* in that, for beliefs in an interval encompassing the confounding belief, the optimal action is to choose a quantity at the boundary of the interval of feasible quantities, and the optimal quantity (as a function of the belief) exhibits a jump from one boundary to the other. In this regime, the true state appears to be tracked fairly well.

However, for the same level of noise, when either the discount rate or the switching intensity is high (in a similarly precise sense) then experimentation is *moderate*, and the monopolist optimally chooses the confounding quantity at the confounding belief. In this regime, the monopolist eventually becomes trapped into choosing quantities on just one side of the confounding quantity (the side which contains the myopic action corresponding to the long-run average state) and the true state is tracked rather poorly. This trap generalises the incomplete learning result of the previous literature to a changing environment. It has the following intuition. When the monopolist does not value information very highly, he chooses the confounding quantity at the confounding belief, and the resulting price contains no information. His updating is then driven exclusively by the possibility of a change in demand, which pulls him in the direction of the long-run average state. So, while one might have thought that the introduction

of state switching would be enough to make the agent always want to seek after the truth, we see that this is in fact not the case.

The key to the two regimes is that, of the agent's two conflicting objectives (current revenue versus information), one is concave in the choice variable, the other convex. Experimentation is extreme if for some beliefs the combined objective is convex, implying corner solutions – this happens when the frequency of change of the environment and the discount rate are low, so the value of information is high. When either of these parameters increases, current revenue becomes more important; eventually, the combined objective is concave at all beliefs, and we have interior solutions, hence moderate experimentation.

At the parameter values where the combined objective just becomes concave throughout, there is a *discontinuous* change in the optimal policy. Thus, a small increase in the variability of the environment can provoke a near cessation of experimentation, with drastic consequences for the process of information acquisition.

These results are new to the literature on single-agent learning in a changing environment. Previous work has either focused on different aspects of the problem, or used frameworks that lent themselves to only limited analytical investigation.

The two papers closest to ours are Kiefer (1989b) and Balvers and Cosimano (1990), both studying a monopolist learning about changing linear demand curves.¹ In a framework with two possible demand curves, Kiefer calculates the value function numerically, illustrates two types of optimal policy (one continuous, one with a jump) and simulates the corresponding sample paths of beliefs and actions. In Balvers and Cosimano's framework, on the other hand, both intercept and slope of the unknown demand curve are given by stochastic processes, so there is in fact a continuum of possible demand curves. This seems more realistic than a two-state model, but the added complexity makes it very hard to obtain analytical results. Moreover, the absence of a confounding action means that their result of sluggish price adjustments has no direct comparison with our main findings.

While the above papers assume a continuous action space, Rustichini and Wolinsky (1995) use a two-armed bandit framework to study monopoly pricing when the buyers' reservation value changes randomly. Their focus is on non-negligible pricing errors even when the frequency of change is negligible. For certain parameter combinations, learning will cease completely even though the state keeps changing. This can be seen as the analogue in a discrete action space of our moderate experimentation trap.

Note that we depart from the model in Kiefer (1989b), and the overwhelming majority of similar models in the economics literature, by formulating the problem in continuous time.² The advantage of this approach is that it allows us to derive sharp analytical results. In particular, we are able to establish key properties of the value function and the optimal policy, and we obtain some analytical comparative statics results. Further,

¹Learning about an unknown and randomly changing relationship between actions and outcomes is also studied in Balvers and Cosimano (1993, 1994).

²Previously, such a formulation has been adopted by Bolton and Harris (1993) and Felli and Harris (1996) who study multi-agent learning problems with a fixed distribution of rewards. See also Bergemann and Välimäki (1996).

in the continuous-time setting it is straightforward to characterise the sample path properties of beliefs and optimal actions in each of the two experimentation regimes.³

The paper is organised as follows. After presenting the model in Section 1, we proceed to analyse the monopolists's decision problem as an optimal control problem with his belief as the state variable: we describe the evolution of this belief over time (Section 2), then introduce the corresponding Bellman equation and use it to characterise the value function and optimal quantities (Section 3). The main results of the paper are in Section 4 where we look at situations in which there is a confounding belief, and the parameter space splits into two regions: one in which experimentation is extreme, the other in which it is moderate. In particular, we consider the limiting cases of no state switching and no discounting. Section 5 then briefly discusses the simpler scenario when there is no confounding belief; experimentation is moderate for all parameter values, and the comparative statics results are particularly sharp. A summary and concluding remarks follow in Section 6. Technical results are collected in a series of appendices.

1 The Model

There are two states, $k = 0$ or 1 . In state k , the expected demand curve (price as a function of quantity) is given by

$$p = \alpha_k - \beta_k q$$

where the α_k, β_k are positive, fixed and known. The realised price is the expected price plus some noise term.

The state changes according to a continuous time Markov process with the transition probabilities

$$\Pr(k_{t+\Delta t} = 0 \mid k_t = 0) = 1 - \lambda_0 \Delta t + o(\Delta t), \quad \Pr(k_{t+\Delta t} = 1 \mid k_t = 0) = \lambda_0 \Delta t + o(\Delta t),$$

$$\Pr(k_{t+\Delta t} = 0 \mid k_t = 1) = \lambda_1 \Delta t + o(\Delta t), \quad \Pr(k_{t+\Delta t} = 1 \mid k_t = 1) = 1 - \lambda_1 \Delta t + o(\Delta t)$$

where $\lambda_i \geq 0$ ($i = 0, 1$) are known. In particular,

$$\Pr(k_s = i \forall s \in [t, t + \Delta t] \mid k_t = i) = \exp(-\lambda_i \Delta t);$$

see Karlin and Taylor (1981, p.146).

At each moment of time, the monopolist chooses a quantity q_t from an interval $Q = [q_{\min}, q_{\max}]$ of feasible quantities. We assume that production has constant marginal cost, normalised to zero without loss of generality, so revenue equals profit. While we impose a non-negativity constraint on quantities, we do not do so for prices. A negative price associated with a large quantity is simply interpreted as a payment required to make consumers absorb that quantity.

³For example, the incomplete learning result from McLennan (1984) follows immediately by simply setting the state transition rates to zero.

Having produced q_t , the monopolist observes a price which is a noisy signal of whether $k_t = 0$ or 1. More precisely, the increment in total revenue from setting quantity q_t is

$$dR_t = q_t [(\alpha_{k_t} - \beta_{k_t} q_t) dt + \sigma dZ_t]$$

where Z is a standard Wiener process independent of the process k , and $\sigma > 0$, fixed and known. Thus, $dR_t = q_t dP_t$ with the cumulative price process P given by

$$dP_t = (\alpha_{k_t} - \beta_{k_t} q_t) dt + \sigma dZ_t.$$

This is the process which the agent observes. Consequently, an admissible strategy $\mathbf{q} = \{q_t\}$ for the monopolist is such that the action taken at time t depends only on the price history up to that time. The set of all admissible strategies is denoted by \mathcal{Q} . (See Appendix A for a formal definition.)

The monopolist's initial belief about the state of demand is characterised by π , his subjective probability that $k_0 = 1$. Given this belief, the agent's objective is to choose a strategy $\mathbf{q} \in \mathcal{Q}$ so as to maximise

$$\begin{aligned} u^{\mathbf{q}}(\pi) &= \mathbb{E}_{\pi} \left[\int_0^{\infty} r e^{-rt} dR_t \right] \\ &= \mathbb{E}_{\pi} \left[\int_0^{\infty} r e^{-rt} q_t [(\alpha_{k_t} - \beta_{k_t} q_t) dt + \sigma dZ_t] \right] \end{aligned}$$

where $r > 0$ is the discount rate, fixed and known.⁴ Up to the multiplication by r , which expresses the total pay-off in per period terms, $u^{\mathbf{q}}(\pi)$ is the expected present value of the revenue flow from strategy \mathbf{q} . Note that we can also write

$$u^{\mathbf{q}}(\pi) = \mathbb{E}_{\pi} \left[\int_0^{\infty} r e^{-rt} q_t [\alpha_{k_t} - \beta_{k_t} q_t] dt \right]$$

since the stochastic integral with respect to the Wiener process Z has zero expectation.

2 Beliefs

Following a strategy $\mathbf{q} \in \mathcal{Q}$ and observing the associated cumulative price process P , the monopolist updates his beliefs about the state of demand in a Bayesian fashion. Let π_t denote the subjective probability he assigns to state 1 at time t , that is, the conditional probability that $k_t = 1$ given the history of the process P up to t .

By the law of iterated expectations, we have

$$u^{\mathbf{q}}(\pi) = \mathbb{E}_{\pi} \left[\int_0^{\infty} r e^{-rt} \mathbb{E}_{\pi_t} [q_t (\alpha_{k_t} - \beta_{k_t} q_t)] dt \right]$$

where

$$\mathbb{E}_{\pi_t} [q_t (\alpha_{k_t} - \beta_{k_t} q_t)] = q_t [(1 - \pi_t)\alpha_0 + \pi_t\alpha_1 - ((1 - \pi_t)\beta_0 + \pi_t\beta_1)q_t]$$

⁴Later, we shall also consider the undiscounted case, when $r = 0$.

is the expected revenue, given the observed price history, for quantity q_t . To simplify the notation, we introduce the functions

$$\begin{aligned}\alpha(\pi) &= (1 - \pi)\alpha_0 + \pi\alpha_1, \\ \beta(\pi) &= (1 - \pi)\beta_0 + \pi\beta_1\end{aligned}$$

which describe the expected intercept and slope of the demand curve given the belief π , and

$$R(\pi, q) = q [\alpha(\pi) - \beta(\pi)q]$$

which is the corresponding expected revenue from setting quantity q . Thus, we have the representation

$$u^{\mathbf{q}}(\pi) = \mathbb{E}_{\pi} \left[\int_0^{\infty} r e^{-rt} R(\pi_t, q_t) dt \right] \quad (1)$$

which does not involve the stochastic variable k_t any more; instead, expected total pay-off is described as a function of beliefs alone.

This suggests looking at strategies based exclusively on the information contained in beliefs. In the next section we show that optimal strategies are in fact stationary Markov strategies, namely ones where the quantity chosen at time t is a (time-invariant) function of the belief at that time, that is, $q_t = q(\pi_t)$. However, first we have to investigate how beliefs evolve over time.

To this end, we define

$$\lambda(\pi) = (1 - \pi)\lambda_0 - \pi\lambda_1$$

and

$$\Sigma(\pi, q) = \sigma^{-1}\pi(1 - \pi)(\Delta\alpha - \Delta\beta q)$$

where $\Delta\alpha = \alpha_1 - \alpha_0$ is the difference in intercepts and $\Delta\beta = \beta_1 - \beta_0$ the difference in slopes between the two expected demand curves. Then, it follows from Liptser and Shiriyayev (1977, Chapter 9) that given a strategy $\mathbf{q} \in \mathcal{Q}$, the beliefs evolve according to the filtering equation

$$d\pi_t = \lambda(\pi_t) dt + \Sigma(\pi_t, q_t) dZ_t^{\mathbf{q}} \quad (2)$$

where $dZ_t^{\mathbf{q}}$ is the increment of a Wiener process. In other words, the change in beliefs $d\pi_t$ is normally distributed with mean $\lambda(\pi_t) dt$ and variance $\Sigma^2(\pi_t, q_t) dt$.

Equation (2) emphasises the two separate forces which drive the updating. The drift term $\lambda(\pi_t) dt$ takes account of the possibility that the state may change over the next infinitesimal period of time. Given the current belief π , the monopolist assigns probability $1 - \pi$ to state 0, hence probability $(1 - \pi)\lambda_0$ to a transition from state 0 to state 1 over the next instant dt ; in the same way, he assigns probability $\pi\lambda_1$ to a transition from state 1 to state 0. The first possibility increases the probability of being in state 1 after the time dt has elapsed, the second reduces it, and the combined effect leads to the drift term in (2). If at least one of the transition intensities λ_0, λ_1 is nonzero, the linear function λ is downward sloping and vanishes at the *invariant belief*

$$\tilde{\pi} = \frac{\lambda_0}{\lambda_0 + \lambda_1}.$$

In view of this, we let $\Lambda = \lambda_0 + \lambda_1$ and rewrite this function as

$$\lambda(\pi) = -\Lambda (\pi - \tilde{\pi}).$$

This representation shows that state switching introduces mean reversion into the evolution of beliefs. Throughout the paper, we shall fix an invariant belief $\tilde{\pi}$ and use the parameter Λ to measure the intensity of demand curve switches, and hence the instability of the environment in which the monopolist operates.

The diffusion term $\Sigma(\pi_t, q_t) dZ_t^q$ captures the influence of the observed price signal on the evolution of beliefs. Z^q being a Wiener process, this part of the updating is completely unpredictable. Intuitively, this expresses the fact that the current belief already incorporates everything that there is to know, so any change must come as a surprise. The representation

$$dZ_t^q = \sigma^{-1} \left((\alpha_{k_t} - \beta_{k_t} q_t) dt + \sigma dZ_t - [\alpha(\pi_t) - \beta(\pi_t) q_t] dt \right) \quad (3)$$

from Liptser and Shirayev (1977, Chapter 9) confirms this, showing that the change in beliefs depends on the difference between the realised price, $(\alpha_{k_t} - \beta_{k_t} q_t) dt + \sigma dZ_t$, and the expected price, $[\alpha(\pi_t) - \beta(\pi_t) q_t] dt$. The greater the spread $\Delta\alpha - \Delta\beta q_t$ between the two demand curves, and the lower the noise level σ , the more informative is the price signal, and the more pronounced is the change of beliefs after the signal is observed. There is, however, the possibility of a completely uninformative signal: if $\hat{q} = \Delta\alpha/\Delta\beta$ is a feasible quantity, then $\Sigma(\pi, \hat{q}) = 0$, so the diffusion term would vanish were the agent to choose the quantity \hat{q} . All this is relevant of course only if the agent is not subjectively certain of the current state. For $\pi = 0$ or 1 , the agent rules out any possibility of learning from the price signal: $\Sigma(0, q) = 0$ and $\Sigma(1, q) = 0$, so the diffusion term vanishes no matter which action is taken.

Finally, note that we can simplify the expression on the right-hand side of (3) to $\sigma^{-1}(k_t - \pi_t)(\Delta\alpha - \Delta\beta q_t) dt + dZ_t$. Using this to replace dZ_t^q in (2), we obtain

$$d\pi_t = \left\{ \lambda(\pi_t) + \sigma^{-2} \pi_t (1 - \pi_t) (k_t - \pi_t) (\Delta\alpha - \Delta\beta q_t)^2 \right\} dt + \Sigma(\pi_t, q_t) dZ_t. \quad (4)$$

Looking at the term which contains the factor $k_t - \pi_t$, we see that whenever the signal is informative and the agent is not already subjectively certain, his belief is pulled towards the truth.⁵

3 The Bellman Equation

The representation (1) for the pay-off $u^q(\pi)$, the filtering equation (2) for the evolution of beliefs and the fact that Z^q is a Wiener process allow us to consider the monopolist's decision problem as a problem of optimal control of a diffusion process, the diffusion in question being the process of beliefs. Following the standard approach to this type

⁵Note the difference in perspective between equations (2) and (4): the former describes the evolution of beliefs from the perspective of the agent in our model, the latter from that of the modeller who observes the hidden variable k_t .

of control problem,⁶ we now turn to the corresponding value function and Bellman equation.

As usual, the value function is defined as

$$u^*(\pi) = \sup_{\mathbf{q} \in \mathcal{Q}} u^{\mathbf{q}}(\pi) \quad (5)$$

for $\pi \in [0, 1]$. It is clearly bounded and, being the upper envelope of linear pay-off functions $u^{\mathbf{q}}$, it is also continuous and convex, convexity expressing the fact that information is valuable to the agent.⁷ (See Appendix B for details.)

Standard results imply that the value function has further regularity properties, principally that it has a continuous first derivative on $[0, 1]$, and a non-negative second derivative almost everywhere on $]0, 1[$. Moreover, u^* is a solution of the Bellman equation

$$\max_{q \in \mathcal{Q}} \left\{ \frac{1}{2} \Sigma^2(\pi, q) u''(\pi) + \lambda(\pi) u'(\pi) - r u(\pi) + r R(\pi, q) \right\} = 0 \quad (6)$$

almost everywhere; see Appendix C for details.

We give a brief, heuristic derivation of the Bellman equation. From the Principle of Optimality, we see that u^* satisfies

$$u(\pi) = \max_{q \in \mathcal{Q}} \left\{ r R(\pi, q) dt + e^{-r dt} \mathbb{E}_{\pi} [u(\pi + d\pi)] \right\} \quad (7)$$

where the first term is the expected current pay-off, and the second term is the discounted expected continuation value. With regard to the latter, we can approximate $e^{-r dt}$ by $1 - r dt$, and, when u is sufficiently differentiable, Itô's lemma gives us

$$\mathbb{E}_{\pi} [u(\pi + d\pi)] = u(\pi) + u'(\pi) \mathbb{E}_{\pi} [d\pi] + \frac{1}{2} u''(\pi) \mathbb{E}_{\pi} [(d\pi)^2].$$

From equation (2), we see that $\mathbb{E}_{\pi} [d\pi] = \lambda(\pi) dt$ and $\mathbb{E}_{\pi} [(d\pi)^2] = \Sigma^2(\pi, q) dt$. The discounted expected continuation value is therefore

$$(1 - r dt) \left(u(\pi) + \lambda(\pi) u'(\pi) dt + \frac{1}{2} \Sigma^2(\pi, q) u''(\pi) dt \right).$$

Substituting this into (7) and ignoring terms of order $(dt)^2$, we obtain

$$u(\pi) = \max_{q \in \mathcal{Q}} \left\{ r R(\pi, q) dt + u(\pi) - r u(\pi) dt + \lambda(\pi) u'(\pi) dt + \frac{1}{2} \Sigma^2(\pi, q) u''(\pi) dt \right\}$$

which, after simplifying, yields the Bellman equation (6).

The Bellman equation is our main tool for constructing optimal strategies which will in fact be stationary Markov strategies. Such a Markov strategy is derived from

⁶See for instance Fleming and Rishel (1975) and Krylov (1980).

⁷If the agent is uncertain about his prior information, and hence about his prior belief, then he can only gain from resolving this uncertainty before choosing a strategy. If, for instance, the prior belief is π_1 with probability η and π_2 with probability $1 - \eta$, then the agent obtains $u^*(\eta\pi_1 + (1 - \eta)\pi_2)$ if the uncertainty remains unresolved, while he can achieve the expected pay-off $\eta u^*(\pi_1) + (1 - \eta)u^*(\pi_2)$ if the uncertainty is resolved. The latter dominates the former if and only if u^* is convex.

a policy function $q : [0, 1] \rightarrow Q$ which selects the quantity $q_t = q(\pi_t)$ when π_t is the belief at time t . A policy function is admissible if this procedure leads to an admissible strategy $\mathbf{q} \in \mathcal{Q}$ for any given initial belief π_0 ; in Appendix A, we present some regularity conditions which ensure that a given policy function is admissible. Now suppose we have a solution u of the Bellman equation (subject to boundary conditions which we shall discuss after the next subsection) and an admissible policy function $q : [0, 1] \rightarrow Q$ such that

$$u(\pi) \in \arg \max_{q \in Q} \left\{ \frac{1}{2} \Sigma^2(\pi, q) u''(\pi) + \lambda(\pi) u'(\pi) - r u(\pi) + r R(\pi, q) \right\}$$

for all π . Then a standard argument, the verification theorem given in Appendix C, implies that u is the value function and q an optimal policy.

We now discuss the economics behind the Bellman equation.

3.1 Interpretation

Some economic insights can be gained from rewriting the Bellman equation as

$$u(\pi) = \lambda(\pi) \frac{u'(\pi)}{r} + \max_{q \in Q} \left\{ \Sigma^2(\pi, q) \frac{u''(\pi)}{2r} + R(\pi, q) \right\} \quad (8)$$

where the maximisation problem immediately indicates the fundamental trade-off between information gathering and myopic profit maximisation. We look at the three terms on the right-hand side of (8) in turn.

The first term, $\lambda(\pi)u'(\pi)/r$, represents the value of state switches. According to (2), $\lambda(\pi)$ indicates the magnitude and direction of the likely change in belief due to possible state switching. This piece of (passively acquired) information has the shadow price $u'(\pi)/r$. The resulting contribution to the value function is positive if the belief is expected to move in the direction which increases value.

The next term, $\Sigma^2(\pi, q)u''(\pi)/2r$, represents the value of information actively acquired by the agent. Indeed, the discussion after equation (3) above shows that $\Sigma^2(\pi, q)$ provides a measure for the informativeness of the price signal obtained from setting the quantity q . This informativeness is valued with the shadow price $u''(\pi)/2r$. Note that for $\Delta\beta \neq 0$, the value of information is a strictly convex quadratic in q with a global minimum of 0 at $\hat{q} = \Delta\alpha/\Delta\beta$. In particular, the value of information increases strictly with the distance between q and \hat{q} .

The last term, $R(\pi, q)$, represents the myopic pay-off.⁸ Note that $R(\pi, q)$ is a

⁸We refer here to a strong form of myopia which assumes that the current belief will persist forever. If this were the case, setting the quantity q forever would indeed yield

$$\int_0^\infty r e^{-rt} R(\pi, q) dt = R(\pi, q).$$

Alternatively, we can think of this as the expected pay-off for $r = \infty$. As r tends to infinity, the distribution on \mathbb{R}_+ with density $r e^{-rt}$ degenerates to a point mass at $t = 0$, and the agent becomes myopic in so far as he does not care for the future any more: for $r = \infty$, $u^{\mathbf{q}}(\pi) = R(\pi, q_0)$.

concave quadratic in q with a global maximum of

$$m(\pi) = \max_q R(\pi, q) = \frac{\alpha(\pi)^2}{4\beta(\pi)}$$

at the quantity

$$q^m(\pi) = \arg \max_q R(\pi, q) = \frac{\alpha(\pi)}{2\beta(\pi)}.$$

We call the functions m and q^m the *myopic optimum pay-off* and the *myopic policy function*, respectively, and we denote the range of q^m by Q^m . Throughout the paper, we assume that $Q^m \subseteq Q$, so the myopically optimal quantities are always in the choice set of the monopolist. As

$$R(\pi, q) = m(\pi) - \beta(\pi)[q - q^m(\pi)]^2, \tag{9}$$

the myopic pay-off decreases strictly as the distance between q and $q^m(\pi)$ increases.

So, the agent's problem is to choose a quantity that maximises the sum of the value of information actively acquired and the myopic pay-off. This sum is also a quadratic in q and its convexity/concavity depends on whether or not the convexity of the value of information term dominates the concavity of the myopic pay-off. This is the key to the discontinuity in optimal behaviour which we will find in Section 4.

3.2 Experimentation and Information

The agent is said to *experiment* if he deviates from the action that would maximise his myopic pay-off. Such a deviation entails an opportunity cost in the form of a loss in current expected rewards; see (9). On the other hand, a deviation from the myopic optimum might render the price signal more informative. We can thus rephrase the above trade-off as follows: In choosing a quantity, the monopolist has to weigh the opportunity cost of experimentation against its long-term informational benefits.

Economic intuition suggests that the monopolist will not experiment when he is subjectively certain of the current state of demand. For π tending to 0 or 1, we therefore expect that the value of information $\Sigma^2(\pi, q) (u^*)''(\pi)/2r$ tends to zero for all possible quantities q . If this is the case, then the myopic quantity $q^m(0)$ or $q^m(1)$ will be optimal in (8) for $\pi = 0$ or 1, respectively, and formally taking limits in (8), we obtain the following boundary conditions for the value function:

$$u^*(0) - \lambda(0) \frac{(u^*)'(0)}{r} = m(0), \quad u^*(1) - \lambda(1) \frac{(u^*)'(1)}{r} = m(1).$$

This intuition is confirmed in Appendix C where we show that these boundary conditions are indeed satisfied by the value function. Hence the agent does not experiment at the beliefs 0 and 1.

At all other beliefs, evaluating the maximand in (8) for the value function u^* and the quantity $q = q^m(\pi)$ shows that

$$u^*(\pi) - \lambda(\pi) \frac{(u^*)(\pi)'}{r} \geq m(\pi). \tag{10}$$

Moreover, the Bellman equation implies that the agent experiments at a belief π if and only if the inequality (10) is strict at that belief. We will see that there is in fact at most one non-degenerate belief at which this inequality might fail to be strict; this belief is identified next.

3.3 The Confounding Quantity and Belief

Unless stated otherwise, we will make the following

Assumption *The two demand curves intersect at a quantity \hat{q} lying strictly between $q^m(0)$ and $q^m(1)$.*

To be more concrete, we assume without loss of generality that the demand curve in state 1 is steeper than the demand curve in state 0, that is, $\Delta\beta > 0$. With this convention, the assumption amounts to the inequalities $\Delta\alpha > 0$ and

$$\frac{\alpha_0}{2\beta_0} > \frac{\Delta\alpha}{\Delta\beta} > \frac{\alpha_1}{2\beta_1},$$

and the two demand curves intersect at the point with co-ordinates

$$\hat{q} = \frac{\Delta\alpha}{\Delta\beta}, \quad \hat{p} = \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{\Delta\beta}$$

in the (q, p) -plane,⁹ as in Figure 1.

We saw in Section 2 that choosing the quantity \hat{q} leads to a completely uninformative price signal – the expected price for this quantity is \hat{p} regardless of the state of demand. As this constitutes a confounding action in the sense of Easley and Kiefer (1988), we shall refer to \hat{q} as the *confounding quantity*.

If the monopolist were to choose \hat{q} , then he would acquire no information and so for this action to be optimal it must maximise his myopic pay-off; that is, given the belief π , \hat{q} can be optimal only if $\hat{q} = q^m(\pi)$. Straightforward algebra shows strict monotonicity of the myopic policy function, so there is a unique belief at which the monopolist could possibly find it optimal to choose \hat{q} . We denote this belief by $\hat{\pi}$, and, lacking a better name, we may sometimes call it the *confounding belief*. Now, as $q^m(\hat{\pi}) = \hat{q}$, a simple calculation reveals that

$$\hat{\pi} = \frac{\alpha_0}{\Delta\alpha} - 2 \frac{\beta_0}{\Delta\beta}.$$

⁹ Note that under this assumption, the situation faced by the monopolist satisfies the two necessary conditions for experimentation identified by Mirman, Samuelson and Urbano (1993) in a two-period framework: experimentation is informative since a change in quantity affects the informativeness of the price signal ($\Delta\beta \neq 0$); and information is valuable in the sense that different quantities are optimal in the two states ($q^m(0) \neq q^m(1)$). In fact, these two conditions are sufficient for experimentation to occur at almost all beliefs π . To see this, note that $q^m(0) \neq q^m(1)$ implies strict convexity of the myopic pay-off function m , which in turn implies strict convexity of the value function u^* since the Bellman equation rules out linear segments for u^* if m has none. In particular, we have $(u^*)'' > 0$ almost everywhere. As $\Delta\beta \neq 0$, this means that the myopic quantity $q^m(\pi)$ violates the first order condition for the maximisation problem in (8) at almost all π . At the same time, this shows that the inequality (10) is strict almost everywhere.

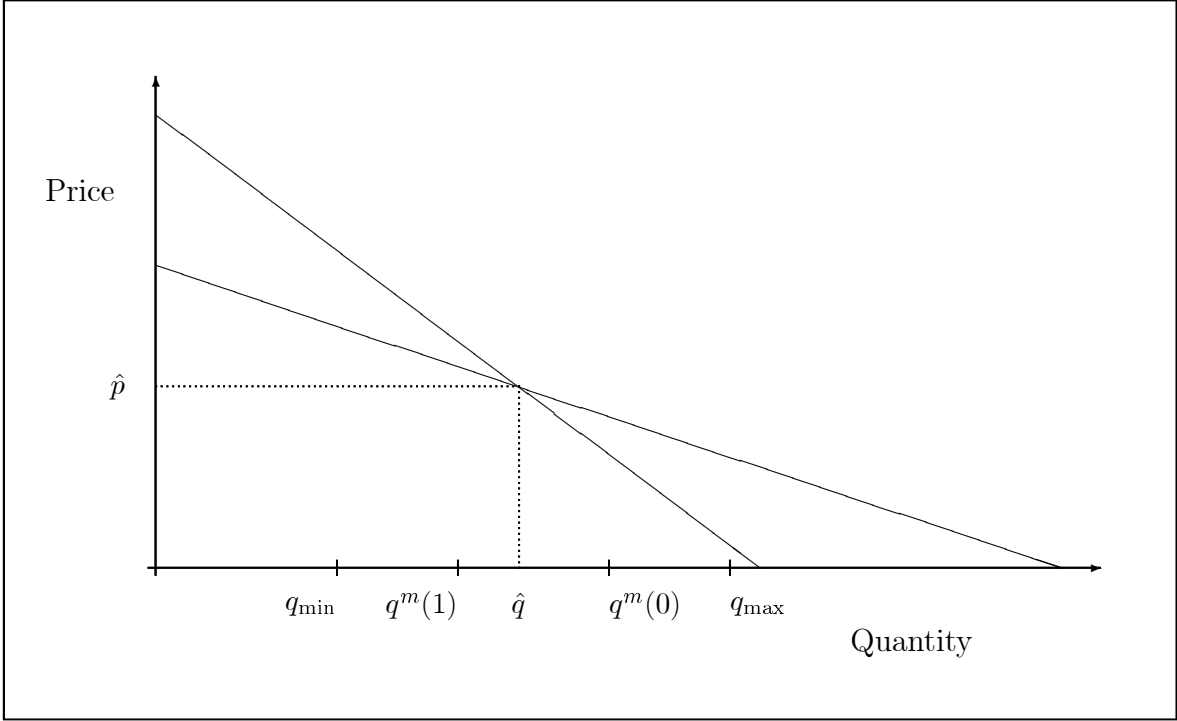


Figure 1: The two demand curves

For future reference, we define $\hat{m} = \hat{p}\hat{q}$. Clearly, $\hat{m} = R(\pi, \hat{q})$, so we can interpret it as the expected revenue, given *any* belief π , from the quantity \hat{q} . In particular, $\hat{m} = m(\hat{\pi})$, and it is easy to verify that this is the global minimum of the myopic pay-off function m ; in fact, m is strictly decreasing on $[0, \hat{\pi}]$ and strictly increasing on $[\hat{\pi}, 1]$.

We noted at the end of Section 3.1 that for each current belief π , the agent maximises the sum of two quadratics, one convex centred on \hat{q} and one concave centred on $q^m(\pi)$. At the belief $\hat{\pi}$, these quadratics are thus centred on the same quantity, as is their sum. Therefore, we expect either *no* experimentation at $\hat{\pi}$ (when the combined quadratic is concave) or *extreme* experimentation (when the combined quadratic is convex) meaning that q_{\max} or q_{\min} is chosen. The Bellman equation indicates further that the case of no experimentation at $\hat{\pi}$ is characterised by the condition

$$u^*(\hat{\pi}) - \lambda(\hat{\pi}) \frac{(u^*)'(\hat{\pi})}{r} = \hat{m},$$

while the case of extreme experimentation corresponds to the inequality

$$u^*(\hat{\pi}) - \lambda(\hat{\pi}) \frac{(u^*)'(\hat{\pi})}{r} > \hat{m}.$$

We shall confirm this below.

3.4 A Differential Equation for the Value Function

The next step in solving the agent's problem is to use the Bellman equation to derive an ordinary differential equation for the value function. The obvious way to do this is in two stages: first calculate optimal quantities in terms of π , $u(\pi)$, $u'(\pi)$ and $u''(\pi)$; then insert these back into the Bellman equation and solve for $u''(\pi)$. However, starting with a simple reformulation of the Bellman equation enables us to get to the desired ODE more directly.

Introducing the notation

$$\tau(\pi) = \Delta\beta^2 \sigma^{-2} \pi^2 (1 - \pi)^2$$

so that $\Sigma^2(\pi, q) = \tau(\pi) [q - \hat{q}]^2$, we can rewrite (8) as

$$u(\pi) - \lambda(\pi) \frac{u'(\pi)}{r} = \max_{q \in Q} \left\{ \tau(\pi) [q - \hat{q}]^2 \frac{u''(\pi)}{2r} + R(\pi, q) \right\}. \quad (11)$$

As $R(\pi, \hat{q}) = \hat{m}$, \hat{q} is suboptimal in (11) as long as $u(\pi) - \lambda(\pi)u'(\pi)/r > \hat{m}$. Under this condition, (11) is then easily seen to be equivalent to

$$\tau(\pi) \frac{u''(\pi)}{2r} = \min_{q \in Q - \{\hat{q}\}} \frac{u(\pi) - \lambda(\pi)u'(\pi)/r - R(\pi, q)}{[q - \hat{q}]^2}, \quad (12)$$

and a quantity $q^* \in Q - \{\hat{q}\}$ attains the maximum in (11) if and only if it attains the minimum in (12).¹⁰

This observation effectively reduces the analysis of the Bellman equation to the analysis of the function

$$G(\pi, v) = \min_{q \in Q - \{\hat{q}\}} \frac{v - R(\pi, q)}{[q - \hat{q}]^2} \quad (13)$$

and the correspondence

$$O(\pi, v) = \arg \min_{q \in Q - \{\hat{q}\}} \frac{v - R(\pi, q)}{[q - \hat{q}]^2}$$

for (π, v) lying in the set

$$\mathcal{A} = \{(\pi, v) \in]0, 1[\times \mathbb{R} : v \geq m(\pi) \text{ and } v > \hat{m}\}.$$

Here, v is a generic variable standing for $u(\pi) - \lambda(\pi)u'(\pi)/r$. Note that the condition $v > \hat{m}$ rules out exactly the point $(\hat{\pi}, \hat{m})$, i.e. the lowest point on the graph of the myopic pay-off function m ; see Figure 2 below.

The function G is well-defined on \mathcal{A} , that is, O is nonempty-valued, and we shall see below that G is continuous on \mathcal{A} .¹¹ This implies that the value function u^* is twice differentiable and solves the ODE

$$\tau(\pi) \frac{u''(\pi)}{2r} = G\left(\pi, u(\pi) - \lambda(\pi) \frac{u'(\pi)}{r}\right) \quad (14)$$

¹⁰A detailed derivation of this equivalence is given in Appendix D.

¹¹We shall give an explicit expression for G which makes continuity obvious. Alternatively, we could show continuity by applying standard arguments which are used in the proof of Berge's Maximum Theorem.

at least on $]0, 1[-\{\hat{\pi}\}$, and on the whole of $]0, 1[$ if $u^*(\hat{\pi}) - \lambda(\hat{\pi})(u^*)'(\hat{\pi})/r > \hat{m}$.¹²

Conversely, we have the following verification theorem. Suppose the function u has a continuous first derivative on $[0, 1]$ and solves (14) on $]0, 1[-\{\hat{\pi}\}$ with the boundary conditions $u(0) - \lambda(0)u'(0)/r = m(0)$ and $u(1) - \lambda(1)u'(1)/r = m(1)$; moreover, suppose that there is an admissible policy function q^* such that

$$q^*(\pi) \in O\left(\pi, u(\pi) - \lambda(\pi)\frac{u'(\pi)}{r}\right)$$

for all $\pi \neq \hat{\pi}$. Then $u = u^*$, and the policy function q^* is optimal. (This follows directly from Proposition C.2 in the Appendix.)

3.5 Optimal Quantities

We turn now to a more explicit analysis of the function G and the optimal quantity correspondence O . We just outline the general structure; details are given in Appendix D. The area \mathcal{A} can be divided into four regions by rays emanating from $(\hat{\pi}, \hat{m})$, as in Figure 2. The regions which border on the curve $v = m(\pi)$ are associated with the minimisation problem in (13) having an interior solution, and the other two are associated with it having a corner solution. In brief, moving clockwise from the left, we shall have: interior solution, corner solution q_{\max} , corner solution q_{\min} , interior solution.

The leftmost ray which separates the first two regions goes up and to the left from $(\hat{\pi}, \hat{m})$ and is determined by the borderline case where the first order condition for the minimisation problem in (13) holds for $q = q_{\max}$. Similarly, the rightmost ray which separates the last two regions goes up and to the right and is determined by the borderline case where the first order condition holds for $q = q_{\min}$. Interior solutions are obtained in the regions (denoted by $\mathcal{A}_{\text{int},\ell}$ and $\mathcal{A}_{\text{int},r}$) which lie below these rays, and are given by¹³

$$O(\pi, v) = q^m(\pi) + \frac{v - m(\pi)}{m(\pi) - \hat{m}} [q^m(\pi) - \hat{q}]. \quad (15)$$

Evaluating the minimand in (13) at these quantities, we find

$$G(\pi, v) = \beta(\pi) \frac{v - m(\pi)}{v - \hat{m}} \quad (16)$$

in the two regions associated with interior solutions.

Corner solutions are obtained in the area between the leftmost and rightmost rays. This area splits into two regions along a third, central ray (denoted by \mathcal{R}_c) which is determined by the borderline case when q_{\max} and q_{\min} are *both* optimal and so give the same value in (13). We have $O(\pi, v) = q_{\max}$ between the left and the central ray, $O(\pi, v) = \{q_{\max}, q_{\min}\}$ along the central ray, and $O(\pi, v) = q_{\min}$ between the central and the right ray. The corresponding expressions for G are

$$G(\pi, v) = \frac{v - R(\pi, q_{\max})}{[q_{\max} - \hat{q}]^2} \quad (17)$$

¹²As $G(\pi, u^*(\pi) - \lambda(\pi)(u^*)'(\pi)/r)$ is continuous in π as long as $u^*(\pi) - \lambda(\pi)(u^*)'(\pi)/r > \hat{m}$, this statement follows directly from Corollary C.1 in the Appendix.

¹³When $O(\pi, v)$ is a singleton, we write $O(\pi, v) = q$ rather than $O(\pi, v) = \{q\}$.

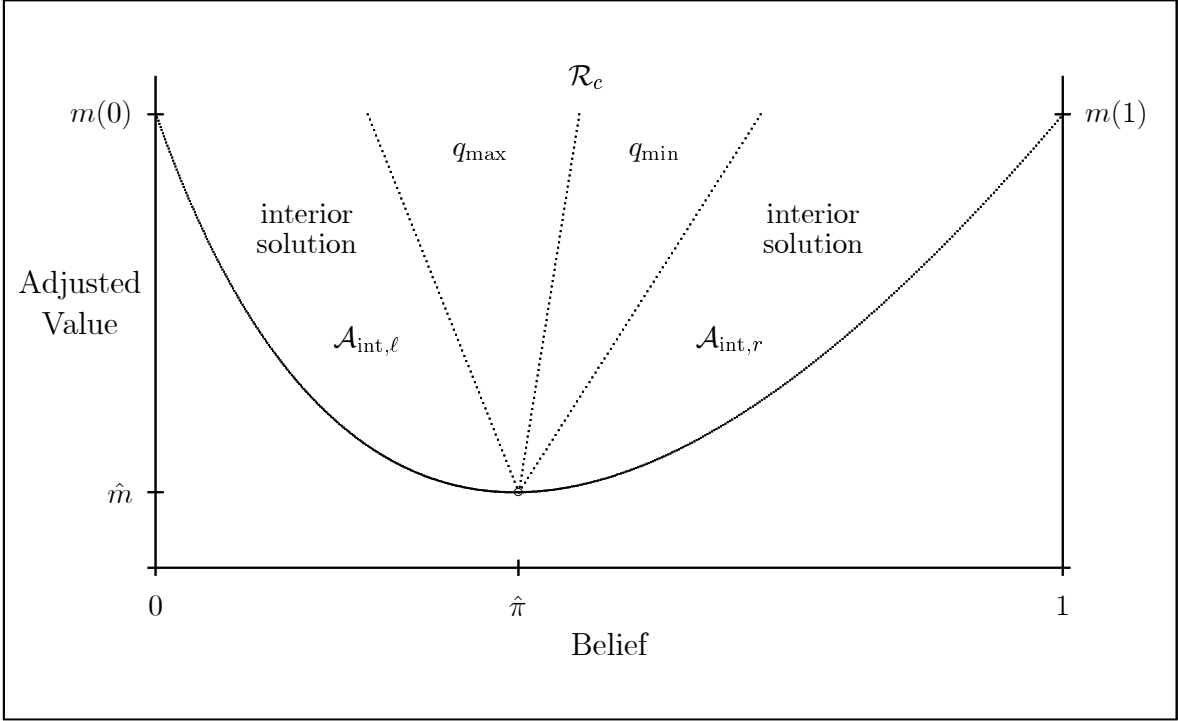


Figure 2: The four regions

The convex curve is the myopic pay-off $v = m(\pi)$.

and

$$G(\pi, v) = \frac{v - R(\pi, q_{\min})}{[\hat{q} - q_{\min}]^2}. \quad (18)$$

Given the representations (16) – (18) on the respective regions, it is now straightforward to verify that G is continuous on \mathcal{A} .¹⁴

3.6 The Adjusted Value Function

The previous subsection shows that optimal quantities depend only on the belief π and the value at π of the function

$$v^*(\pi) = u^*(\pi) - \lambda(\pi) \frac{(u^*)'(\pi)}{r} \quad (19)$$

which we call the *adjusted value function*. Since knowing the adjusted value function will be enough to determine optimal policies, our next step is to transform the ODE

¹⁴Note, however, that it cannot be extended continuously into the point $(\hat{\pi}, \hat{m})$ that we excluded from the set \mathcal{A} . For a sequence of points (π_n, v_n) in \mathcal{A} converging to $(\hat{\pi}, \hat{m})$ along the graph of the myopic pay-off function m , for example, $\lim_{n \rightarrow \infty} G(\pi_n, v_n) = 0$; for a sequence converging to $(\hat{\pi}, \hat{m})$ along the central ray, on the other hand, this limit is $\beta(\hat{\pi})$.

(14) for u^* into an ODE for v^* . Note that the boundary conditions become very simple: v^* and m coincide at the beliefs 0 and 1. Note also that once we know v^* , we can recover u^* by integrating (19), that is, by solving a linear ODE.¹⁵

We formally differentiate the equation $v(\pi) = u(\pi) - \lambda(\pi) u'(\pi)/r$ twice, each time using the relationship $\tau(\pi)u''(\pi)/2r = G(\pi, v(\pi))$ to replace u'' with an expression that involves only π and v . This yields the following second-order ODE for the adjusted value function:

$$\tau(\pi)\frac{v''(\pi)}{2} = rG(\pi, v(\pi)) + \Lambda \left\{ f(\pi)G(\pi, v(\pi)) + (\pi - \tilde{\pi})\frac{d}{d\pi}G(\pi, v(\pi)) \right\} \quad (20)$$

with

$$f(\pi) = 2 - (\pi - \tilde{\pi})\frac{\tau'(\pi)}{\tau(\pi)} = 2 - (\pi - \tilde{\pi})\frac{2(1 - 2\pi)}{\pi(1 - \pi)}.$$

As for differentiability of G , it is easy to check that G is continuously differentiable in the interior of \mathcal{A} with the exception of the central ray separating the regions where q_{\max} or q_{\min} is optimal. We will therefore consider the ODE (20) separately to the left and to the right of that ray.

Summarising the developments so far, we can say that the adjusted value function solves (20) on $]0, 1[$ with the possible exception of the confounding belief $\hat{\pi}$ or any belief where the graph of v^* crosses the central ray. Conversely, if we have a solution (in the sense of the previous sentence) v of (20) with the above boundary conditions and such that $q^*(\pi) = O(\pi, v(\pi))$ defines an admissible policy function, then $v = v^*$ and the policy q^* is optimal. It is mainly this version of a verification theorem that we will use below.

When the optimisation problem in the Bellman equation has an interior solution, $G(\pi, v)$ is given by (16), so the ODE (20) becomes

$$\begin{aligned} \tau(\pi)\frac{v''(\pi)}{2} = & \beta(\pi) \left\{ \left(r + \Lambda \left[f(\pi) + (\pi - \tilde{\pi})\frac{\Delta\beta}{\beta(\pi)} \right] \right) \frac{v(\pi) - m(\pi)}{v(\pi) - \hat{m}} \right. \\ & \left. + \Lambda (\pi - \tilde{\pi}) \left[\frac{v(\pi) - m(\pi)}{v(\pi) - \hat{m}} \right]' \right\} \end{aligned} \quad (21)$$

in this case. Many of the results obtained in the following sections are based on a detailed investigation of this particular differential equation.

While we have derived the above statements for a discount rate $r > 0$ only, they continue to be valid in the limiting case of no discounting ($r = 0$) once we use a definition of the adjusted value function that corresponds to the so-called *catching-up criterion*; we refer the reader to Appendix E for details.¹⁶ The undiscounted case

¹⁵In fact, the method of variation of constants shows that for $\Lambda > 0$,

$$u(\pi) = (r/\Lambda) |\pi - \tilde{\pi}|^{-r/\Lambda} \text{sign}(\pi - \tilde{\pi}) \int_{\tilde{\pi}}^{\pi} |\xi - \tilde{\pi}|^{r/\Lambda - 1} v(\xi) d\xi$$

is the unique bounded solution of the ODE $u(\pi) - \lambda(\pi)u'(\pi)/r = v(\pi)$ for any given continuous function v .

¹⁶See Dutta (1991) for a discussion of undiscounted decision criteria in a discrete time framework.

provides a useful benchmark; in fact, economic intuition suggests that the optimal experimentation strategy of an agent with discount rate $r > 0$ will be somewhere ‘in between’ the two extremes given by myopic behaviour (corresponding to $r = \infty$), on the one hand, and the behaviour of an infinitely patient agent ($r = 0$), on the other hand.¹⁷

4 Experimentation Regimes

The assumption that the confounding quantity \hat{q} lies in the interior of Q^m brings with it two complications. A first complication arises from the fact that it might be optimal at $\hat{\pi}$ to choose \hat{q} . As we saw already, choosing \hat{q} leads to a completely uninformative price signal, makes the diffusion coefficient in the updating equation (2) vanish and thereby causes a singularity in the Bellman equation and the related ODEs. Moreover, there is a ‘break’ in the ODE for the adjusted value function along the central ray.

A second complication arises from the fact that the direction of increasing informativeness of the price signal is ambiguous. Assume for example that the current belief is slightly higher than $\hat{\pi}$, so the myopically optimal quantity is slightly below \hat{q} . The true optimum will usually involve some deviation from the myopic quantity, motivated by the desire to render observed prices more informative, and naïve intuition suggests that the monopolist might wish to move further away from \hat{q} by reducing quantity. However, it could also make sense to *increase* quantity beyond \hat{q} and thus achieve a wider spread between the two possible price distributions there.

For beliefs close to the boundaries of the unit interval, on the other hand, we do expect the naïve intuition to be borne out. Thus, we expect optimal experimentation in *both* directions, involving both quantity expansion for beliefs π close to 0, and quantity reduction for beliefs π close to 1. The optimal policy as a function of beliefs will then have to move downward past \hat{q} as π increases. This raises the following question: does the optimal quantity change continuously, or is there a jump?

Answering this question, we shall first characterise the two regimes of optimal experimentation that can arise. In the *moderate* experimentation regime, the optimal policy selects quantities in Q^m , the range of the myopic policy, only, and it chooses \hat{q} at $\hat{\pi}$. In the *extreme* experimentation regime, each of the quantities q_{\max} and q_{\min} is chosen on a set of beliefs of positive measure; in particular, q_{\max} or q_{\min} will be chosen at $\hat{\pi}$, and the optimal policy will exhibit a jump past \hat{q} from one extreme quantity to the other. These regimes are further distinguished by the sample path behaviour of posterior beliefs and optimal quantities. While extreme experimentation implies that any posterior belief can be reached with positive probability, moderate experimentation restricts posterior beliefs to lie on one side of $\hat{\pi}$ in the long run, so the monopolist ends up producing quantities from only part of Q^m .

We will show that extreme experimentation arises for low values of r , Λ and σ , and moderate experimentation for high values. Near the boundary between the corresponding parameter regions, a small change in any of these parameters can trigger

¹⁷Moreover, it is well known that the undiscounted case tends to be mathematically more tractable than the discounted case. See for instance Bolton and Harris (1993) and Harris (1988).

a change in the experimentation regime, hence a large *discontinuous* change in the monopolist's strategy and the resulting sample path behaviour of beliefs and quantities produced. These results are particularly clear in the limiting cases where at least one of the parameters r and Λ is zero. The simple benchmark where the monopolist uses the catching-up criterion ($r = 0$) and the environment does not change ($\Lambda = 0$) allows a closed-form solution for the optimal policy. In the undiscounted case with state switching ($r = 0$ and $\Lambda > 0$) we will establish the existence of a critical switching intensity Λ^* that separates moderate from extreme experimentation. Similarly, we will find a critical discount rate r^* in the discounted case without state switching ($r > 0$ and $\Lambda = 0$).

Throughout this section, we fix demand curve parameters $\alpha_0, \alpha_1, \beta_0$ and β_1 such that the confounding quantity \hat{q} lies in the interior of Q^m ; an invariant belief $\tilde{\pi} \in]0, 1[-\{\hat{\pi}\}$ is also held fixed.

4.1 Moderate versus Extreme Experimentation

Let Θ be the set of all possible combinations of the three parameters r, Λ and σ , that is, $\Theta = \{(r, \Lambda, \sigma) : r \geq 0, \Lambda \geq 0, \sigma > 0\}$. Let $(D_{\tilde{\pi}}v)(\hat{\pi})$ denote the one-sided derivative of a function v at $\hat{\pi}$ in the direction of $\tilde{\pi}$, and $(D_{\sigma}v)(\hat{\pi})$ the one-sided derivative in the opposite direction. Recall the structure of the ODE (20) for the adjusted value function, and in particular its special case (21) corresponding to interior solutions of the optimisation problem in the Bellman equation.

Theorem 4.1 *The parameter set Θ can be partitioned into two non-empty sets Θ_m and Θ_e such that:*

- (i) *If $(r, \Lambda, \sigma) \in \Theta_m$, then $v^*(\hat{\pi}) = \hat{m}$, $v^* > m$ on $]0, 1[-\{\hat{\pi}\}$, v^* is strictly convex with $(v^*)'' > 0$ on $]0, 1[-\{\hat{\pi}\}$, and $(D_{\tilde{\pi}}v^*)(\hat{\pi}) = 0$. Furthermore, v^* is the unique solution of the ODE (21) on $]0, 1[-\{\hat{\pi}\}$ subject to $v^*(\pi) = m(\pi)$ at $\pi = 0, \hat{\pi}, 1$, $v^* > m$ everywhere else, and $(D_{\tilde{\pi}}v^*)(\hat{\pi}) = 0$.*
- (ii) *If $(r, \Lambda, \sigma) \in \Theta_e$, then $v^* > m$ on the whole of $]0, 1[$, and v^* is the unique differentiable function which solves the ODE (20) on $\{\pi \in]0, 1[: (\pi, v^*(\pi)) \notin \mathcal{R}_c\}$ subject to the conditions $v^*(0) = m(0)$, $v^*(1) = m(1)$ and $v^* > m$ everywhere else.*

PROOF: Let Θ_m be the set of all parameter combinations in Θ for which the corresponding adjusted value function satisfies $v^*(\hat{\pi}) = \hat{m}$, and Θ_e the set of all combinations for which the adjusted value function satisfies $v^*(\hat{\pi}) > \hat{m}$. Since $v^*(\hat{\pi}) \geq \hat{m}$ for all possible parameters, these two sets trivially partition Θ . We shall see later that neither of them is empty.

For $v^*(\hat{\pi}) = \hat{m}$, the statements on convexity and the one-sided derivative $(D_{\tilde{\pi}}v^*)(\hat{\pi})$ are shown in Appendix F; see Proposition F.2. Given $v^*(\hat{\pi}) = \hat{m}$ and the boundary conditions, convexity entails $(\pi, v^*(\pi)) \in \mathcal{A}_{\text{int},\ell} \cup \mathcal{A}_{\text{int},r}$ for all $\pi \in]0, 1[-\{\hat{\pi}\}$. Sections 3.4 and 3.6 therefore imply that v^* solves (21) on $]0, 1[-\{\hat{\pi}\}$ subject to the stated

conditions. If v is another solution of (21) on $]0, 1[- \{\hat{\pi}\}$ with $v(\pi) = m(\pi)$ at $\pi = 0, \hat{\pi}, 1$ and $(D_{\hat{\pi}}v)(\hat{\pi}) = 0$, then the construction of (ϵ) -optimal strategies in the proof of Proposition 4.2 below and our verification theorem imply that $v = v^*$. Finally, if there were a $\pi \neq 0, \hat{\pi}, 1$ such that $v^*(\pi) = m(\pi)$, then $v^* - m$ would have a local minimum there, hence $(v^*)''(\pi) \geq m''(\pi) > 0$. But $v^*(\pi) = m(\pi)$ would imply $(u^*)''(\pi) = 0$ by ODE (14) (or its undiscounted variant), hence $(u^*)''$ would have a local minimum at π . This would require $(u^*)'''(\pi) = 0$ and, since $(v^*)''(\pi)$ is a linear combination of $(u^*)''(\pi)$ and $(u^*)'''(\pi)$, we would have $(v^*)''(\pi) = 0$ – a contradiction.

For $v^*(\hat{\pi}) > \hat{m}$, Sections 3.4 and 3.6 imply that v^* is once continuously differentiable and solves the ODE (20) on $\{\pi \in]0, 1[: (\pi, v^*(\pi)) \notin \mathcal{R}_c\}$. Given another solution v with this property and the same boundary conditions, the arguments in the proof of Proposition 4.1 below together with the verification theorem imply again that $v = v^*$. Finally, the same argument as above shows $v^* > m$ for all beliefs $\pi \neq 0, 1$ with $(\pi, v^*(\pi)) \notin \mathcal{R}_c$; this implies $v^* > m$ on the whole of $]0, 1[$. ■

The following two propositions characterise the optimal experimentation behaviour in scenarios (i) and (ii) of the theorem. We start with (ii), which is the easier case.

Proposition 4.1 (Extreme Experimentation) *If $v^*(\hat{\pi}) > \hat{m}$, then the optimal policy function prescribes each of the extreme quantities q_{\max} and q_{\min} on a set of beliefs of positive measure, and is continuous except for a jump from one extreme quantity to the other at each belief π such that $(\pi, v^*(\pi)) \in \mathcal{R}_c$. The corresponding process of posterior beliefs is regular on $]0, 1[$, that is, starting from any point in this interval, any other point in it may be reached with positive probability.*

PROOF: The policy function q^* obtained by extending $O(\pi, v^*(\pi))$ continuously into $\pi = 0$ and 1 and selecting either q_{\max} or q_{\min} at each π where $(\pi, v^*(\pi)) \in \mathcal{R}_c$ is piecewise continuous with $[q^*(\pi) - \hat{q}]^2$ bounded away from zero, hence admissible by Proposition A.1 and optimal by the verification theorem from Section 3.6. Clearly, $q^*(\pi) = q_{\max}$ on a set of positive measure, and the same is true for q_{\min} . The fact that $[q^*(\pi) - \hat{q}]^2$ is bounded away from zero also implies regularity of the process of posterior beliefs. ■

For parameters in Θ_e , we thus find extreme experimentation in the sense that the quantities q_{\max} and q_{\min} are optimal for non-negligible sets of beliefs. Moreover, optimal quantities are always some distance away from the confounding quantity, so the information content of the price signal observed by the monopolist is bounded away from zero. The resulting process of posterior beliefs can therefore reach any point in the open unit interval with positive probability.

An example of extreme experimentation is shown in Figure 3 which has been calculated for $r = 0.1$ and $\Lambda = 0.05$.¹⁸ The bold line in the upper panel is the graph of the

¹⁸In this and all the subsequent figures, the demand curve parameters are $\alpha_0 = 40, \beta_0 = 2/3, \alpha_1 = 60$ and $\beta_1 = 3/2$, implying $q^m(0) = 30, q^m(1) = 20, \hat{q} = 24$ and $\hat{\pi} = 0.4$. The range of feasible quantities is defined by $q_{\min} = 40/3$ and $q_{\max} = 40$, the noise parameter is $\sigma = 5$, and the invariant belief is $\bar{\pi} = 0.5$. Only r and Λ vary across figures. The adjusted value function is calculated as a numerical solution to the ODE (20) subject to the boundary conditions $v^*(0) = m(0)$ and $v^*(1) = m(1)$, and

adjusted value function v^* , the thin line that of the myopic pay-off function m , and the bold dashed line that of the value function u^* . The upper panel also shows the three rays introduced in Section 3.5. In the lower panel, the bold line is the optimal policy function q^* , while the thin line is the myopic policy q^m . The adjusted value is strictly higher than the myopic pay-off at all non-degenerate beliefs, and as v^* crosses each ray in turn, q^* first reaches q_{\max} , then jumps to q_{\min} , and finally moves away from q_{\min} .

The corresponding sample path behaviour is illustrated in Figure 4. The upper panel shows the evolution of the agent's belief starting from the prior $\pi_0 = 0.25$, the lower panel the associated quantities. The bold dashed line in either panel represents the true state, the initial state being $k_0 = 0$. By the time of the first state change, the agent's belief has predominantly been between 0 and 0.2. After the state change, the belief moves relatively quickly in the direction of the new state and eventually reaches $\hat{\pi}$. This starts a phase of intense experimentation with frequent jumps between q_{\max} and q_{\min} . At the end of this phase, the belief leaves the neighbourhood of $\hat{\pi}$ to move closer to the true state. This pattern is repeated each time the state switches, and the true state is tracked quite well.

Next, we turn to case (i) of Theorem 4.1, $v^*(\hat{\pi}) = \hat{m}$. This case is more complicated since it involves the 'singularity' at $(\hat{\pi}, \hat{m})$ of the ODE for v^* . We formulate the next result for $\Lambda > 0$; we shall obtain the analogous result for $\Lambda = 0$ later, in Section 4.3.

Proposition 4.2 (Moderate Experimentation) *Let $\Lambda > 0$ and $v^*(\hat{\pi}) = \hat{m}$. Then the optimal policy assumes values in Q^m only and selects \hat{q} at $\hat{\pi}$. With probability one, the resulting process of posterior beliefs is confined to one of the subintervals $]0, \hat{\pi}[$ or $]\hat{\pi}, 1[$ in the long run, and the monopolist ends up choosing quantities in either $]\hat{q}, q^m(0)[$ or $]q^m(1), \hat{q}[$ only.*

More precisely, the proof will show that the optimal policy function is differentiable if v^* is differentiable at $\hat{\pi}$, while it has a single jump at $\hat{\pi}$ if v^* has a kink.

PROOF: For $\pi \in]0, 1[- \{\hat{\pi}\}$, we have $(\pi, v^*(\pi)) \in \mathcal{A}_{\text{int},\ell} \cup \mathcal{A}_{\text{int},r}$ by convexity of v^* and hence

$$O(\pi, v^*(\pi)) = q^m(\pi) + \frac{v^*(\pi) - m(\pi)}{m(\pi) - \hat{m}} [q^m(\pi) - \hat{q}] = \hat{q} + \frac{v^*(\pi) - \hat{m}}{m(\pi) - \hat{m}} [q^m(\pi) - \hat{q}]$$

from (15). Strict convexity of v^* also implies that $v^* < \bar{m}_\ell$ on $]0, \hat{\pi}[$ and $v^* < \bar{m}_r$ on $]\hat{\pi}, 1[$ with \bar{m}_ℓ and \bar{m}_r being the functions whose graphs are the straight lines joining the point $(\hat{\pi}, \hat{m})$ with the points $(0, m(0))$ and $(1, m(1))$, respectively. It is straightforward to verify that $O(\pi, \bar{m}_\ell(\pi)) = q^m(0)$ on $]0, \hat{\pi}[$ and $O(\pi, \bar{m}_r(\pi)) = q^m(1)$ on $]\hat{\pi}, 1[$. Since $v^* > m$ on $]0, 1[$, we conclude that $q^m(\pi) < O(\pi, v^*(\pi)) < q^m(0)$ on $]0, \hat{\pi}[$ and $q^m(\pi) > O(\pi, v^*(\pi)) > q^m(1)$ on $]\hat{\pi}, 1[$. In particular, $O(\pi, v^*(\pi))$ assumes values in Q^m only.

Straightforward algebra shows that

$$O(\pi, v^*(\pi)) = \hat{q} - \frac{2}{\Delta\alpha} \frac{v^*(\pi) - \hat{m}}{\pi - \hat{\pi}},$$

optimal quantities are then determined through the optimal policy correspondence. Details of the numerical procedure are reported in Appendix H.

so $O(\pi, v^*(\pi)) \rightarrow \hat{q}$ as $\pi \rightarrow \hat{\pi} \pm$ iff $(D_{\pm}v^*)(\hat{\pi}) = 0$.

If $(v^*)'(\hat{\pi}) = 0$, $O(\pi, v^*(\pi))$ can therefore be extended to a continuous policy function $q^* : [0, 1] \rightarrow Q^m$ with $q^*(\hat{\pi}) = \hat{q}$. In fact, the policy is differentiable with bounded derivative, hence Lipschitz continuous. This is obvious for beliefs different from $\hat{\pi}$; differentiability at $\hat{\pi}$ follows from the representation

$$\frac{q^*(\pi) - \hat{q}}{\pi - \hat{\pi}} = \frac{v^*(\pi) - \hat{m}}{m(\pi) - \hat{m}} \frac{q^m(\pi) - \hat{q}}{\pi - \hat{\pi}}$$

and the fact that the ratio $[v^*(\pi) - \hat{m}]/[m(\pi) - \hat{m}]$ tends to a finite limit as $\pi \rightarrow \hat{\pi}$ (see Proposition F.2). The policy q^* is admissible by Proposition A.1, hence optimal by the verification theorem from Section 3.6. Turning to the belief process resulting from this policy, let us assume $\tilde{\pi} > \hat{\pi}$ for concreteness. Starting from a prior belief π_0 in the subinterval $[\hat{\pi}, 1]$, all posterior beliefs π_t will remain in the open subinterval $]\hat{\pi}, 1[$ because the inequality $\lambda(\hat{\pi}) > 0$ makes the belief $\hat{\pi}$ an entrance boundary. (Since $\lambda(1) < 0$ for $\tilde{\pi} \neq 1$, the right boundary of the unit interval is always an entrance boundary.) If $\pi_0 < \hat{\pi}$, on the other hand, the process of beliefs will, with probability one, reach $\hat{\pi}$ in finite time and then move into the subinterval $]\hat{\pi}, 1[$.

Next, suppose that v^* has a kink at $\hat{\pi}$. To be concrete, we assume again that $\tilde{\pi} > \hat{\pi}$, so $(D_-v^*)(\hat{\pi}) < 0$ and $(D_+v^*)(\hat{\pi}) = 0$. Let q^* be the policy function obtained by extending $O(\pi, v^*(\pi))$ continuously into $\pi = 0$ and 1 and setting $q^*(\hat{\pi}) = \hat{q}$; again, this function takes values in Q^m only. By Proposition F.2, the ratio $[v^*(\pi) - \hat{m}]/[m(\pi) - \hat{m}]$ tends to a finite limit as $\pi \rightarrow \hat{\pi}+$. As above, this implies that the restriction of the policy function q^* to the interval $[\hat{\pi}, 1]$ is Lipschitz continuous, hence admissible. Moreover, if the prior belief π_0 lies in this subinterval and the monopolist uses the policy q^* , then all posterior beliefs π_t will remain in $]\hat{\pi}, 1[$ by the same argument as above. By the verification theorem, the policy q^* is thus optimal for all prior beliefs $\pi_0 \geq \hat{\pi}$.

From the above expression for $O(\pi, v^*(\pi))$, we see that q^* approaches the limit

$$q^*(\hat{\pi}-) = \hat{q} - \frac{2}{\Delta\alpha} (D_-v^*)(\hat{\pi}) \in]\hat{q}, q^m(0)]$$

from the left of $\hat{\pi}$. On the subinterval $[0, \hat{\pi}[$, the function q^* is (locally) Lipschitz continuous, so the existence result underlying Proposition A.1 implies that, starting from any prior belief $\pi_0 < \hat{\pi}$, the policy q^* generates a unique stochastic process of beliefs up to the first time $\hat{\pi}$ is reached; with probability one, this happens in finite time. From then on, the process of beliefs is uniquely determined by the restriction of q^* to $[\hat{\pi}, 1]$. This establishes admissibility of the policy q^* on the whole of the unit interval,¹⁹ and optimality follows from the verification theorem. \blacksquare

¹⁹The technical details required to make this argument fully rigorous are beyond the scope of this paper. One can avoid these complications altogether without affecting the main results by constructing ϵ -optimal policies. Given an arbitrary $\epsilon > 0$, define $\delta = \epsilon/(1 + \max_{q \in Q^m} [q - \hat{q}]^2)$. Since $(D_-v^*)(\hat{\pi}) < 0$, $\tau(\pi)(u^*)''(\pi)/(2r)$ converges to $\beta(\hat{\pi})$ as $\pi \rightarrow \hat{\pi}-$. (We only deal with the case $r > 0$ here; a similar argument can be given for the undiscounted case.) Next, the continuity of v^* implies that $R(\pi, q) - v^*(\pi)$ converges to $-\beta(\hat{\pi})[q - \hat{q}]^2$ as $\pi \rightarrow \hat{\pi}$, and this convergence is uniform in $q \in Q^m$. So

The proposition shows that optimal experimentation is moderate for parameters in Θ_m , meaning that the monopolist restricts himself to quantities in Q^m . The optimal policy function approaches the confounding quantity from at least one side of $\hat{\pi}$, the side where $\tilde{\pi}$ lies. In a changing environment, this implies that starting from any prior belief lying on the same side of $\hat{\pi}$ as $\tilde{\pi}$, the process of posterior beliefs will stay on this side forever; starting from a prior belief on the other side of $\hat{\pi}$, the process of posterior beliefs will cross $\hat{\pi}$ almost surely in finite time and then be confined to the side where $\tilde{\pi}$ lies. Eventually, the monopolist's beliefs will thus be 'trapped' on one side of $\hat{\pi}$, although he knows that the state of demand will continue to switch from time to time. This result is the analogue, in a changing environment, of the possibility of cessation of learning in an unchanging environment as identified by Rothschild (1974), McLennan (1984), Easley and Kiefer (1988), Aghion, Bolton, Harris and Jullien (1991) and others; cf. our discussion of the case $\Lambda = 0$ below.

Figure 5 shows an example of moderate experimentation, calculated for $r = 0.1$ and $\Lambda = 0.2$. Again, the bold line in the upper panel is the adjusted value function v^* , the thin line the myopic pay-off function m , and the bold dashed line the value function u^* . In the lower panel, the bold line is the optimal policy function q^* , and the thin line the myopic policy q^m . The adjusted value function v^* touches the myopic pay-off at its lowest point. Consequently, the optimal policy q^* never selects quantities outside the range of the myopic policy, spanned by $q^m(0) = 30$ and $q^m(1) = 20$. Furthermore, it appears that v^* is differentiable at $\hat{\pi} = 0.4$, and q^* moves smoothly through $\hat{q} = 24$.

Figure 6 illustrates the corresponding sample path behaviour. The upper panel shows how the belief process is trapped after its transit through $\hat{\pi}$; in particular, the true state (again represented by a bold dashed line) is tracked poorly. As a consequence, the resulting path of optimal quantities in the lower panel remains in the range from $q^m(1) = 20$ to $\hat{q} = 24$ after its first passage through the confounding quantity.

The moderate experimentation scenario where v^* has a kink at $\hat{\pi}$ is particularly interesting.²⁰ If $\tilde{\pi} > \hat{\pi}$, say, and $(D_-v^*)(\hat{\pi}) > 0$, the optimal policy approaches a limit different from \hat{q} as $\pi \rightarrow \pi^-$; see the proof of Proposition 4.2. This is due to the fact that to the left of $\hat{\pi}$, $(v^*)'$ is bounded away from zero, so $(u^*)''$ is relatively high, implying a high value of information. Intuitively, we can interpret this as follows. With a posterior belief slightly to the left of $\hat{\pi}$, the agent anticipates that once his belief crosses $\hat{\pi}$, he will not find it profitable to experiment in a way that would allow his belief to cross $\hat{\pi}$ from the right to the left again. Therefore, he experiments relatively strongly so as to

we can find $\rho > 0$ such that $\tau(\pi)(u^*)''(\pi)/(2r) \geq \beta(\hat{\pi}) - \delta$ and $R(\pi, q) - v^*(\pi) \geq -\beta(\hat{\pi})[q - \hat{q}]^2 - \delta$ for all $\pi \in [\hat{\pi} - \rho, \hat{\pi}]$ and all $q \in Q^m$. The Lipschitz continuous (hence admissible) policy function $q_\epsilon : [0, 1] \rightarrow Q^m$ which coincides with q^* on $[0, \hat{\pi} - \rho] \cup [\hat{\pi}, 1]$ and whose graph joins the points $(\hat{\pi} - \rho, q^*(\hat{\pi} - \rho))$ and $(\hat{\pi}, \hat{q})$ by a straight line satisfies

$$\tau(\pi) \frac{(u^*)''(\pi)}{2r} [q_\epsilon(\pi) - \hat{q}]^2 + R(\pi, q_\epsilon(\pi)) - v^*(\pi) \geq -\epsilon$$

on $[\hat{\pi} - \rho, \hat{\pi}]$, hence is ϵ -optimal by Proposition C.2. The resulting long-run behaviour of beliefs and actions is as in the proposition.

²⁰Note that this is the only case where the value function u^* fails to be twice continuously differentiable on the whole of $]0, 1[$.

give his belief a chance to avoid the trap for now and move away from $\hat{\pi}$ to the left, should the current true state be $k = 0$.

The above results show that the optimal experimentation behaviour depends *qualitatively* on the parameters r , Λ and σ , and the two regimes of experimentation lead to markedly different behaviour of beliefs in the long run. These results also imply that a small change in one of the parameters r , Λ and σ can trigger a large *discontinuous* change in the monopolist's strategy and the resulting stream of quantities that he produces. We will see this particularly clearly in the special cases which we study below. Finally, note that the more variable the environment, the less variable are the agent's actions, and *vice versa*.

4.2 Sufficient Conditions and Critical Parameter Values

We would naturally expect moderate experimentation for high values of r , Λ and σ , and extreme experimentation for low values. In fact, high values of these parameters are bound to reduce the value of information so much that the monopolist rationally assesses the loss in current revenue from experimenting strongly near $\hat{\pi}$ to be higher than the loss in future revenues resulting from sometimes being trapped on the 'wrong' side of $\hat{\pi}$. Before formulating a result to this effect, we first want to point out that the choice of the interval Q of feasible quantities is irrelevant for the partition of the parameter set Θ into Θ_m and Θ_e . This is a direct corollary of Theorem 4.1.

Corollary 4.1 *Given a parameter combination (r, Λ, σ) , experimentation is moderate (extreme) for some $Q \supseteq Q^m$ if and only if it is moderate (extreme) for all $Q \supseteq Q^m$.*

PROOF: This follows immediately from the fact that if $v^*(\hat{\pi}) = \hat{m}$, then v^* is characterised by the ODE (21) which does not depend on the quantities q_{\max} and q_{\min} as long as $Q = [q_{\min}, q_{\max}] \supseteq Q^m$. ■

We are therefore free to choose the interval Q in a convenient way when we look for sufficient conditions for either type of experimentation. This is exploited in the proof of the following result.

Proposition 4.3 *There are positive constants $\gamma_m \geq \gamma_e$ and $\kappa_m \geq \kappa_e$ such that optimal experimentation is moderate with a differentiable policy function for all $r \geq 0$, $\Lambda \geq 0$ and $\sigma > 0$ satisfying*

$$\frac{r}{\kappa_m} + \frac{\Lambda}{\gamma_m} \geq \frac{1}{\sigma^2},$$

and extreme for all $r \geq 0$, $\Lambda \geq 0$ and $\sigma > 0$ satisfying

$$\frac{r}{\kappa_e} + \frac{\Lambda}{\gamma_e} < \frac{1}{\sigma^2}.$$

PROOF: By Proposition G.3, there are positive constants c_1 and c_2 such that for all $r \geq 0$, $\Lambda \geq 0$ and $\sigma > 0$ satisfying $r + c_1\Lambda \geq c_2/\sigma^2$, there exists a continuous function

$v : [0, 1] \rightarrow \mathbb{R}$ which solves (21) on $]0, 1[- \{\hat{\pi}\}$ with $v(\pi) = m(\pi)$ at $\pi = 0, \hat{\pi}, 1$, $v > m$ everywhere else, and $v - \hat{m} \leq 2(m - \hat{m})$ in a neighbourhood of $\hat{\pi}$. In particular, v is differentiable at $\hat{\pi}$ with $v'(\hat{\pi}) = 0$. The uniqueness part of Theorem 4.1 (i) implies $v = v^*$. Experimentation is therefore moderate with a continuous optimal policy for all these parameter combinations. We can thus set $\kappa_m = c_2$ and $\gamma_m = c_2/c_1$.

Turning to extreme experimentation, we assume without loss of generality that $q_c = \hat{q}$, where q_c is the midpoint of the interval Q . This makes the central ray \mathcal{R}_c vertical and simplifies the construction of solutions to the ODE (20) via the techniques of Appendix G. In fact, Proposition G.4 shows that there are positive constants c_3, c_4 and c_5 such that for all $r \geq 0, \Lambda \geq 0$ and $\sigma > 0$ satisfying $c_3 r + c_4 \Lambda \leq c_5/\sigma^2$, there exists a continuous function $v : [0, 1] \rightarrow \mathbb{R}$ with the following properties: v is once continuously differentiable and satisfies $m < v < \bar{m}$ on $]0, 1[$; $v(0) = m(0)$ and $v(1) = m(1)$; and v solves (20) on $]0, 1[- \{\hat{\pi}\}$. By the uniqueness part of Theorem 4.1 (ii), we have $v = v^*$, so experimentation is extreme for these parameter combinations, and we can take $\kappa_e = c_5/c_3$ and $\gamma_e = c_5/c_4$. \blacksquare

Note that the constructive approach used in the proof (based on Propositions G.3 and G.4) yields explicit formulae for constants $\gamma_m, \kappa_m, \gamma_e$ and κ_e with the above properties. The proof also shows that under the stated condition for moderate experimentation, the adjusted value function is always differentiable at $\hat{\pi}$. In particular, a kink in v^* can only occur in an ‘intermediate’ range of parameter combinations.

Using the fact that v^* is strictly convex whenever $v^*(\hat{\pi}) = \hat{m}$, we can derive a more precise characterisation of the boundary between the parameter regions associated with moderate and extreme experimentation. Before formulating this result, we note from the ODE (20) that v^* depends on r, Λ and σ only through the two products $\kappa = r\sigma^2$ and $\gamma = \Lambda\sigma^2$. This reduces the parameter space effectively to the non-negative orthant \mathbb{R}_+^2 , which splits into a region of moderate experimentation, $K_m = \{(r\sigma^2, \Lambda\sigma^2) : (r, \Lambda, \sigma) \in \Theta_m\}$, and a region of extreme experimentation, $K_e = \{(r\sigma^2, \Lambda\sigma^2) : (r, \Lambda, \sigma) \in \Theta_e\}$. As the following result shows, the boundary between these two sets cuts each ray through the origin in a single point. Thus, we can ‘trace’ this boundary by varying the slope of the ray.

Proposition 4.4 *Let R be a ray in \mathbb{R}_+^2 emanating from the origin $(0, 0)$. Then, there is a unique point $(\kappa^\dagger, \gamma^\dagger) \in R - \{(0, 0)\}$ such that experimentation is extreme for all $(\kappa, \gamma) \in R$ with $(\kappa, \gamma) < (\kappa^\dagger, \gamma^\dagger)$, and moderate for all $(\kappa, \gamma) \in R$ with $(\kappa, \gamma) \geq (\kappa^\dagger, \gamma^\dagger)$.*

PROOF: Fixing a point $(\kappa_0, \gamma_0) \in R - \{(0, 0)\}$, we can parameterise the ray R by the mapping $\mu \mapsto (\mu\kappa_0, \mu\gamma_0)$ where $\mu \geq 0$. We write $v^*[\mu]$ for the adjusted value function associated with the parameters $(\mu\kappa_0, \mu\gamma_0)$. Define $M = \{\mu \geq 0 : v^*[\mu](\hat{\pi}) = \hat{m}\}$ and $\mu^\dagger = \inf M$. By Proposition 4.3, M is non-empty, and μ^\dagger is finite and positive. By Proposition F.2, $v^*[\mu]$ is strictly convex for all $\mu \in M$, so we can replicate the comparative statics argument in the proof of Theorem 5.2 below to conclude that if $\mu \in M$ and $\mu' > \mu$, then $\mu' \in M$. Combined with a standard continuity argument, this shows that $M = [\mu^\dagger, \infty[$. \blacksquare

We have the following two corollaries to the above proposition for the limiting cases where either Λ or r is zero. Given σ , we find thresholds $r^* > 0$ (when $\Lambda = 0$) and $\Lambda^* > 0$ (when $r = 0$) which separate extreme from moderate experimentation. This illustrates once more that a small change in the discount rate or in the variability of the environment can trigger a discontinuous qualitative change in optimal behaviour.

Corollary 4.2 *Let $\Lambda = 0$. Then there exists a unique real number $\kappa^* > 0$ such that optimal experimentation is extreme if $r < \kappa^*/\sigma^2$, and moderate if $r \geq \kappa^*/\sigma^2$.*

PROOF: Apply Proposition 4.4 to the ray $\mathbb{R}_+ \times \{0\}$ in \mathbb{R}_+^2 . ■

Corollary 4.3 *Let $r = 0$. Then there is a unique real number $\gamma^* > 0$ such that optimal experimentation is extreme if $\Lambda < \gamma^*/\sigma^2$, and moderate if $\Lambda \geq \gamma^*/\sigma^2$.*

PROOF: Apply Proposition 4.4 to the ray $\{0\} \times \mathbb{R}_+$ in \mathbb{R}_+^2 . ■

We mentioned above that $\kappa_m, \kappa_e, \gamma_m$ and γ_e as in Proposition 4.3 can be calculated explicitly. This yields explicit upper and lower bounds for each of κ^* and γ^* .

Thus the boundary between the two experimentation regimes links κ^* on the $r\sigma^2$ -axis with γ^* on the $\Lambda\sigma^2$ -axis, and lies between the lines $r/\kappa_e + \Lambda/\gamma_e = 1/\sigma^2$ and $r/\kappa_m + \Lambda/\gamma_m = 1/\sigma^2$. Furthermore, given any point $(\kappa^\dagger, \gamma^\dagger)$ on the boundary, the set of points $\{(\kappa, \gamma) : \kappa \geq \mu\kappa^\dagger, \gamma = \mu\gamma^\dagger, \mu \geq 1\}$ lies within K_m (cf. the comparative statics results in Theorem 5.2 below).

Note that within the L-shaped region $\{(\kappa, \gamma) : \kappa < \kappa^* \text{ or } \gamma < \gamma^*\}$ there is a potential trade-off between the discount rate and the switching intensities, given a noise intensity σ : for any $r < \kappa^*/\sigma^2$, moderate experimentation can be avoided by Λ being sufficiently low, and similarly, for any $\Lambda < \gamma^*/\sigma^2$, moderate experimentation can be avoided by r being sufficiently low.

4.3 No State Switching ($r > 0, \Lambda = 0$)

The discounted case with $\Lambda = 0$ is simple since we do not have to make the transformation from u^* to v^* , but can work with u^* itself. Our next result provides a detailed characterisation of the optimal experimentation behaviour.

Proposition 4.5 *Let $\Lambda = 0$.*

In the case of extreme experimentation, the optimal policy is continuous except for a single jump between q_{\max} and q_{\min} , and the process of posterior beliefs, if started from a prior belief in $]0, 1[$, converges to the true state of demand with probability one.

In the case of moderate experimentation, the optimal policy is continuous, and the process of posterior beliefs exhibits the following long-run behaviour given a prior belief $\pi_0 \neq 0, \hat{\pi}, 1$ and the true state of demand k : with probability one, beliefs converge to

- $\hat{\pi}$ if $0 < \pi_0 < \hat{\pi}$ and $k = 1$, or $\hat{\pi} < \pi_0 < 1$ and $k = 0$;
- 0 or $\hat{\pi}$ if $0 < \pi_0 < \hat{\pi}$ and $k = 0$;

- $\hat{\pi}$ or 1 if $\hat{\pi} < \pi_0 < 1$ and $k = 1$.

In particular, there is always a positive probability that beliefs will converge to $\hat{\pi}$.

PROOF: For $u^*(\hat{\pi}) > \hat{m}$, we obtain the optimal policy exactly as in the proof of Proposition 4.1. Since u^* is strictly convex, its graph crosses \mathcal{R}_c only once, so this policy has indeed just one jump.

If $u^*(\hat{\pi}) = \hat{m}$, on the other hand, we have $(u^*)'(\hat{\pi}) = 0$ and $\tau(\pi)(u^*)''(\pi)/2r = \beta(\pi)[u^*(\pi) - m(\pi)]/[u^*(\pi) - \hat{m}] < \beta(\pi)$ on $]0, 1[-\{\hat{\pi}\}$. For all $\pi \neq \hat{\pi}$, there are ξ and η strictly between $\hat{\pi}$ and π such that $[u^*(\pi) - \hat{m}]/[m(\pi) - \hat{m}] = (u^*)''(\xi)/m''(\zeta)$; as $(u^*)''$ is bounded on $]0, 1[-\{\hat{\pi}\}$, so is the quotient on the left-hand side. Therefore, $O(\pi, u^*(\pi))$ extends to a Lipschitz continuous policy function which is optimal by the verification theorem from Section 3.4.

The updating equation (4) shows that the process of posterior beliefs generated by the optimal policy is a supermartingale if the true state is $k = 0$, and a submartingale if $k = 1$. Since the process is bounded, this implies almost sure convergence. The stated long-run behaviour is now established by means of the standard boundary classification for diffusion processes; cf. Karlin and Taylor (1981, Chapter 15, Sections 6-7). ■

The proposition shows that for sufficiently high discount rates, there is a positive probability that beliefs will settle down at a point where the agent has not learnt the true state. This is a particular case of the general incomplete learning result obtained in the literature on optimal learning in an unchanging environment and referred to after Proposition 4.2. Our setup with continuous time and a one-dimensional state space makes this result particularly stark: if the prior belief lies on the ‘wrong’ side of $\hat{\pi}$, moderate experimentation will cause beliefs to converge to $\hat{\pi}$ with probability one! Here, exactly as in the case of a changing environment, the monopolist, when choosing to experiment moderately, effectively reduces the state space to a smaller interval. In this sense, a ‘local’ decision, how to experiment at or near the belief $\hat{\pi}$, has drastic ‘global’ consequences.

Finally, note that the case of a changing environment is richer in that it allows for a form of moderate experimentation where the monopolist experiments much more on one side of $\hat{\pi}$ than on the other. This asymmetry reflects the fact that state switching introduces mean reversion into the updating equation (2), thereby destroying the martingale property of beliefs; this does not happen when $\Lambda = 0$.

4.4 Maximum Experimentation ($r = 0$, $\Lambda = 0$)

For given σ , an agent who uses the catching-up criterion in an unchanging environment clearly has the strongest possible incentive to experiment. Consequently, if extreme experimentation is to occur at all, it must occur for $r = 0$ and $\Lambda = 0$. We have seen above that this is indeed the case. We now derive the optimal policy in a simple closed form. Recall the construction of the left, central and right rays in Section 3.5, and let q_c denote the centre of the interval $[q_{\min}, q_{\max}]$.

Proposition 4.6 *Let $r = 0$ and $\Lambda = 0$. Let π_ℓ , π_c and π_r be the beliefs where the graph of the full-information pay-off function \bar{m} intersects the left, central and right*

rays, respectively, that is,

$$\begin{aligned}\bar{m}(\pi_\ell) &= \hat{m} - \frac{1}{2} \Delta\alpha [q_{\max} - \hat{q}] (\pi_\ell - \hat{\pi}), \\ \bar{m}(\pi_r) &= \hat{m} + \frac{1}{2} \Delta\alpha [\hat{q} - q_{\min}] (\pi_r - \hat{\pi}),\end{aligned}$$

and

$$\pi_c = \hat{\pi} + \frac{2}{\Delta\alpha} \frac{q_c - \hat{q}}{[q_{\max} - \hat{q}][\hat{q} - q_{\min}]} (\bar{m}(\pi_c) - \hat{m}).$$

Then the policy function $\bar{q} : [0, 1] \rightarrow Q$ defined by

$$\bar{q}(\pi) = \begin{cases} q^m(\pi) + \frac{\bar{m}(\pi) - m(\pi)}{m(\pi) - \hat{m}} [q^m(\pi) - \hat{q}] & \text{for } 0 \leq \pi < \pi_\ell, \\ q_{\max} & \text{for } \pi_\ell \leq \pi < \pi_c, \\ q_{\max} \text{ or } q_{\min} & \text{for } \pi = \pi_c, \\ q_{\min} & \text{for } \pi_c < \pi \leq \pi_r, \\ q^m(\pi) + \frac{\bar{m}(\pi) - m(\pi)}{m(\pi) - \hat{m}} [q^m(\pi) - \hat{q}] & \text{for } \pi_r < \pi \leq 1, \end{cases}$$

which is continuous except for a jump at π_c , is optimal.

PROOF: \bar{q} is measurable with $[\bar{q}(\pi) - \hat{q}]^2$ bounded away from zero, hence admissible by Proposition A.1. Optimality follows from the undiscounted variant of the verification theorem from Section 3.4 and the fact that $\bar{q}(\pi) \in O(\pi, \bar{m}(\pi))$ for $0 < \pi < 1$. ■

Note that this result holds independently of the value of the parameter σ . The intuition for this is simple. In the absence of state switching, a change in σ amounts to a mere rescaling of the time axis. As the objective of an agent using the catching-up criterion is invariant to such a rescaling, the optimal policy remains the same.

4.5 Further Findings

Beyond the results reported above, our numerical simulations suggest additional properties of the adjusted value function and the optimal policy. We discuss them briefly in this subsection.

First, the adjusted value function appears to be strictly convex even in the extreme experimentation regime; see for example Figure 3. This implies in particular that the graph of v^* crosses the central ray only once, so extreme experimentation entails just one jump in the optimal policy function. Theorem 4.1 establishes strict convexity for moderate experimentation only. It would be desirable to have a general, economically intuitive proof of this property that does not rely on a detailed (and rather complicated) investigation of the ODE for v^* as in Appendix F. Given that the adjusted value function for $\Lambda > 0$ satisfies the same boundary conditions as the (true) value function for $\Lambda = 0$ (both coincide with the myopic pay-off for $\pi = 0$ and 1), one could for example try to reinterpret v^* as the value function of a transformed optimal experimentation

problem *without* state switching, in which case convexity would follow immediately by standard arguments. So far, we have not been able to make progress in this direction.

Second, granted strict convexity of v^* , we also have the strict inequality $v^* < \bar{m}$ on $]0, 1[$. This inequality reflects the intuitive fact that the incentive to experiment is highest for a monopolist who operates in an unchanging environment and does not discount. Again, a proof based on this intuition would be desirable.

Third, our numerical results suggest that v^* is strictly decreasing (on $]0, 1[$ for extreme experimentation, on $]0, 1[- \{\hat{\pi}\}$ for moderate experimentation) in each of the parameters r , Λ and σ . As a consequence, the extent to which the agent experiments, measured by the distance $|q^*(\pi) - q^m(\pi)|$, is strictly decreasing in each of these parameters as long as $q^*(\pi) \notin \{q_{\min}, q^m(1), \hat{q}, q^m(0), q_{\max}\}$. In the extreme experimentation regime, moreover, the set of beliefs at which q_{\max} or q_{\min} is optimal shrinks in response to an increase in one of the parameters, and is always a subset of $[\pi_\ell, \pi_r]$ as in Proposition 4.6. The intuition behind these comparative statics is clear. As the discount rate increases, the future becomes less important to the agent. The value of information falls, and with it the agent's willingness to sacrifice current revenue for potential future gains from experimentation. A higher level of noise, on the other hand, renders the price signal less informative, which reduces the expected gain from any given deviation from the myopic optimum, and thus the incentive to experiment. Finally, a higher frequency of state switches increases the risk of information becoming obsolete, so the trade-off between current revenues and potential gains from experimentation shifts again in favour of the former. For the moderate experimentation regime, we can use the techniques of Appendix G to prove analytically that $|q^*(\pi) - q^m(\pi)|$ is always strictly decreasing on $]0, 1[- \{\hat{\pi}\}$ in r and σ , and strictly decreasing in Λ in the undiscounted case ($r = 0$); cf. the proof of Theorem 5.2 below.

5 No Confounding Belief

We now turn briefly to the simpler case of optimal experimentation when the two demand curves do not intersect in the interior of Q^m , and so there is no longer a belief $\hat{\pi} \in]0, 1[$ such that $q^m(\hat{\pi}) = \hat{q}$.²¹ That is, we drop the standing assumption made in Section 3.3, and instead make the following

Assumption *The two demand curves do not intersect in the range of the myopic policy function ($\hat{q} \notin Q^m$), experimentation is informative ($\Delta\beta \neq 0$) and information is valuable ($\alpha_0\beta_1 - \alpha_1\beta_0 \neq 0$).*

We continue to assume without loss of generality that the demand curve in state 1 is steeper than the demand curve in state 0, that is, $\Delta\beta > 0$. As in Section 3.3, this new assumption is sufficient for experimentation to occur at almost all beliefs; see the footnote on page 11.

²¹To avoid cumbersome case distinctions, we will ignore the border-line cases $\hat{q} = q^m(0)$ and $\hat{q} = q^m(1)$ in what follows. It is easy to check, though, that the results given below remain valid in these two scenarios.

It turns out that under the new assumption, experimentation is always in *one* direction and *moderate*. That is, the monopolist either always reduces production relative to the myopic benchmark, or he always expands production, and he never chooses any quantities outside the range Q^m of the myopic policy function. In particular, the relevant ODE for the adjusted value function will be given by (21), and optimal quantities by (15).

When $r = 0$ and $\Lambda = 0$, the optimal policy can again be written down in closed form:

$$\bar{q}(\pi) = q^m(\pi) + \frac{\bar{m}(\pi) - m(\pi)}{m(\pi) - \hat{m}} [q^m(\pi) - \hat{q}];$$

since $\bar{q}(\pi) = O(\pi, \bar{m}(\pi))$ for $0 < \pi < 1$, optimality of this policy function follows exactly as in the proof of Proposition 4.6. When $r > 0$ or $\Lambda > 0$, we have the following theorem.

Theorem 5.1 (Moderate Experimentation, One Direction) *Let $r > 0$ or $\Lambda > 0$. Then the adjusted value function v^* satisfies $m < v^* < \bar{m}$ on $]0, 1[$, and it is analytic, strictly convex and the unique solution of the ODE (21) subject to the conditions $v^*(0) = m(0)$, $v^*(1) = m(1)$, and $v^* > m$ everywhere else. The optimal policy function,*

$$q^*(\pi) = q^m(\pi) + \frac{v^*(\pi) - m(\pi)}{m(\pi) - \hat{m}} [q^m(\pi) - \hat{q}],$$

is analytic, takes values in Q^m only, and satisfies the following inequalities on $]0, 1[$: $q^m < q^ < \bar{q}$ if $\hat{q} < Q^m$, and $\bar{q} < q^* < q^m$ if $\hat{q} > Q^m$.*

PROOF: Applying Proposition G.2, we obtain a continuous function $v : [0, 1] \rightarrow \mathbb{R}$ which solves (21) with $m < v < \bar{m}$ on $]0, 1[$. This function is analytic on $]0, 1[$ by the Cauchy-Kowalewski theorem.

Using the verification theorem from Section 3.6 (and its undiscounted variant), we see that $v = v^*$ and the policy q^* is optimal. This also establishes the uniqueness part of the theorem. Moreover, we have $(v^*)'' > 0$ on $]0, 1[$ by Proposition F.1.

The stated inequalities for q^* follow directly from the fact that $m < v^* < \bar{m}$ on $]0, 1[$. The derivative of the function \bar{q} is

$$\bar{q}'(\pi) = -\frac{[q^m(0) - \hat{q}][q^m(1) - \hat{q}]\Delta\beta\hat{p}}{2\beta(\pi)^2[q^m(\pi) - \hat{q}]^2} = \frac{[q^m(0) - \hat{q}][q^m(1) - \hat{q}]}{[q^m(\pi) - \hat{q}]^2} (q^m)'(\pi);$$

as either $\hat{q} < Q^m$ or $\hat{q} > Q^m$, this is either strictly positive or strictly negative throughout and of the same sign as the derivative of q^m . As \bar{q} and q^m coincide at either end of the unit interval, the range of \bar{q} and hence the range of q^* equals Q^m . ■

Thus, experimentation is indeed in one direction for $\hat{q} \notin Q^m$: if \hat{q} lies to the left of Q^m , then $q^* > q^m$ on $]0, 1[$, which means that the agent experiments by increasing quantity; if, on the other hand, \hat{q} lies to the right of Q^m , then $q^* < q^m$ on $]0, 1[$, so the agent experiments by decreasing quantity. The intuition behind this quantity expansion or reduction is straightforward: the monopolist deviates from the myopic quantity by

moving away from \hat{q} and in the (now unambiguously defined) direction of widening spreads between the two possible demand curves, thus making price observations more informative.²² Note also that the process of posterior beliefs is now always regular since the difference $q^*(\pi) - \hat{q}$, and hence the informativeness of the price signal, is bounded away from zero.

Convexity of v^* turns out to be crucial for the comparative statics of optimal experimentation, to which we turn next.

Theorem 5.2 *Given fixed demand curve parameters $\alpha_0, \alpha_1, \beta_0, \beta_1$ such that $\hat{q} \notin Q^m$ and a fixed $\tilde{\pi}$, the distance $|q^*(\pi) - q^m(\pi)|$ is*

- *strictly decreasing in r ;*
- *strictly decreasing in σ ;*
- *strictly decreasing in Λ if $r = 0$*

for all $\pi \in]0, 1[$.

PROOF: Let $v^*[r, \Lambda, \sigma]$ denote the adjusted value function for any given combination of parameters $r \geq 0, \Lambda \geq 0$ and $\sigma > 0$, and let $\text{ODE}[r, \Lambda, \sigma]$ be the differential equation (21) for these parameter values.

Consider discount rates $r_1 < r_2$. Since $v^*[r_1, \Lambda, \sigma] > m$ on $]0, 1[$, the right-hand side of (21) with $v = v^*[r_1, \Lambda, \sigma]$ is strictly increasing in r at each $\pi \in]0, 1[$. Being a *solution* of $\text{ODE}[r_1, \Lambda, \sigma]$, $v^*[r_1, \Lambda, \sigma]$ is thus a strict *supersolution* of $\text{ODE}[r_2, \Lambda, \sigma]$. (See Appendix G for a definition of supersolution.) As in Proposition G.2, therefore, there exists a continuous function $v : [0, 1] \rightarrow \mathbb{R}$ which solves $\text{ODE}[r_2, \Lambda, \sigma]$ with $m < v < v^*[r_1, \Lambda, \sigma]$ on $]0, 1[$. As $v^*[r_1, \Lambda, \sigma] < \bar{m}$ on $]0, 1[$, the uniqueness part of Theorem 5.1 implies $v = v^*[r_2, \Lambda, \sigma]$. The comparative statics result with respect to the discount rate now follows from the observation that optimal quantities are increasing in adjusted values.

The other comparative statics results follow in the same way. In fact, since the second derivative of the adjusted value function is strictly positive on $]0, 1[$, $v^*[r, \Lambda, \sigma_1]$ is a strict supersolution of $\text{ODE}[r, \Lambda, \sigma_2]$ for $\sigma_1 < \sigma_2$, and $v^*[0, \Lambda_1, \sigma]$ is a strict supersolution of $\text{ODE}[0, \Lambda_2, \sigma]$ for $\Lambda_1 < \Lambda_2$. ■

The intuition behind these comparative statics results has already been discussed in Section 4.5.²³

²²It is less intuitive, though, that experimentation should always be moderate.

²³Clearly, the comparative statics with respect to Λ should also pertain when $r > 0$. Our numerical simulations confirm this conjecture, but we have not been able to provide an analytical proof so far. Note that by the same argument as in the proof of Theorem 5.2, it would be sufficient to show that

$$\left[f(\pi) + (\pi - \tilde{\pi}) \frac{\Delta\beta}{\beta(\pi)} \right] \frac{v^*(\pi) - m(\pi)}{v^*(\pi) - \hat{m}} + (\pi - \tilde{\pi}) \left[\frac{v^*(\pi) - m(\pi)}{v^*(\pi) - \hat{m}} \right]' > 0$$

on $]0, 1[$, which is equivalent to $(v^*)'' > (u^*)''$ on the open unit interval. All the numerical solutions that we have calculated satisfy this condition.

6 Conclusion

We have studied the behaviour of a monopolist who learns about randomly changing demand by choosing a stream of quantities, observing the prices they generate, and updating his beliefs accordingly. As the action space is continuous, a small amount of information can be obtained at a small opportunity cost. Given the changing environment, therefore, experimentation will occur even in the long run although, as we have seen, the scope of actions may become restricted.

We formulated the problem in continuous time, which lead us via the Bellman equation to an ordinary differential equation for the adjusted value function. The advantages of this approach are three-fold: (a) key properties of the value function and optimal policy can be established analytically, as can some comparative statics results, even though a closed-form solution is generally not obtainable; (b) the sample path properties of beliefs and optimal actions are easy to characterise; (c) it is straightforward to solve the differential equation of interest numerically, enabling us to illustrate the analytical results and suggest further plausible properties of the solution.

Our analysis focused on the most interesting case where the confounding quantity lies between the myopically optimal quantities for the two possible demand curves. We found two qualitatively different experimentation regimes. For low discount rates and low probabilities of a demand curve switch, optimal experimentation is extreme: the maximal and minimal feasible quantities are chosen a non-negligible fraction of time; the optimal policy function exhibits a jump from one extreme quantity to the other; and the true state is tracked fairly well. For high discount rates or high probabilities of a demand curve switch, on the other hand, experimentation is moderate: the quantities chosen are bounded away from the extremes; the monopolist behaves like a myopic agent at the confounding belief; and he eventually restricts his choices to a subset of the space of feasible quantities.

A transition from one regime to the other in response to a change in the model parameters involves a discontinuous change of optimal policy. This suggests that agents in a changing environment may reduce their investment in information drastically if the frequency of change (or the interest rate) passes a critical threshold. However, there is a region in which a trade-off between interest rates and stability can be exploited: a moderately high interest rate need not trigger sluggish investment provided that the underlying environment is sufficiently stable; conversely, the low-investment effect of a changing environment can be overcome by a sufficiently low interest rate.

As to the robustness of our results, it is clear that the linearity of demand curves is inessential: any pair of demand curves with a confounding quantity and a price spread that increases monotonically as we move further away from the confounding quantity (at least locally) would lead to the same conclusions. (Also, we could have used price rather than quantity as the choice variable with the appropriate specification of noise.)

The fact that we allow only two possible demand curves turns out to be more restrictive. While the discontinuous dependence of optimal policies on the discount and switching rates does not hinge on this assumption,²⁴ the moderate experimentation

²⁴If we consider three linear demand curves, for example, the space of possible posterior beliefs is

trap can only occur with three or more demand curves if there is again a confounding quantity, i.e. if they have a common intersection point. In this sense, the trap is not generic – but then neither are the incomplete learning results of the previous literature on unchanging environments. The case of two reward functions is worth investigating since it identifies the circumstances which produce an experimentation trap, and shows that the mere introduction of state switching is not enough to root out incomplete learning.

The continuous-time assumption is less crucial. The behaviour of posterior beliefs under moderate experimentation will be less clear-cut in discrete time, since discrete adjustments could allow the belief to jump back and forth past the confounding belief. However, adjustments towards the long-run average state become stronger and more frequent the further the current belief is away from the long-run average state, so excursions out of the trap can be expected to be infrequent and short. As we let the period length shrink, the resulting sample path behaviour will then become very close to that in our model.

Appendix

A Admissible Strategies and Policy Functions

We first provide a precise definition of the set \mathcal{Q} of admissible strategies. Assume that the Brownian motion Z and the Markov process k are given on some complete probability space and are both adapted to the filtration $\{\mathcal{F}_t\}$. Let \mathcal{Q}_0 denote the set of all processes $\mathbf{q} = \{q_t\}$ which take values in Q , the interval of feasible quantities, and are adapted to the aforementioned filtration. Each $\mathbf{q} \in \mathcal{Q}_0$ gives rise to a unique cumulative price process $P^{\mathbf{q}}$. The information contained in prices is summarised by $\{\mathcal{F}_t^{\mathbf{q}}\}$, the filtration generated by $P^{\mathbf{q}}$. A process $\mathbf{q} \in \mathcal{Q}_0$ is an *admissible strategy* if q_t is adapted to the filtration $\{\mathcal{F}_t^{\mathbf{q}}\}$.

Admissible policy functions can now be defined as follows. The function $q : [0, 1] \rightarrow Q$ is an *admissible policy function* if for any given initial belief π_0 , there is a unique strategy $\mathbf{q} \in \mathcal{Q}$ (with associated process of beliefs $\{\pi_t\}$) such that $q_t = q(\pi_t)$ for all t .

The following result provides conditions under which a given policy function is admissible.

Proposition A.1 *A policy function $q : [0, 1] \rightarrow Q$ is admissible if at least one of the following conditions holds:*

- (a) *q is Lipschitz continuous.*
- (b) *q is measurable, and there exists a $\delta > 0$ such that $[\Delta\alpha - \Delta\beta q(\pi)]^2 > \delta$ for all π .*

PROOF: Suppose that (a) holds. Then an extension to a standard existence theorem implies that the stochastic differential equation

$$d\pi_t = \left\{ \lambda(\pi_t) + \sigma^{-1} \Sigma(\pi_t, q(\pi_t)) \left(\alpha_{k_t} - \beta_{k_t} q(\pi_t) - [\alpha(\pi_t) - \beta(\pi_t) q(\pi_t)] \right) \right\} dt + \Sigma(\pi_t, q(\pi_t)) dZ_t \tag{A.1}$$

which is obtained from combining (2) and (3) has a unique solution π for any given starting value $\pi_0 \in [0, 1]$; cf. Liptser and Shirayayev (1977, p.330).²⁵ Define the strategy \mathbf{q} by $q_t = q(\pi_t)$ and consider

the two-dimensional simplex; along its edges, we are back in a scenario with two demand curves – so we have the discontinuity.

²⁵This is in fact a strong solution. A weak solution would be enough for our purposes.

the associated price process $dP_t = (\alpha_{k_t} - \beta_{k_t} q_t) dt + \sigma dZ_t$. Section 2 implies that the corresponding process of beliefs $\pi_t^{\mathbf{q}} = \mathbb{E}[k_t | \mathcal{F}_t^{\mathbf{q}}]$ also solves (A.1) with initial value π_0 . By the uniqueness part of Liptser and Shiriyayev (1977, Theorem 9.2), the processes π and $\pi^{\mathbf{q}}$ coincide, so π is indeed the process of beliefs associated with the strategy \mathbf{q} . The policy function q thus generates a unique strategy in \mathcal{Q} .

Now suppose (b). Given any initial value π_0 , Krylov (1980, Theorem 2.6.1) implies that the stochastic differential equation

$$d\pi_t = \lambda(\pi_t) dt + \Sigma(\pi_t, q(\pi_t)) dZ_t$$

has a weak solution (π, Z^0) with Z^0 a Wiener process. We extend the corresponding filtered probability space in such a way that it supports an independent Markov process $\{k_t\}$ taking values in $\{0, 1\}$ with transition probabilities as in Section 1. Consider the bounded process

$$\eta_t = \sigma^{-1} \left(\alpha_{k_t} - \beta_{k_t} q(\pi_t) - [\alpha(\pi_t) - \beta(\pi_t) q(\pi_t)] \right).$$

By Girsanov's theorem, there is a new measure under which

$$Z_t = Z_t^0 - \int_0^t \eta_s ds$$

is a Wiener process; cf. Revuz and Yor (1991). In other words, (π, Z) is a weak solution to the stochastic differential equation (A.1). Admissibility of the policy function q is now shown in exactly the same way as in the first part of this proof. \blacksquare

B Some Properties of the Value Function

Consider the value function u^* as defined in (5).

Proposition B.1 *The value function u^* is continuous and convex.*

PROOF: For fixed $\mathbf{q} \in \mathcal{Q}$, $u^{\mathbf{q}}$ is linear in π . Indeed,

$$\begin{aligned} u^{\mathbf{q}}(\pi) &= \pi \mathbb{E}_{k_0=1} \left[\int_0^\infty r e^{-rt} q_t [\alpha_{k_t} - \beta_{k_t} q_t] dt \right] \\ &\quad + (1 - \pi) \mathbb{E}_{k_0=0} \left[\int_0^\infty r e^{-rt} q_t [\alpha_{k_t} - \beta_{k_t} q_t] dt \right]. \end{aligned}$$

For $\pi = \eta \pi_1 + (1 - \eta) \pi_2$ with $0 \leq \eta \leq 1$, we therefore have

$$\begin{aligned} u^{\mathbf{q}}(\pi) &= \eta u^{\mathbf{q}}(\pi_1) + (1 - \eta) u^{\mathbf{q}}(\pi_2) \\ &\leq \eta u^*(\pi_1) + (1 - \eta) u^*(\pi_2) \end{aligned}$$

by the definition of the value function. Taking the supremum on the left-hand side proves convexity. A convex function is continuous on the interior of its domain, so we only have to show continuity at $\pi = 0$ and $\pi = 1$. Suppose for example that the value function is not continuous at $\pi = 0$. Due to convexity, this can only mean $u^*(0) > u^*(0+)$. By definition of the value function, there exists a policy $\mathbf{q} \in \mathcal{Q}$ such that $u^{\mathbf{q}}(0) > u^*(0+)$. But then $u^{\mathbf{q}}(\pi) > u^*(\pi)$ for small $\pi > 0$, which is a contradiction. The right boundary $\pi = 1$ is dealt with in the same way. \blacksquare

Convexity implies the existence of a left-hand derivative $D_- u^*$ on $]0, 1[$ and a right-hand derivative $D_+ u^*$ on $[0, 1[$, both being non-decreasing functions, the former left-continuous, the latter right-continuous, with $D_- u^* \leq D_+ u^*$ on their common domain.

Lemma B.1 *The one-sided derivatives $D_- u^*$ and $D_+ u^*$ are bounded.*

PROOF: We see from the representation of the pay-off function $u^{\mathbf{q}}$ in the previous proof that there is a constant $K > 0$ such that $|(u^{\mathbf{q}})'(\pi)| \leq K$ for all $\mathbf{q} \in \mathcal{Q}$ and all π . Now, suppose that $(D_- u^*)(\pi_1) < -K$ for some belief $\pi_1 > 0$. Then there is a $\pi_2 < \pi_1$ such that $u^*(\pi_1) - u^*(\pi_2) < -K(\pi_1 - \pi_2)$, i.e., $u^*(\pi_2) > u^*(\pi_1) + K(\pi_1 - \pi_2)$. By definition of the value function, we can find a strategy $\mathbf{q} \in \mathcal{Q}$ with $u^*(\pi_2) \geq u^{\mathbf{q}}(\pi_2) > u^*(\pi_1) + K(\pi_1 - \pi_2)$. But then the linearity of $u^{\mathbf{q}}$ implies $u^{\mathbf{q}}(\pi_1) \geq u^{\mathbf{q}}(\pi_2) - K(\pi_1 - \pi_2) > u^*(\pi_1)$, which is a contradiction. Using a similar argument for the right-hand derivative, we obtain $-K \leq D_- u^* \leq D_+ u^* \leq K$ on $]0, 1[$. Due to left- and right-continuity, respectively, this also proves that $(D_- u^*)(1)$ and $(D_+ u^*)(0)$ are bounded in absolute value by K . ■

C The Value Function as a Solution of the Bellman Equation

Proposition C.1 *The value function u^* has a continuous first derivative on $[0, 1]$, and possesses a locally bounded generalised second derivative $u_2^* \geq 0$ such that*

$$(u^*)'(\pi_2) - (u^*)'(\pi_1) = \int_{\pi_1}^{\pi_2} u_2^*(\pi) d\pi \quad (\text{C.1})$$

for all π_1 and π_2 . Moreover,

$$\max_{q \in \mathcal{Q}} \left\{ \frac{1}{2} \Sigma^2(\pi, q) u_2^*(\pi) + \lambda(\pi) (u^*)'(\pi) - r u^*(\pi) + r R(\pi, q) \right\} = 0 \quad (\text{C.2})$$

almost everywhere on $]0, 1[$.

PROOF: Krylov (1980, Theorem 6, p.289) implies that u^* has two locally bounded generalised derivatives, u_1^* and u_2^* . By definition, this means that

$$\int_0^1 \phi(\pi) u_1^*(\pi) d\pi = - \int_0^1 \phi'(\pi) u^*(\pi) d\pi$$

and

$$\int_0^1 \phi(\pi) u_2^*(\pi) d\pi = \int_0^1 \phi''(\pi) u^*(\pi) d\pi$$

for all functions ϕ that are infinitely differentiable and of compact support in $]0, 1[$. On the other hand, u^* is convex by Proposition B.1. As its left-hand derivative $D_- u^*$ is left-continuous and non-decreasing, one can define a measure μ on $]0, 1[$ via $\mu[\pi_1, \pi_2[= (D_- u^*)(\pi_2) - (D_- u^*)(\pi_1)$. This measure represents the second derivative of u^* in the sense of a distribution:

$$\int_0^1 \phi''(\pi) u^*(\pi) d\pi = \int_0^1 \phi(\pi) d\mu(\pi)$$

for every function ϕ that is infinitely differentiable and of compact support in $]0, 1[$. Moreover, this property characterises μ uniquely; cf. Krylov (1980, p.49). Comparing it with the definition of the generalised second derivative u_2^* , we conclude that $d\mu = u_2^* d\pi$. In particular,

$$(D_- u^*)(\pi_2) - (D_- u^*)(\pi_1) = \int_{\pi_1}^{\pi_2} u_2^*(\pi) d\pi$$

for all $\pi_1, \pi_2 \in]0, 1[$. This implies that $D_- u^*$ is continuous, so u^* is continuously differentiable on the open unit interval with $(u^*)' = D_- u^*$. By Proposition B.1, $(u^*)'(\pi)$ has a continuous extension to the whole of $[0, 1]$.

As to the last part of the proposition, Krylov (1980, Theorem 6, p.289) implies that

$$\max_{q \in \mathcal{Q}} \left\{ \frac{1}{2} \Sigma^2(\pi, q) u_2^*(\pi) + \lambda(\pi) u_1^*(\pi) - r u^*(\pi) + r R(\pi, q) \right\} = 0$$

holds almost everywhere on $]0, 1[$. The proof is completed by replacing u_1^* with $(u^*)'$. ■

The representation (C.1) implies

Corollary C.1 u^* is almost everywhere twice differentiable, and $(u^*)'' = u_2^*$ almost everywhere. Moreover, u^* is twice continuously differentiable on any open set where u_2^* has a continuous version, i.e., coincides with a continuous function almost everywhere.

Applying (C.2) with $q = q^m(\pi)$ and dividing through by r , we see immediately that

$$u^*(\pi) - \lambda(\pi) \frac{(u^*)'(\pi)}{r} \geq m(\pi)$$

almost everywhere. By continuity of u^* and $(u^*)'$, we can conclude that this inequality holds in fact on the whole of $[0, 1]$. As to the boundary of the unit interval, we have the following result.

Corollary C.2 The value function satisfies the boundary conditions

$$\begin{aligned} u^*(0) - \lambda(0) \frac{(u^*)'(0)}{r} &= m(0), \\ u^*(1) - \lambda(1) \frac{(u^*)'(1)}{r} &= m(1). \end{aligned}$$

PROOF: We first note that (C.2) implies

$$\frac{1}{2} \max_{q \in Q} \Sigma^2(\pi, q) u_2^*(\pi) + \lambda(\pi) (u^*)'(\pi) - r u^*(\pi) + r m(\pi) \geq 0$$

and hence

$$u_2^*(\pi) \geq \frac{2r\sigma^2}{\pi^2(1-\pi)^2} \frac{u^*(\pi) - \lambda(\pi)(u^*)'(\pi)/r - m(\pi)}{\max_{q \in Q} [\Delta\alpha - \Delta\beta q]^2}$$

for almost all π . Now suppose that $u^*(0) - \lambda(0)(u^*)'(0)/r > m(0)$. Using the continuity of $u^*(\pi) - \lambda(\pi)(u^*)'(\pi)/r$ and the inequality just derived, we can find $K > 0$ and $\epsilon > 0$ such that $u_2^*(\pi) \geq K\pi^{-2}$ almost everywhere on $[0, \epsilon]$. But then

$$(u^*)'(\pi) = (u^*)'(\epsilon) - \int_{\pi}^{\epsilon} u_2^*(\xi) d\xi \leq (u^*)'(\epsilon) - K \int_{\pi}^{\epsilon} \frac{d\xi}{\xi^2} = (u^*)'(\epsilon) - K \left[\frac{1}{\pi} - \frac{1}{\epsilon} \right] \longrightarrow -\infty$$

as $\pi \rightarrow 0$, which contradicts the boundedness of $(u^*)'$. The boundary condition at $\pi = 1$ follows by the same argument. ■

The next result is a so-called *verification theorem*, providing sufficient conditions for a given solution of the Bellman equation to be the value function, and for a given policy function to be optimal or ϵ -optimal.

Proposition C.2 Let u be a once continuously differentiable function on $[0, 1]$ with a generalised second derivative $u_2 \geq 0$ on $]0, 1[$ such that

$$u'(\pi_2) - u'(\pi_1) = \int_{\pi_1}^{\pi_2} u_2(\pi) d\pi$$

for all π_1 and π_2 , and $\pi^2(1-\pi)^2 u_2(\pi) \rightarrow 0$ as $\pi \rightarrow 0$ and $\pi \rightarrow 1$, respectively. If

$$\max_{q \in Q} \left\{ \frac{1}{2} \Sigma^2(\pi, q) u_2(\pi) + \lambda(\pi) u'(\pi) - r u(\pi) + r R(\pi, q) \right\} = 0$$

on $]0, 1[$, then the following statements hold true:

- (a) $u(\pi) \geq u^{\mathbf{q}}(\pi)$ for all $\mathbf{q} \in Q$ and all π , that is, $u \geq u^*$.
- (b) Let $\epsilon > 0$. If $q : [0, 1] \rightarrow Q$ is an admissible policy function satisfying

$$\frac{1}{2} \Sigma^2(\pi, q(\pi)) u_2(\pi) + \lambda(\pi) u'(\pi) - r u(\pi) + r R(\pi, q(\pi)) \geq -\epsilon r \tag{C.3}$$

for all π , then the strategy \mathbf{q}_π obtained by following this policy from any given initial belief π is ϵ -optimal, i.e., $u^{\mathbf{q}_\pi}(\pi) \geq u(\pi) - \epsilon$. In particular, $u^* \geq u - \epsilon$.

(c) If there is an admissible policy function as in (b) for any $\epsilon > 0$, then u is the value function: $u = u^*$.

(d) If $q^* : [0, 1] \rightarrow Q$ is an admissible policy function such that for every π , the quantity $q^*(\pi)$ attains the supremum in (C.2), then this policy function is optimal. For any π ,

$$u(\pi) = u^*(\pi) = \max_{\mathbf{q} \in \mathcal{Q}} u^{\mathbf{q}}(\pi) = u^{\mathbf{q}^*}(\pi)$$

where \mathbf{q}^* is the strategy obtained by following this policy from the initial belief π .

PROOF: Let the initial belief be $\pi_0 = \pi$. For an arbitrary strategy $\mathbf{q} \in \mathcal{Q}$ consider the stochastic process $M^{\mathbf{q}}$ given by

$$M_T^{\mathbf{q}} = \int_0^T r e^{-rt} R(\pi_t, q_t) dt + e^{-rT} u(\pi_T).$$

By a generalisation of Itô's lemma,

$$\begin{aligned} M_T^{\mathbf{q}} &= M_0^{\mathbf{q}} \\ &+ \int_0^T e^{-rt} \left\{ \frac{1}{2} \Sigma^2(\pi_t, q_t) u_2(\pi_t) + \lambda(\pi_t) u'(\pi_t) - r u(\pi_t) + r R(\pi_t, q_t) \right\} dt \\ &+ \sigma^{-1} \int_0^T e^{-rt} \pi_t (1 - \pi_t) (\Delta\alpha - \Delta\beta q_t) dZ_t^{\mathbf{q}}; \end{aligned}$$

cf. Rogers and Williams (1987, Lemma IV.45.9, p.105). Now, (C.2) implies that the expression under the first integral is non-positive, so $M^{\mathbf{q}}$ is a supermartingale. In other words, $E_{\pi}[M_T^{\mathbf{q}}] \leq M_0^{\mathbf{q}}$ or

$$u(\pi) \geq E_{\pi} \left[\int_0^T r e^{-rt} R(\pi_t, q_t) dt \right] + e^{-rT} E_{\pi}[u(\pi_T)].$$

Letting T go to infinity, we see that the first term on the right hand side becomes $u^{\mathbf{q}}(\pi)$, while the second term tends to zero. This proves part (a). Next, let $\epsilon \geq 0$, and consider a policy function $q : [0, 1] \rightarrow Q$ satisfying (C.3) on the whole of its domain. If \mathbf{q} is the strategy obtained by following this policy from the initial belief π , then

$$E_{\pi}[M_T^{\mathbf{q}}] \geq M_0^{\mathbf{q}} - \epsilon \int_0^T r e^{-rt} dt.$$

Letting $T \rightarrow \infty$ yields $u^{\mathbf{q}}(\pi) \geq u(\pi) - \epsilon$. Parts (b), (c) and (d) follow immediately. ■

D Analysing the Bellman Equation

In Section 3.4 of the main text we initially rewrote the Bellman equation in the form

$$v = \max_{q \in Q} \{ \tau(\pi) s [q - \hat{q}]^2 + R(\pi, q) \}$$

where the variable v is standing in for $u(\pi) - \lambda(\pi)u'(\pi)/r$ and s is representing $u''(\pi)/2r$. The first task here is to show that this problem can be reformulated as

$$\tau(\pi) s = \min_{q \in Q - \{\hat{q}\}} \frac{v - R(\pi, q)}{[q - \hat{q}]^2}$$

for triplets (π, v, s) with $s \geq 0$ and (π, v) lying in the set

$$\mathcal{A} = \{(\pi, v) \in]0, 1[\times \mathbb{R} : v \geq m(\pi) \text{ and } v > \hat{m}\}.$$

(As noted in the main text, the condition $v > \hat{m}$ only bites if \hat{q} lies in the interior of Q^m , in which case it rules out exactly the point $(\hat{\pi}, \hat{m})$, which in turn excludes the possibility of \hat{q} being optimal.)

To derive the reformulation, define the functions

$$B[\pi, v, s, q] = \tau(\pi) s [q - \hat{q}]^2 + R(\pi, q) - v$$

and

$$B^*[\pi, v, s] = \max_{q \in Q} B[\pi, v, s, q],$$

and rewrite the Bellman equation as

$$B^*[\pi, v, s] = 0.$$

Then, for all triplets (π, v, s) with $s \geq 0$ and $(\pi, v) \in \mathcal{A}$, the equation $B^*[\pi, v, s] = 0$ is equivalent to $\max_{q \in Q - \{\hat{q}\}} B[\pi, v, s, q] = 0$. Now, we have $[q - \hat{q}]^2 > 0$ on $Q - \{\hat{q}\}$, so $\max_{q \in Q - \{\hat{q}\}} B[\pi, v, s, q] = 0$ if and only if

$$\max_{q \in Q - \{\hat{q}\}} \frac{B[\pi, v, s, q]}{[q - \hat{q}]^2} = 0,$$

which in turn is equivalent to

$$\tau(\pi) s = \min_{q \in Q - \{\hat{q}\}} \frac{v - R(\pi, q)}{[q - \hat{q}]^2}. \quad (\text{D.1})$$

Moreover, a quantity $q^* \in Q - \{\hat{q}\}$ satisfies $B[\pi, v, s, q^*] = B^*[\pi, v, s] = 0$ if and only if (D.1) holds and q^* minimises $[v - R(\pi, q)]/[q - \hat{q}]^2$. Thus, the original problem and its reformulation are equivalent on \mathcal{A} .

Next we shall prove the claim made in Section 3.5 that, when \hat{q} lies in the interior of the interval Q^m , the area \mathcal{A} can be divided into four sub-regions by the following rays emanating from $(\hat{\pi}, \hat{m})$:

$$\begin{aligned} \mathcal{R}_\ell &= \{(\pi, v) \in \mathcal{A} : v = \hat{m} - \frac{1}{2} \Delta\alpha [q_{\max} - \hat{q}] (\pi - \hat{\pi})\}, \\ \mathcal{R}_r &= \{(\pi, v) \in \mathcal{A} : v = \hat{m} + \frac{1}{2} \Delta\alpha [\hat{q} - q_{\min}] (\pi - \hat{\pi})\}, \\ \mathcal{R}_c &= \left\{ (\pi, v) \in \mathcal{A} : \pi = \hat{\pi} + \frac{2}{\Delta\alpha} \frac{q_c - \hat{q}}{[q_{\max} - \hat{q}] [\hat{q} - q_{\min}]} (v - \hat{m}) \right\}, \end{aligned}$$

the sub-regions being associated with cases in which the above optimisation problems have interior or corner solutions.

Before proceeding, we use equation (9) to replace $R(\pi, q)$ in the optimisation problems to obtain

$$v - m(\pi) = \max_{q \in Q} \{ \tau(\pi) s [q - \hat{q}]^2 - \beta(\pi) [q - q^m(\pi)]^2 \} \quad (\text{D.2})$$

and its reformulation

$$\tau(\pi) s = \min_{q \in Q - \{\hat{q}\}} \left\{ \frac{\beta(\pi) [q - q^m(\pi)]^2 + v(\pi) - m(\pi)}{[q - \hat{q}]^2} \right\}. \quad (\text{D.3})$$

We first provide a preliminary lemma showing the regions of \mathcal{A} where the appropriate second order condition for the above equivalent problems is satisfied. Note that the relationships $m(\pi) = \hat{m} + \beta(\pi) [q^m(\pi) - \hat{q}]^2$ and $\beta(\pi) [q^m(\pi) - \hat{q}] = -\frac{1}{2} \Delta\alpha (\pi - \hat{\pi})$ are used in a number of the algebraic manipulations.

Lemma D.1 *Let \hat{q} lie in the interior of Q^m . For $(\pi, v, s) \in \mathcal{A} \times \mathbb{R}_+$, the second order condition for the minimisation problem in (D.3) is satisfied if and only if (π, v) lies below $\mathcal{R}_{2\ell}$ or below \mathcal{R}_{2r} , where*

$$\mathcal{R}_{2\ell} = \{(\pi, v) \in \mathcal{A} : v = \hat{m} - \Delta\alpha [q_{\max} - \hat{q}] (\pi - \hat{\pi})\}$$

and

$$\mathcal{R}_{2r} = \{(\pi, v) \in \mathcal{A} : v = \hat{m} + \Delta\alpha [\hat{q} - q_{\min}] (\pi - \hat{\pi})\}.$$

PROOF: The second order condition for the minimisation problem in (D.3) is satisfied wherever the second order condition for the maximisation problem in (D.2) is satisfied, and it is clear from (D.2) that the latter holds if and only if $\tau(\pi) s - \beta(\pi) < 0$.

Using the inequality $v < \hat{m} - \Delta\alpha [q_{\max} - \hat{q}] (\pi - \hat{\pi})$ in (D.2) leads, after some manipulation, to

$$\max_{q \in Q} \{ [\tau(\pi) s - \beta(\pi)] [q - \hat{q}]^2 - 2\beta(\pi) [q^m(\pi) - \hat{q}] [q_{\max} - q] \} < 0.$$

Evaluating the maximand at $q = q_{\max}$ gives us $[\tau(\pi) s - \beta(\pi)] [q_{\max} - \hat{q}]^2 < 0$ and so $\tau(\pi) s - \beta(\pi) < 0$.

On the other hand, using the inequality $v \geq \hat{m} - \Delta\alpha [q_{\max} - \hat{q}] (\pi - \hat{\pi})$ for $\pi \leq \hat{\pi}$ we arrive at

$$\max_{q \in Q} \{ [\tau(\pi) s - \beta(\pi)] [q - \hat{q}]^2 - 2\beta(\pi) [q^m(\pi) - \hat{q}] [q_{\max} - q] \} \geq 0.$$

The term $2\beta(\pi) [q^m(\pi) - \hat{q}] [q_{\max} - q]$ is non-negative for $\pi \leq \hat{\pi}$ so in this case $\tau(\pi) s - \beta(\pi) \geq 0$.

This proves the assertion concerning $\mathcal{R}_{2\ell}$. The case for \mathcal{R}_{2r} is proved in the same way simply by replacing q_{\max} by q_{\min} and $\pi \leq \hat{\pi}$ by $\pi \geq \hat{\pi}$. \blacksquare

The next lemma shows that the above optimisation problems have an interior solution if and only if (π, v) lies below \mathcal{R}_ℓ or below \mathcal{R}_r .

Lemma D.2 *Let \hat{q} lie in the interior of Q^m . For $(\pi, v, s) \in \mathcal{A} \times \mathbb{R}_+$, the minimisation problem in (D.3) has an interior solution if and only if $(\pi, v) \in \mathcal{A}_{\text{int},\ell} \cup \mathcal{A}_{\text{int},r}$, where*

$$\mathcal{A}_{\text{int},\ell} = \{ (\pi, v) \in \mathcal{A} : v < \hat{m} - \frac{1}{2} \Delta\alpha [q_{\max} - \hat{q}] (\pi - \hat{\pi}) \}$$

and

$$\mathcal{A}_{\text{int},r} = \{ (\pi, v) \in \mathcal{A} : v < \hat{m} + \frac{1}{2} \Delta\alpha [\hat{q} - q_{\min}] (\pi - \hat{\pi}) \}.$$

Moreover, the minimising quantity is given by

$$q^m(\pi) + \frac{v(\pi) - m(\pi)}{m(\pi) - \hat{m}} [q^m(\pi) - \hat{q}]$$

and the corresponding minimum is

$$\beta(\pi) \frac{v(\pi) - m(\pi)}{v(\pi) - \hat{m}}.$$

PROOF: In light of the preceding lemma, the minimisation problem in (D.3) has an interior solution if and only if the first order condition is satisfied when (π, v) lies below $\mathcal{R}_{2\ell}$ or below \mathcal{R}_{2r} . Note that $\mathcal{A}_{\text{int},\ell}$ lies below $\mathcal{R}_{2\ell}$ and $\mathcal{A}_{\text{int},r}$ lies below \mathcal{R}_{2r} .

The first order condition for the minimisation problem in (D.3) is satisfied by the quantity

$$q = q^m(\pi) + \frac{v - m(\pi)}{\beta(\pi) [q^m(\pi) - \hat{q}]} . \tag{D.4}$$

In the borderline cases, this first order condition holds for $q = q_{\max}$ or q_{\min} . With q^\dagger denoting either q_{\max} or q_{\min} this can be characterised by

$$q^\dagger = q^m(\pi) + \frac{v - m(\pi)}{\beta(\pi) [q^m(\pi) - \hat{q}]}$$

and rearranged to give

$$v = m(\pi) + \beta(\pi) [q^m(\pi) - \hat{q}] [q_{\max} - q^m(\pi)] > m(\pi) \quad \text{iff } \pi < \hat{\pi}$$

and

$$v = m(\pi) - \beta(\pi) [q^m(\pi) - \hat{q}] [q^m(\pi) - q_{\min}] > m(\pi) \quad \text{iff } \pi > \hat{\pi}$$

where the inequalities are obvious if one notes that $q^m(\pi) - \hat{q}$ has the opposite sign to $\pi - \hat{\pi}$.²⁶ We have the alternative formulations

$$v = \hat{m} - \frac{1}{2} \Delta\alpha [q_{\max} - \hat{q}] (\pi - \hat{\pi})$$

and

$$v = \hat{m} + \frac{1}{2} \Delta\alpha [\hat{q} - q_{\min}] (\pi - \hat{\pi}).$$

Now, the first order condition holds for some $q \in]q_{\min}, q_{\max}[$ if and only if

$$q_{\min} < q^m(\pi) + \frac{v - m(\pi)}{\beta(\pi) [q^m(\pi) - \hat{q}]} < q_{\max};$$

the first inequality is equivalent to $(\pi, v) \in \mathcal{A}_{\text{int},r}$, and the second to $(\pi, v) \in \mathcal{A}_{\text{int},\ell}$.

The expression given for the minimising quantity is simply a manipulation of the right-hand side of (D.4), which when substituted into (D.3) yields the expression for the corresponding minimum. \blacksquare

Finally, we show that the optimisation problems have the unique corner solution q_{\max} if (π, v) lies on or above \mathcal{R}_ℓ but to the left of \mathcal{R}_c , and the unique corner solution q_{\min} if (π, v) lies on or above \mathcal{R}_r but to the right of \mathcal{R}_c ; for $(\pi, v) \in \mathcal{R}_c$ both corner solutions are optimal.

Lemma D.3 *Let \hat{q} lie in the interior of Q^m . For $(\pi, v, s) \in \mathcal{A} \times \mathbb{R}_+$, the minimisation problem in (D.3) has the corner solution q_{\max} if and only if $(\pi, v) \in \mathcal{A}_{\max} \cup \mathcal{R}_c$, and the corner solution q_{\min} if and only if $(\pi, v) \in \mathcal{A}_{\min} \cup \mathcal{R}_c$, where*

$$\mathcal{A}_{\max} = \left\{ (\pi, v) \in \mathcal{A} : v \geq \hat{m} - \frac{1}{2} \Delta\alpha [q_{\max} - \hat{q}] (\pi - \hat{\pi}) \right. \\ \left. \text{and } \pi < \hat{\pi} + \frac{2}{\Delta\alpha} \frac{q_c - \hat{q}}{[q_{\max} - \hat{q}] [\hat{q} - q_{\min}]} (v - \hat{m}) \right\}$$

and

$$\mathcal{A}_{\min} = \left\{ (\pi, v) \in \mathcal{A} : v \geq \hat{m} + \frac{1}{2} \Delta\alpha [\hat{q} - q_{\min}] (\pi - \hat{\pi}) \right. \\ \left. \text{and } \pi > \hat{\pi} + \frac{2}{\Delta\alpha} \frac{q_c - \hat{q}}{[q_{\max} - \hat{q}] [\hat{q} - q_{\min}]} (v - \hat{m}) \right\}.$$

PROOF: In light of the previous lemma, we know that corner solutions will prevail in the sub-regions under consideration. Also, it is easy to see from the alternative parameterisation of the central ray, namely

$$v = \hat{m} + \frac{1}{2} \Delta\alpha \frac{[q_{\max} - \hat{q}] [\hat{q} - q_{\min}]}{q_c - \hat{q}} (\pi - \hat{\pi})$$

for $q_c \neq \hat{q}$, where $q_c = \frac{1}{2} (q_{\min} + q_{\max})$, that \mathcal{R}_c lies between \mathcal{R}_ℓ and \mathcal{R}_r .

In this region, the borderline case arises when q_{\max} and q_{\min} are both optimal and give the same value of $\tau(\pi) s$ in (D.3). This is the case if and only if

$$\frac{\beta(\pi) [q_{\max} - q^m(\pi)]^2 + v - m(\pi)}{[q_{\max} - \hat{q}]^2} = \frac{\beta(\pi) [q_{\min} - q^m(\pi)]^2 + v - m(\pi)}{[q_{\min} - \hat{q}]^2}$$

and simplification leads to

$$\pi = \hat{\pi} + \frac{2}{\Delta\alpha} \frac{q_c - \hat{q}}{[q_{\max} - \hat{q}] [\hat{q} - q_{\min}]} (v - \hat{m}).$$

Thus both extreme quantities are optimal for $(\pi, v) \in \mathcal{R}_c$.

It follows immediately that q_{\max} is uniquely optimal for $(\pi, v) \in \mathcal{A}_{\max}$, to the left of \mathcal{R}_c , and q_{\min} is uniquely optimal for $(\pi, v) \in \mathcal{A}_{\min}$, to the right of \mathcal{R}_c . \blacksquare

²⁶Using these two inequalities, it is easy to see that \mathcal{R}_ℓ cuts the axis $\pi = 0$ above $m(0)$, and that \mathcal{R}_r cuts the vertical line $\pi = 1$ above $m(1)$.

E The Undiscounted Case

In the absence of discounting, the monopolist uses the *catching-up criterion* to choose amongst admissible strategies: given a prior belief π , he looks for a strategy $\mathbf{q}^* \in \mathcal{Q}$ which, in the long run, does at least as well as any other strategy in the sense that $\liminf_{T \rightarrow \infty} \mathbb{E}_\pi [R_T^{\mathbf{q}^*} - R_T^{\mathbf{q}}] \geq 0$ for all $\mathbf{q} \in \mathcal{Q}$, where

$$R_T^{\mathbf{q}} = \int_0^T q_t [(\alpha_{k_t} - \beta_{k_t} q_t) dt + \sigma dZ_t]$$

is the process of cumulative revenues. The agent achieves this goal by maximising the *transient pay-off*, that is, total expected revenue net of the highest possible long-run average pay-off. Indeed, it can be shown that a strategy which achieves the maximum transient pay-off is catching-up optimal.

E.1 No State Switching ($r = 0, \Lambda = 0$)

Let $\Lambda = 0$, so the state of demand is fixed over time. Then, the monopolist can achieve a long-run average pay-off arbitrarily close to the full-information pay-off, that is, $m(0)$ if the true state is $k = 0$, and $m(1)$ if the true state is $k = 1$. In fact, it suffices to follow any admissible policy which coincides with the myopic policy q^m for beliefs close to 0 and 1, and is bounded away from the confounding quantity \hat{q} in case the latter lies in the interior of Q^m , the range of the myopic policy. By the martingale convergence theorem and the standard boundary classification for diffusion processes, beliefs will converge to the truth with probability one,²⁷ and the quantity chosen will approach the quantity which is optimal for the true demand. Given the initial belief π , the agent's objective is therefore to maximise the transient pay-off

$$u^{\mathbf{q}}(\pi) = \mathbb{E}_\pi \left[\int_0^\infty [R(k, q_t) - m(k)] dt \right]$$

where k is the unknown state of demand. By the law of iterated expectations,

$$u^{\mathbf{q}}(\pi) = \mathbb{E}_\pi \left[\int_0^\infty [R(\pi_t, q_t) - \bar{m}(\pi_t)] dt \right]$$

where

$$\bar{m}(\pi) = \pi m(1) + (1 - \pi) m(0)$$

is the *ex ante full-information pay-off*.

Standard results imply that the value function $u^*(\pi) = \sup_{\mathbf{q} \in \mathcal{Q}} u^{\mathbf{q}}(\pi)$ solves the Bellman equation

$$\max_{q \in Q} \left\{ \frac{1}{2} \Sigma^2(\pi, q) u''(\pi) - \bar{m}(\pi) + R(\pi, q) \right\} = 0 \quad (\text{E.1})$$

subject to the boundary conditions $u(0) = u(1) = 0$. Moreover, if a function u solves (E.1) with these boundary conditions, and there is an admissible policy function $q : [0, 1] \rightarrow Q$ such that

$$q(\pi) \in \arg \max_{q \in Q} \left\{ \frac{1}{2} \Sigma^2(\pi, q) u''(\pi) - \bar{m}(\pi) + R(\pi, q) \right\} \quad (\text{E.2})$$

for all π , then $u = u^*$ and the given policy is optimal.

In view of the results of Section 3.4, property (E.2) is equivalent to $q(\pi) \in O(\pi, \bar{m}(\pi))$. Moreover, the affine function \bar{m} trivially solves the ODE (20) for $r = 0$ and $\Lambda = 0$ with $\bar{m}(0) = m(0)$ and $\bar{m}(1) = m(1)$ as boundary values. We can therefore simply define the adjusted value function as $v^* = \bar{m}$.

²⁷We are assuming here that the agent does not assign prior probability zero to the true state. See Karlin and Taylor (1981) for the classification of boundary points.

E.2 State Switching ($r = 0, \Lambda > 0$)

We now assume $\Lambda > 0$. Let θ^* be the highest long-run average pay-off achievable with a strategy $\mathbf{q} \in \mathcal{Q}$. According to the introductory remarks to this section, the monopolist's objective is then to maximise

$$u^{\mathbf{q}}(\pi) = \mathbb{E}_{\pi} \left[\int_0^{\infty} [R(k_t, q_t) - \theta^*] dt \right] = \mathbb{E}_{\pi} \left[\int_0^{\infty} [R(\pi_t, q_t) - \theta^*] dt \right],$$

the transient pay-off as measured against the benchmark θ^* .

It can be shown that θ^* and the value function $u^*(\pi) = \sup_{\mathbf{q} \in \mathcal{Q}} u^{\mathbf{q}}(\pi)$ solve the Bellman equation

$$\max_{q \in \mathcal{Q}} \left\{ \frac{1}{2} \Sigma^2(\pi, q) u''(\pi) + \lambda(\pi) u'(\pi) - \theta + R(\pi, q) \right\} = 0 \quad (\text{E.3})$$

almost everywhere subject to the boundary conditions $\theta - \lambda(0) u'(0) = m(0)$ and $\theta - \lambda(1) u'(1) = m(1)$.²⁸ Conversely, if a real number θ and a function u solve (E.3) with the stated boundary conditions, and there is an admissible policy function $q : [0, 1] \rightarrow \mathcal{Q}$ such that

$$q(\pi) \in \arg \max_{q \in \mathcal{Q}} \left\{ \frac{1}{2} \Sigma^2(\pi, q) u''(\pi) + \lambda(\pi) u'(\pi) - \theta + R(\pi, q) \right\} \quad (\text{E.4})$$

for all π , then $\theta = \theta^*$, u is the value function up to a constant of integration, and the given policy is optimal.

Given the above boundary conditions and the fact that (E.4) is equivalent to $q(\pi) \in O(\pi, \theta - \lambda(\pi) u'(\pi))$, we define the adjusted value function by $v^*(\pi) = \theta^* - \lambda(\pi) (u^*)'(\pi)$. As the analysis in Appendix D remains valid for $r = 0$ with v now standing for $\theta - \lambda(\pi) u'(\pi)$ and s representing $u''(\pi)/2$, we can argue as in Section 3.6, and show that this v^* is indeed a solution (with the caveats stated there) of the ODE (20) for $r = 0$ and $\Lambda > 0$.

F Convexity of the Adjusted Value Function

Fix $\tilde{\pi} \in]0, 1[$, $\Lambda > 0$ and $r \geq 0$. For $k > 0$, let $w_k :]0, 1[- \{\tilde{\pi}\} \rightarrow \mathbb{R}$ be the function defined implicitly by the equation

$$G(\pi, w_k(\pi)) = k \tau(\pi) |\pi - \tilde{\pi}|^{-2-r/\Lambda}$$

with G given in equation (13). Then we have the following facts:

- $w_k \geq m$, with equality at $\pi = 0$ and 1 , and a strict inequality everywhere else;
- w_k has a pole at $\pi = \tilde{\pi}$;
- $w_k''(\pi) > 0$ unless $(\pi, w_k(\pi)) \in \mathcal{R}_c$ (see Lemma F.2 below);
- $r G(\pi, w_k(\pi)) + \Lambda \left\{ f(\pi) G(\pi, w_k(\pi)) + (\pi - \tilde{\pi}) \frac{d}{d\pi} G(\pi, w_k(\pi)) \right\} = 0$ for all $\pi \neq \tilde{\pi}$.

This last property explains our interest in the family of functions w_k . In fact, since $\frac{d}{d\pi} G(\pi, v(\pi))$ is strictly increasing in $v'(\pi)$, we have the following characterisation for the curvature of v^* on the set $\{\pi : v^*(\pi) > m(\pi)\}$.²⁹

Lemma F.1 *Let π be such that $v^*(\pi) > m(\pi)$ and $(\pi, v^*(\pi)) \notin \mathcal{R}_c$. Then $(v^*)''(\pi) > 0$ if $\pi = \tilde{\pi}$. If $\pi \neq \tilde{\pi}$, let w be that function w_k which coincides with v^* at π . Then $(v^*)''(\pi) > 0$ if and only if either $\pi < \tilde{\pi}$ and $(v^*)'(\pi) < w'(\pi)$, or $\pi > \tilde{\pi}$ and $(v^*)'(\pi) > w'(\pi)$.*

²⁸Moreover, u^* possesses the same regularity properties as the value function in the discounted case; arguments similar to those given in Appendices B and C apply. The discussion in Sections 3.2 and 3.3 regarding conditions for experimentation and the confounding quantity carries over as well.

²⁹Note that if $v^*(\pi) = m(\pi)$ with $\pi \neq \tilde{\pi}$, then trivially $(v^*)''(\pi) > 0$ since $m''(\pi) > 0$.

PROOF: This follows directly from the fact that v^* solves the ODE (20) at π . ■

We shall use this to prove that the adjusted value function is strictly convex whenever experimentation is moderate.

Proposition F.1 *If \hat{q} is not in the interior of Q^m , then $(v^*)'' > 0$ on $]0, 1[$.*

PROOF: Suppose that $(v^*)''(\pi^\dagger) \leq 0$ with $0 < \pi^\dagger < 1$; without loss of generality, we assume that $\pi^\dagger < \tilde{\pi}$. Let $k^\dagger > 0$ be such that the function $w^\dagger = w_{k^\dagger}$ coincides with v^* at π^\dagger . By the above lemma, $(v^*)'(\pi^\dagger) \geq (w^\dagger)'(\pi^\dagger)$. Now, the strict convexity of the functions w_k implies that $(v^*)'(\pi) > w'_k(\pi)$ for $\pi < \pi^\dagger$ sufficiently close to π^\dagger and $k < k^\dagger$ sufficiently close to k^\dagger . So $(v^*)''$ is strictly negative immediately to the left of π^\dagger . On the other hand, v^* cannot be strictly concave on the whole of $]0, \pi^\dagger[$ since this would imply $(v^*)' > w'$ on $]0, \pi^\dagger[$ and hence $v^*(0) < w^\dagger(0) = m(0)$. Therefore, there must be positive $\pi < \pi^\dagger$ such that $(v^*)''(\pi) = 0$ again; let π^\ddagger be the biggest such π , and w^\ddagger that function w_k which is tangent to v^* at π^\ddagger . On $[\pi^\ddagger, \pi^\dagger]$, w^\ddagger and w^\dagger are strictly convex, while v^* is strictly concave. This implies $w^\ddagger(\pi^\ddagger) < w^\dagger(\pi^\ddagger)$ and $w^\ddagger(\pi^\dagger) > w^\dagger(\pi^\dagger)$, so w^\ddagger and w^\dagger must intersect somewhere on $]\pi^\ddagger, \pi^\dagger[$ – a contradiction. The same argument can be used to the right of $\tilde{\pi}$. ■

Proposition F.2 *Suppose that \hat{q} lies in the interior of Q^m and $v^*(\hat{\pi}) = \hat{m}$. Then v^* is strictly convex with $(v^*)'' > 0$ on $]0, 1[- \{\hat{\pi}\}$, and at most one of the one-sided derivatives $(D_-v^*)(\hat{\pi})$ and $(D_+v^*)(\hat{\pi})$ can be different from zero. In fact, if $\hat{\pi} < \tilde{\pi}$, then $(D_+v^*)(\hat{\pi}) = 0$ and $(D_-v^*)(\hat{\pi}) \leq 0$; and if $\hat{\pi} > \tilde{\pi}$, then $(D_-v^*)(\hat{\pi}) = 0$ and $(D_+v^*)(\hat{\pi}) \geq 0$. If v^* is differentiable at $\hat{\pi}$, then the ratio $[v^*(\pi) - \hat{m}]/[m(\pi) - \hat{m}]$ converges to a finite limit as $\pi \rightarrow \hat{\pi}$. If v^* has a kink at $\hat{\pi}$, then $[v^*(\pi) - \hat{m}]/[m(\pi) - \hat{m}]$ converges to a finite limit as π approaches $\hat{\pi}$ from the direction of $\tilde{\pi}$.*

PROOF: We first convince ourselves that we can assume $q_c = \hat{q}$ without loss of generality. In fact, suppose we have shown the stated properties for the adjusted value function v^* in this particular case. Then strict convexity of v^* implies that $v^* < \bar{m}_\ell$ on $]0, \hat{\pi}[$ and $v^* < \bar{m}_r$ on $]\hat{\pi}, 1[$ with \bar{m}_ℓ and \bar{m}_r being the functions whose graphs are the straight lines joining the point $(\hat{\pi}, \hat{m})$ with the points $(0, m(0))$ and $(1, m(1))$, respectively. In particular, the graph of v^* lies entirely in the closure of $\mathcal{A}_{\text{int}, \ell} \cup \mathcal{A}_{\text{int}, r}$. Arguing exactly as in the proof of Proposition 4.2, we see that v^* is the adjusted value function for any q_{max} and q_{min} such that $[q_{\text{min}}, q_{\text{max}}] \supseteq Q^m$. So we have the stated properties of the adjusted value function for $q_c \neq \hat{q}$ as well.

Suppose therefore that $q_c = \hat{q}$, implying that the central ray \mathcal{R}_c is vertical. This simplifies the following analysis since it rules out intersections between the graph of v^* and \mathcal{R}_c , so we do not have to worry about the ‘break’ in (20) along \mathcal{R}_c .

Below, we will make repeated use of the following observation:

$$(v^*)'(\pi) - 2\Lambda(\pi - \tilde{\pi})G(\pi, v^*(\pi))/\tau(\pi) \text{ is continuous on } [0, 1]. \quad (\text{F.1})$$

In fact, the definition of the adjusted value function in (19) and the ODE (14) together with its undiscounted variant imply that this expression equals $(1 + \Lambda/r)(u^*)'(\pi)$ if $r > 0$, and $\Lambda(u^*)'(\pi)$ otherwise. So (F.1) follows from continuity of $(u^*)'$.

We can now turn to the proof of convexity of v^* . For the sake of concreteness, we assume that $\hat{\pi} < \tilde{\pi}$. Again, this is without loss of generality, since we could always relabel the demand curves.

We consider the subinterval left of $\hat{\pi}$ first. Suppose that $(v^*)'' \leq 0$ on $]0, \hat{\pi}[$. For any π in this interval, the function w_k that satisfies $w_k(\pi) = v^*(\pi)$ must then have $w'_k(\pi) \leq (v^*)'(\pi)$ by the above lemma. In fact, the equality $w'_k(\pi) = (v^*)'(\pi)$ is precluded since it would imply convexity of v^* immediately to the right of π . But $w'_k(\pi) < (v^*)'(\pi)$ cannot hold on the whole of $]0, \hat{\pi}[$ either: once w_k has crossed v^* from above it has to cross it from below later so as to reach the value $w_k(\hat{\pi}) > \hat{m}$. Thus, we must have $\inf\{\pi \in]0, \hat{\pi}[: (v^*)''(\pi) > 0\} < \hat{\pi}$. The same argument as in the previous proof now shows that this infimum is 0.

Arguing once more as in the previous proof, we also see that $(v^*)'' > 0$ on $]\tilde{\pi}, 1[$. Now let $\pi^\dagger = \inf\{\pi > \hat{\pi} : (v^*)''(\pi) > 0\}$. We know that $(v^*)''(\tilde{\pi}) > 0$, so $\hat{\pi} \leq \pi^\dagger < \tilde{\pi}$. Suppose $\pi^\dagger > \hat{\pi}$. Arguing again as in the previous proof, we can show that $(v^*)'' < 0$ immediately to the left of π^\dagger ;

moreover, $(v^*)''$ must be negative on the whole of $]\hat{\pi}, \pi^\dagger[$ since the existence of a π^\ddagger in this interval with $(v^*)''(\pi^\ddagger) \geq 0$ would lead to a contradiction. Thus, the one-sided derivatives of v^* at $\hat{\pi}$ are well defined, and we have $(D_-v^*)(\hat{\pi}) \leq 0$ and $(D_+v^*)(\hat{\pi}) > 0$ since $v^* \geq m$, $m'(\hat{\pi}) = 0$ and v^* is strictly concave immediately to the right of $\hat{\pi}$. In view of (F.1), we conclude that $G(\pi, v^*(\pi))$ has one-sided limits at $\hat{\pi}$ with $\lim_{\pi \rightarrow \hat{\pi}^-} G(\pi, v^*(\pi)) > \lim_{\pi \rightarrow \hat{\pi}^+} G(\pi, v^*(\pi))$. The explicit representation for G in (16)–(18) shows that these limits lie in the interval $[0, \beta(\hat{\pi})]$. However, $(D_+v^*)(\hat{\pi}) > 0$ implies $\lim_{\pi \rightarrow \hat{\pi}^+} G(\pi, v^*(\pi)) = \beta(\hat{\pi})$ and hence $\lim_{\pi \rightarrow \hat{\pi}^-} G(\pi, v^*(\pi)) > \beta(\hat{\pi})$ – a contradiction. Therefore $\pi^\dagger = \hat{\pi}$ and $(v^*)'' > 0$ on the whole of $]0, 1[- \{\hat{\pi}\}$. Moreover, we have shown that $(D_+v^*)(\hat{\pi}) = 0$.

The left-hand derivative $(D_-v^*)(\hat{\pi})$ exists also and is non-positive. (F.1) therefore implies the existence of one-sided limits $\lim_{\pi \rightarrow \hat{\pi}^-} G(\pi, v^*(\pi)) \geq \lim_{\pi \rightarrow \hat{\pi}^+} G(\pi, v^*(\pi))$ in the interval $[0, \beta(\hat{\pi})]$, the inequality being strict iff $(D_-v^*)(\hat{\pi}) < 0$. Having established convexity of v^* , we also know that its graph lies entirely in the region associated with interior quantities. So the relevant expression for the function G is $G(\pi, v) = \beta(\pi)[v - m(\pi)]/[v - \hat{m}]$. We can now prove the rest of the proposition.

If $(D_-v^*)(\hat{\pi}) < 0$, then $\lim_{\pi \rightarrow \hat{\pi}^+} G(\pi, v^*(\pi)) < \lim_{\pi \rightarrow \hat{\pi}^-} G(\pi, v^*(\pi)) = \beta(\hat{\pi})$ where the equality follows by L'Hôpital's rule. Since $[v - \hat{m}]/[m(\pi) - \hat{m}] = \beta(\pi)/[\beta(\pi) - G(\pi, v)]$, this proves that $[v^*(\pi) - \hat{m}]/[m(\pi) - \hat{m}]$ tends to a finite limit as $\pi \rightarrow \hat{\pi}^+$.

If $(D_-v^*)(\hat{\pi}) = 0$, on the other hand, then $G(\pi, v^*(\pi))$ approaches the same limit from both sides of $\hat{\pi}$. If this limit is strictly smaller than $\beta(\hat{\pi})$, then the quotient $[v^*(\pi) - \hat{m}]/[m(\pi) - \hat{m}]$ converges to a finite limit as $\pi \rightarrow \hat{\pi}$ (cf. the expression for this quotient given in previous paragraph). Suppose therefore that $\lim_{\pi \rightarrow \hat{\pi}} G(\pi, v^*(\pi)) = \beta(\hat{\pi})$, hence $\lim_{\pi \rightarrow \hat{\pi}} [v^*(\pi) - \hat{m}]/[m(\pi) - \hat{m}] = +\infty$. If $(d/d\pi)G(\pi, v^*(\pi))$ stays bounded above in a neighbourhood of $\hat{\pi}$, then $(v^*)''$ also remains bounded because of ODE (20); but then the mean value theorem implies that the quotient $[v^*(\pi) - \hat{m}]/[m(\pi) - \hat{m}]$ stays bounded as well. So $(d/d\pi)G(\pi, v^*(\pi))$ must be unbounded above as π approaches $\hat{\pi}$. Given that $\hat{\pi} < \tilde{\pi}$, (20) now implies that $(v^*)''$ is unbounded below – a contradiction to convexity. \blacksquare

We still have to show that the functions w_k are themselves convex.

Lemma F.2 *For $k > 0$, the function w_k is strictly convex at all $\pi \in]0, 1[- \{\tilde{\pi}\}$ such that $(\pi, w_k(\pi)) \notin \mathcal{R}_c$.*

PROOF: We fix a $k > 0$ and simply write w for the corresponding function w_k .

We first consider π such that $(\pi, w(\pi))$ lies in one of the regions associated with an interior quantity. It is straightforward to show that in these regions, w satisfies the first-order ODE

$$w'(\pi) = \frac{w(\pi) - \hat{m}}{m(\pi) - \hat{m}} [m'(\pi) - K(\pi)(w(\pi) - m(\pi))]$$

with

$$K(\pi) = \frac{2 + r/\Lambda}{\pi - \tilde{\pi}} + \frac{\Delta\beta}{\beta(\pi)} - \frac{2(1 - 2\pi)}{\pi(1 - \pi)}.$$

Differentiating both sides and using the ODE to replace $w'(\pi)$, we find that $w''(\pi)$ is a quadratic in $w(\pi) - m(\pi)$ multiplied by a positive factor:

$$w'' = [a(w - m)^2 + b(w - m) + c] d$$

with

$$\begin{aligned} a &= 2K^2, \\ b &= (K^2 - K')(m - \hat{m}) - 2Km', \\ c &= (m - \hat{m})m'', \\ d &= (w - \hat{m})/(m - \hat{m})^2, \end{aligned}$$

where we have suppressed the dependence of the functions K , m and w on π . Clearly $a > 0$ and, since $m > \hat{m}$ and m is convex, we also have $c > 0$. Thus, if $b \geq 0$, then $w'' > 0$ and we are done.

Suppose therefore that $b < 0$. We have to show that the above quadratic in $w - m$ has no real roots. This is the same as showing $b^2 - 4ac < 0$, or equivalently $(b + 2\sqrt{ac})(b - 2\sqrt{ac}) < 0$. Since we are

dealing with the case $b < 0$, the second factor is negative, so all we have to show is that $b + 2\sqrt{ac} > 0$, i.e.,

$$(K^2 - K')(m - \hat{m}) - 2Km' + 2\sqrt{2K^2(m - \hat{m})m''} > 0$$

when $(K^2 - K')(m - \hat{m}) - 2Km' < 0$.

First, we can show that $K^2 - K' > 0$. Indeed, $K^2 - K' = [(r/\Lambda)^2 + (r/\Lambda)f_1 + 2f_2]/(\pi - \tilde{\pi})^2$ with

$$\begin{aligned} f_1(\pi) &= \{\tilde{\pi}(1 - \pi)[(1 - \pi)(\beta_0 + \beta(\pi)) + (1 - \pi)\beta_0 + \beta(\pi)] \\ &\quad + (1 - \tilde{\pi})\pi[\pi(\beta_1 + \beta(\pi)) + \pi\beta_1 + \beta(\pi)]\}/\{\pi(1 - \pi)\beta(\pi)\} \end{aligned}$$

and

$$f_2(\pi) = \frac{(\beta_0\tilde{\pi}(1 - \pi)^2 + \beta_1(1 - \tilde{\pi})\pi^2)^2 + 2\tilde{\pi}(1 - \tilde{\pi})\pi(1 - \pi)\beta(\pi)^2}{(\pi(1 - \pi)\beta(\pi))^2}$$

which are positive by inspection. This leads to the further simplification that the only case we need consider is when $Km' > 0$.

It will be more convenient to rewrite $m(\pi)$ and its derivatives in terms of $q^m(\pi)$ and \hat{q} as follows:

$$m - \hat{m} = \beta(q^m - \hat{q})^2, \quad m' = -\Delta\beta q^m (q^m - \hat{q}), \quad m'' = \frac{\Delta\beta^2}{2\beta} (2q^m - \hat{q})^2$$

where again we have suppressed the dependence of the functions β , m and q^m on π . Having made these substitutions, the expression which we wish to show is positive becomes

$$\beta(K^2 - K')(q^m - \hat{q})^2 + 2\Delta\beta \{Kq^m (q^m - \hat{q}) + |K| |q^m - \hat{q}| |2q^m - \hat{q}|\},$$

and we need consider only the case when $K(q^m - \hat{q}) < 0$.

Case 1: $K < 0$, $q^m - \hat{q} > 0$. The expression in braces becomes

$$-|K|q^m (q^m - \hat{q}) + |K|(q^m - \hat{q})(2q^m - \hat{q}) = |K|(q^m - \hat{q})^2,$$

so we are done.

Case 2: $K > 0$, $q^m - \hat{q} < 0$. This time the expression in braces becomes

$$-Kq^m |q^m - \hat{q}| + K|q^m - \hat{q}| |2q^m - \hat{q}| = K|q^m - \hat{q}| (|2q^m - \hat{q}| - q^m),$$

and we have two subcases.

Case 2i: $2q^m \geq \hat{q}$. The above expression becomes

$$K|q^m - \hat{q}| (2q^m - \hat{q} - q^m) = -K(q^m - \hat{q})^2$$

and the whole expression which we wish to show is positive becomes

$$\beta(K^2 - K')(q^m - \hat{q})^2 - 2\Delta\beta K(q^m - \hat{q})^2 = [\beta(K^2 - K') - 2\Delta\beta K](q^m - \hat{q})^2.$$

Case 2ii: $2q^m < \hat{q}$. Now the expression for Case 2 becomes

$$K|q^m - \hat{q}| (-2q^m + \hat{q} - q^m) = K(\hat{q} - q^m)(\hat{q} - 3q^m)$$

and the whole expression which we wish to show is positive becomes

$$\left[\beta(K^2 - K') + 2\Delta\beta K \frac{\hat{q} - 3q^m}{\hat{q} - q^m} \right] (q^m - \hat{q})^2.$$

But $q^m - \hat{q} < 0$ and $2q^m < \hat{q}$ imply that $(\hat{q} - 3q^m)/(\hat{q} - q^m) > -1$, and so the above expression is greater than

$$[\beta(K^2 - K') - 2\Delta\beta K](q^m - \hat{q})^2$$

which is just the expression that we found in Case 2i.

Therefore, Case 2 comes down to showing that the term in square brackets is positive. In fact, it can be written as

$$\begin{aligned} & \frac{2\beta(\pi) [\tilde{\pi}^2(1-\pi)^4 + (1-\tilde{\pi})^2\pi^4 + 2\tilde{\pi}(1-\tilde{\pi})\pi(1-\pi)(1+\pi(1-\pi))]}{\pi^2(1-\pi)^2(\pi-\tilde{\pi})^2} \\ & + \frac{\beta(\pi)(r/\Lambda) [\tilde{\pi}(1-\pi)(1+3(1-\pi)) + (1-\tilde{\pi})\pi(1+3\pi)]}{\pi(1-\pi)(\pi-\tilde{\pi})^2} \\ & + \frac{\beta(\pi)(r/\Lambda)^2}{(\pi-\tilde{\pi})^2} \end{aligned}$$

which is positive by inspection. Thus $w''(\pi) > 0$ in the regions associated with interior solutions.

We still have to consider π such that $(\pi, w(\pi))$ lies in a region associated with exactly one of the extreme quantities. This quantity, which can be either q_{\max} or q_{\min} , will be denoted by q^\dagger . It is straightforward to show that in such a region, w satisfies

$$w'(\pi) = q^\dagger (\Delta\alpha - \Delta\beta q^\dagger) - L(\pi) (w(\pi) - R(\pi, q^\dagger))$$

with

$$L(\pi) = \frac{2+r/\Lambda}{\pi-\tilde{\pi}} - \frac{2(1-2\pi)}{\pi(1-\pi)},$$

and that

$$w''(\pi) = [L^2(\pi) - L'(\pi)] (w(\pi) - R(\pi, q^\dagger)).$$

As $w(\pi) > R(\pi, q^\dagger)$ by construction, we only have to show that $L^2 - L' > 0$. This follows from the representation $L^2 - L' = [(r/\Lambda)^2 + (r/\Lambda)g_1 + 2g_2]/(\pi - \tilde{\pi})^2$ where

$$g_1(\pi) = \frac{\tilde{\pi}(1-\pi)[3(1-\pi)+1] + (1-\tilde{\pi})\pi[3\pi+1]}{\pi(1-\pi)}$$

and

$$g_2(\pi) = \frac{(\tilde{\pi}(1-\pi)^2 + (1-\tilde{\pi})\pi^2)^2 + 2\tilde{\pi}(1-\tilde{\pi})\pi(1-\pi)}{\pi^2(1-\pi)^2}$$

are clearly positive; this representation is obtained by setting $\Delta\beta = 0$ in the above expression for $K^2 - K'$. ■

G Two-Point Boundary Value Problems

Consider a second-order differential equation of the form

$$\pi^2(1-\pi)^2 v'' = F[\pi, v, v'] \tag{G.1}$$

on some open interval $I =]\pi_\ell, \pi_r[\subseteq]0, 1[$. We are interested in finding a solution to this ODE which assumes prespecified values at the two boundary points of the interval.

The existence theorem presented below requires the concept of a sub- or supersolution to this ODE. Let v be a real-valued continuous function on $\bar{I} = [\pi_\ell, \pi_r]$ with a continuous first derivative on I . Define functions $\underline{D}v', \overline{D}v' : I \rightarrow \mathbb{R} \cup \{\pm\infty\}$ by

$$(\underline{D}v')(\pi) = \liminf_{h \rightarrow 0} \frac{v'(\pi+h) - v'(\pi-h)}{2h}, \quad (\overline{D}v')(\pi) = \limsup_{h \rightarrow 0} \frac{v'(\pi+h) - v'(\pi-h)}{2h}.$$

(Note that for twice differentiable v , the functions $\underline{D}v'$ and $\overline{D}v'$ coincide with v'' .) The function v is called a *subsolution* of the ODE (G.1) if $\pi^2(1-\pi)^2 \underline{D}v' \geq F[\pi, v, v']$ on I . Similarly, v is called a *supersolution* if $\pi^2(1-\pi)^2 \overline{D}v' \leq F[\pi, v, v']$ on I . We speak of a *strict* subsolution or supersolution if the respective inequality is strict on I .

Fix functions $\underline{v}, \overline{v} : \bar{I} \rightarrow \mathbb{R}$ satisfying $\underline{v} \leq \overline{v}$ on \bar{I} . Given any subinterval $J \subseteq \bar{I}$, we say that the function F on the right-hand side of (G.1) is *regular on J with respect to \underline{v} and \overline{v}* if it is continuous on $\mathcal{D}_J = \{(\pi, v_0, v_1) \in J \times \mathbb{R} \times \mathbb{R} : \underline{v}(\pi) \leq v_0 \leq \overline{v}(\pi)\}$ and there is a constant C_J depending only on J such that $|F[\pi, v_0, v_1]| \leq C_J(1 + |v_1|)$ on \mathcal{D}_J .

Proposition G.1 *Let $0 < \pi_\ell < \pi_r < 1$. Assume that $\underline{v} : \bar{I} \rightarrow \mathbb{R}$ is a subsolution of (G.1), $\bar{v} : \bar{I} \rightarrow \mathbb{R}$ a supersolution, and $\underline{v} \leq \bar{v}$. If F is regular with respect to \underline{v} and \bar{v} on $\bar{I} = [\pi_\ell, \pi_r]$, then for any $v_\ell \in [\underline{v}(\pi_\ell), \bar{v}(\pi_\ell)]$ and $v_r \in [\underline{v}(\pi_r), \bar{v}(\pi_r)]$, there is a continuous function $v : \bar{I} \rightarrow \mathbb{R}$ which solves (G.1) on I with $\underline{v} \leq v \leq \bar{v}$ and satisfies the boundary conditions $v(\pi_\ell) = v_\ell$ and $v(\pi_r) = v_r$. Moreover, if \underline{v} is a strict subsolution, then $v > \underline{v}$ on I , and if \bar{v} is a strict supersolution, then $v < \bar{v}$ on I .*

PROOF: The existence of such a solution v follows directly from Bernfeld and Lakshmikantham (1974, Theorem 1.5.1). Now assume that \underline{v} is a strict subsolution and that there is a belief $\tilde{\pi} \in I$ such that $v(\tilde{\pi}) = \underline{v}(\tilde{\pi})$. Then the function $v - \underline{v}$ has a local minimum at $\tilde{\pi}$, so $v'(\tilde{\pi}) = \underline{v}'(\tilde{\pi})$ and $v''(\tilde{\pi}) \geq (\underline{D}\underline{v})'(\tilde{\pi})$. Yet $\tilde{\pi}^2(1 - \tilde{\pi})^2 v''(\tilde{\pi}) = F[\tilde{\pi}, v(\tilde{\pi}), v'(\tilde{\pi})] = F[\tilde{\pi}, \underline{v}(\tilde{\pi}), \underline{v}'(\tilde{\pi})] < \tilde{\pi}^2(1 - \tilde{\pi})^2 (\underline{D}\underline{v})'(\tilde{\pi})$ — a contradiction. The case of a strict supersolution \bar{v} is dealt with in the same way. \blacksquare

We will need the following corollary of this result.

Corollary G.1 *Given $\pi_\ell < \pi_c < \pi_r$ in $]0, 1[$, consider the ODEs*

$$\pi^2(1 - \pi)^2 v'' = F_\ell[\pi, v, v'] \quad (\text{G.2})$$

on $]\pi_\ell, \pi_c[$ and

$$\pi^2(1 - \pi)^2 v'' = F_r[\pi, v, v'] \quad (\text{G.3})$$

on $]\pi_c, \pi_r[$. Let $\underline{v}_\ell : [\pi_\ell, \pi_c] \rightarrow \mathbb{R}$ be a subsolution of (G.2), $\bar{v}_\ell : [\pi_\ell, \pi_c] \rightarrow \mathbb{R}$ a supersolution of (G.2), $\underline{v}_r : [\pi_c, \pi_r] \rightarrow \mathbb{R}$ a subsolution of (G.3) and $\bar{v}_r : [\pi_c, \pi_r] \rightarrow \mathbb{R}$ a supersolution of (G.3) such that $\underline{v}_\ell \leq \bar{v}_\ell$, $\underline{v}_r \leq \bar{v}_r$, $\underline{v}_\ell(\pi_c) = \underline{v}_r(\pi_c) < \bar{v}_\ell(\pi_c) = \bar{v}_r(\pi_c)$, $\underline{v}'_\ell(\pi_c-) \leq \underline{v}'_r(\pi_c+)$ and $\bar{v}'_\ell(\pi_c-) \geq \bar{v}'_r(\pi_c+)$. Assume that F_ℓ is regular with respect to \underline{v}_ℓ and \bar{v}_ℓ on each closed interval contained in $]\pi_\ell, \pi_c[$, and F_r is regular with respect to \underline{v}_r and \bar{v}_r on each closed interval contained in $]\pi_c, \pi_r[$. Then there is a differentiable function $v :]\pi_\ell, \pi_r[\rightarrow \mathbb{R}$ which solves (G.2) on $]\pi_\ell, \pi_c[$ and (G.3) on $]\pi_c, \pi_r[$ such that $\underline{v}_\ell \leq v \leq \bar{v}_\ell$ on $]\pi_\ell, \pi_c[$ and $\underline{v}_r \leq v \leq \bar{v}_r$ on $]\pi_c, \pi_r[$.

PROOF: Piecing together \underline{v}_ℓ , \underline{v}_r and \bar{v}_ℓ , \bar{v}_r in the obvious way, we get continuous functions $\underline{v}, \bar{v} :]\pi_\ell, \pi_r[\rightarrow \mathbb{R}$. Let $\epsilon > 0$ be such that $\pi_\ell + \epsilon < \pi_c < \pi_r - \epsilon$. We shall construct numbers $\underline{a}_n, \bar{a}_n \in [\underline{v}(\pi_c), \bar{v}(\pi_c)]$ and functions $\underline{v}_n, \bar{v}_n$ with the following properties for all $n = 1, 2, \dots$:

- (i) $\underline{a}_n < \bar{a}_n$;
- (ii) $\underline{a}_{n+1} \geq \underline{a}_n$ and $\bar{a}_{n+1} \leq \bar{a}_n$;
- (iii) $\underline{v}_n, \bar{v}_n :]\pi_\ell + \epsilon, \pi_r - \epsilon[\rightarrow \mathbb{R}$ are continuous and solve (G.2) on $]\pi_\ell + \epsilon, \pi_c[$ and (G.3) on $]\pi_c, \pi_r - \epsilon[$ subject to $\underline{v}_n(\pi_\ell + \epsilon) = \bar{v}_n(\pi_\ell + \epsilon) = \bar{v}(\pi_\ell + \epsilon)$, $\underline{v}_n(\pi_r - \epsilon) = \bar{v}_n(\pi_r - \epsilon) = \bar{v}(\pi_r - \epsilon)$, $\underline{v}_n(\pi_c) = \underline{a}_n$, $\bar{v}_n(\pi_c) = \bar{a}_n$;
- (iv) $\underline{v}'_n(\pi_c-) \leq \underline{v}'_n(\pi_c+)$ and $\bar{v}'_n(\pi_c-) \geq \bar{v}'_n(\pi_c+)$;
- (v) $\underline{v} \leq \underline{v}_n < \bar{v}_n \leq \bar{v}$ on $]\pi_\ell + \epsilon, \pi_r - \epsilon[$;
- (vi) $\underline{v}_{n+1} \geq \underline{v}_n$ and $\bar{v}_{n+1} \leq \bar{v}_n$.

For $n = 1$, we set $\underline{a}_1 = \underline{v}(\pi_c)$ and $\bar{a}_1 = \bar{v}(\pi_c)$, so (i) holds. Using Proposition G.1 separately to the left and right of π_c , we find a function \underline{v}_1 satisfying (iii) and $\underline{v} \leq \underline{v}_1 \leq \bar{v}$, and a function \bar{v}_1 satisfying (iii) and $\underline{v}_1 \leq \bar{v}_1 \leq \bar{v}$. Property (iv) is then obvious, and a simple argument similar to the one given at the end of the previous proof shows (v). Suppose we have constructed $\underline{a}_n, \bar{a}_n, \underline{v}_n$ and \bar{v}_n with (i) and (iii)–(v). If $\underline{v}'_n(\pi_c-) = \underline{v}'_n(\pi_c+)$ or $\bar{v}'_n(\pi_c-) = \bar{v}'_n(\pi_c+)$, we simply set $\underline{a}_{n+1} = \underline{a}_n$, $\bar{a}_{n+1} = \bar{a}_n$, $\underline{v}_{n+1} = \underline{v}_n$ and $\bar{v}_{n+1} = \bar{v}_n$. Otherwise, we consider $a = (\underline{a}_n + \bar{a}_n)/2$ and a continuous function $v :]\pi_\ell + \epsilon, \pi_r - \epsilon[\rightarrow \mathbb{R}$ which satisfies $\underline{v}_n \leq v \leq \bar{v}_n$ and solves (G.2) on $]\pi_\ell + \epsilon, \pi_c[$ and (G.3) on $]\pi_c, \pi_r - \epsilon[$ subject to $v(\pi_\ell + \epsilon) = \bar{v}(\pi_\ell + \epsilon)$, $v(\pi_r - \epsilon) = \bar{v}(\pi_r - \epsilon)$, $v(\pi_c) = a$. Such a function exists by Proposition G.1, and it is again straightforward to see that $\underline{v}_n < v < \bar{v}_n$ on $]\pi_\ell + \epsilon, \pi_r - \epsilon[$. If $v'(\pi_c-) \leq v'(\pi_c+)$, we set $\underline{a}_{n+1} = a$, $\bar{a}_{n+1} = \bar{a}_n$, $\underline{v}_{n+1} = v$ and $\bar{v}_{n+1} = \bar{v}_n$; if $v'(\pi_c-) > v'(\pi_c+)$, we set $\underline{a}_{n+1} = \underline{a}_n$, $\bar{a}_{n+1} = a$, $\underline{v}_{n+1} = \underline{v}_n$ and $\bar{v}_{n+1} = v$. This procedure clearly implies (i)–(vi).

If none of the functions \underline{v}_n or \bar{v}_n is differentiable at π_c , then the sequences (\underline{a}_n) and (\bar{a}_n) converge to a common limit a_∞ , and by Bernfeld and Lakshmikantham (1974, Corollary 1.5.1), the sequences (\underline{v}_n) and (\bar{v}_n) have subsequences converging uniformly to functions $\underline{v}_\infty \leq \bar{v}_\infty$ which solve (G.2) and (G.3) on the respective open intervals, with the corresponding subsequences of (\underline{v}'_n) and (\bar{v}'_n) converging to \underline{v}'_∞ and \bar{v}'_∞ , respectively. As $\underline{v}_\infty(\pi_c) = \bar{v}_\infty(\pi_c) = a_\infty$, we have $\underline{v}'_\infty(\pi_c-) \geq \bar{v}'_\infty(\pi_c-)$ and $\underline{v}'_\infty(\pi_c+) \leq \bar{v}'_\infty(\pi_c+)$. On the other hand, $\underline{v}'_\infty(\pi_c-) \leq \underline{v}'_\infty(\pi_c+)$ and $\bar{v}'_\infty(\pi_c-) \geq \bar{v}'_\infty(\pi_c+)$, hence $\underline{v}'_\infty(\pi_c-) \geq \bar{v}'_\infty(\pi_c-) \geq \bar{v}'_\infty(\pi_c+) \geq \underline{v}'_\infty(\pi_c+) \geq \underline{v}'_\infty(\pi_c-)$, implying equality throughout.

For any small $\epsilon > 0$, we can therefore always find a function v_ϵ on $[\pi_\ell + \epsilon, \pi_r - \epsilon]$ which solves (G.2) on $]\pi_\ell + \epsilon, \pi_c[$ and (G.3) on $]\pi_c, \pi_r - \epsilon[$, is differentiable at π_c , and satisfies $v_\epsilon(\pi_\ell + \epsilon) = \bar{v}(\pi_\ell + \epsilon)$, $v_\epsilon(\pi_r - \epsilon) = \bar{v}(\pi_r - \epsilon)$, and $\underline{v} \leq v_\epsilon \leq \bar{v}$ everywhere else.

Finally, consider a sequence $v_k : [\pi_\ell + \epsilon_k, \pi_r - \epsilon_k] \rightarrow \mathbb{R}$ of such functions for small positive numbers $(\epsilon_k)_{k=1,2,\dots}$ converging monotonically to 0. By Bernfeld and Lakshmikantham (1974, Theorem 1.4.1), there is an $N_k > 0$ such that $|v'| \leq N_k$ on $[\pi_\ell + \epsilon_k, \pi_r - \epsilon_k]$ for *any* solution lying between \underline{v} and \bar{v} on this interval. Thus for any fixed integer $K \geq 1$ and all $k \geq K$, v_k is a solution satisfying $\underline{v} \leq v_k \leq \bar{v}$ and $|v'_k| \leq N_K$ on $[\pi_\ell + \epsilon_K, \pi_r - \epsilon_K]$, so the sequences $(v_k)_{k \geq K}$ and $(v'_k)_{k \geq K}$ are both uniformly bounded and equicontinuous on that interval. Employing the standard diagonalisation argument, we obtain a subsequence which converges uniformly on all compact subintervals of $]\pi_\ell, \pi_r[$ to a function v with the desired properties. \blacksquare

Our analysis of the Bellman equation lead us to the following second-order differential equation for the adjusted value function v^* :

$$\tau(\pi) \frac{v''(\pi)}{2} = r G(\pi, v(\pi)) + \Lambda \left\{ f(\pi) G(\pi, v(\pi)) + (\pi - \tilde{\pi}) \frac{d}{d\pi} G(\pi, v(\pi)) \right\} \quad (\text{G.4})$$

with

$$f(\pi) = 2 - (\pi - \tilde{\pi}) \frac{\tau'(\pi)}{\tau(\pi)}$$

and G defined by equation (13). We saw that G is continuously differentiable in the area \mathcal{A} with the exception of the central ray in case \hat{q} lies in the interior of Q^m ; if this is the case, we consider the ODE separately to the left and to the right of the central ray. Throughout, we will assume that at least one of the parameters r and Λ is strictly positive.

Lemma G.1 *The myopic pay-off function m is a strict subsolution of (G.4) on $]0, 1[$ if \hat{q} is not in the interior of Q^m , and on $]0, 1[- \{\hat{\pi}\}$ otherwise.*

PROOF: $m'' > 0$, and we have $G(\pi, m(\pi)) = 0$ on the stated sets of beliefs. \blacksquare

Lemma G.2 *The full information pay-off function \bar{m} is a strict supersolution of (G.4).*

PROOF: $\bar{m}'' = 0$, so we have to show that the right-hand side of (G.4) with $v(\pi)$ replaced by $\bar{m}(\pi) = (1 - \pi) m(0) + \pi m(1)$ is positive.

Suppose first that $(\pi, \bar{m}(\pi))$ lies to the left of \mathcal{R}_ℓ or to the right of \mathcal{R}_r . Then that right-hand side becomes

$$\begin{aligned} r \beta(\pi) \bar{H}(\pi) &+ \lambda_0 (1 - \pi) \left[\frac{\beta_0 + \beta(\pi)}{\pi} \bar{H}(\pi) - \beta(\pi) \bar{H}'(\pi) \right] \\ &+ \lambda_1 \pi \left[\frac{\beta_1 + \beta(\pi)}{1 - \pi} \bar{H}(\pi) + \beta(\pi) \bar{H}'(\pi) \right] \end{aligned}$$

with $\bar{H}(\pi) = [\bar{m}(\pi) - m(\pi)] / [\bar{m}(\pi) - \hat{m}]$. The first term is clearly positive. The expressions in square brackets associated with λ_0 and λ_1 simplify to $h_0(\pi) / (\bar{m}(\pi) - \hat{m})^2$ and $h_1(\pi) / (\bar{m}(\pi) - \hat{m})^2$ respectively, where h_0 and h_1 are quadratics in π :

$$\begin{aligned} h_0(\pi) &= K [m(1) - \hat{m} + [m(0) - m(1)] (1 - \pi)^2], \\ h_1(\pi) &= K [m(0) - \hat{m} + [m(1) - m(0)] \pi^2] \end{aligned}$$

with $K = \beta_0 \beta_1 [q^m(0) - q^m(1)]^2$. Thus, $h_0(0) = h_1(0) = K [m(0) - \hat{m}]$ and $h_0(1) = h_1(1) = K [m(1) - \hat{m}]$, so h_0 and h_1 are both non-negative at each end of the unit interval. As the two quadratics are strictly monotonic on $[0, 1]$, they are both non-negative over the entire unit interval.

Next consider π such that $(\pi, \overline{m}(\pi))$ lies between the rays \mathcal{R}_ℓ and \mathcal{R}_r . In such a region, the right-hand side for \overline{m} can be written as

$$r \overline{G}(\pi) + \lambda_0 (1 - \pi) \left[\frac{2}{\pi} \overline{G}(\pi) - \overline{G}'(\pi) \right] + \lambda_1 \pi \left[\frac{2}{1 - \pi} \overline{G}(\pi) + \overline{G}'(\pi) \right]$$

where $\overline{G}(\pi) = (\overline{m}(\pi) - m(\pi) + \beta(\pi)[q^\dagger - q^m(\pi)]^2)/[q^\dagger - \hat{q}]^2$ and the quantity q^\dagger is either q_{\max} or q_{\min} . Again, the first term is clearly positive. The expressions in square brackets associated with λ_0 and λ_1 simplify to $\ell_0(\pi)/(\pi [q^\dagger - \hat{q}]^2)$ and $\ell_1(\pi)/((1 - \pi)[q^\dagger - \hat{q}]^2)$ respectively, where ℓ_0 and ℓ_1 are the following linear functions:

$$\begin{aligned} \ell_0(\pi) &= (\beta_0 [q^\dagger - q^m(0)]^2 + \beta_1 [q^\dagger - q^m(1)]^2) + (\beta_0 [q^\dagger - q^m(0)]^2 - \beta_1 [q^\dagger - q^m(1)]^2) (1 - \pi), \\ \ell_1(\pi) &= (\beta_1 [q^\dagger - q^m(1)]^2 + \beta_0 [q^\dagger - q^m(0)]^2) + (\beta_1 [q^\dagger - q^m(1)]^2 - \beta_0 [q^\dagger - q^m(0)]^2) \pi. \end{aligned}$$

By inspection, these functions are positive on the unit interval. ■

Define

$$\overline{m}_\ell(\pi) = \frac{\hat{\pi} - \pi}{\hat{\pi}} m(0) + \frac{\pi}{\hat{\pi}} \hat{m}$$

for $0 \leq \pi \leq \hat{\pi}$, and

$$\overline{m}_r(\pi) = \frac{1 - \pi}{1 - \hat{\pi}} \hat{m} + \frac{\pi - \hat{\pi}}{1 - \hat{\pi}} m(1)$$

for $\hat{\pi} \leq \pi \leq 1$. The graphs of these functions are the rays joining $(\hat{\pi}, \hat{m})$ with $(0, m(0))$ and $(1, m(1))$, respectively.

Lemma G.3 *Let \hat{q} lie in the interior of Q^m . Then the functions $\overline{m}_\ell : [0, \hat{\pi}] \rightarrow \mathbb{R}$ and $\overline{m}_r : [\hat{\pi}, 1] \rightarrow \mathbb{R}$ are strict supersolutions of (G.4).*

PROOF: The functions \overline{m}_ℓ and \overline{m}_r are linear, so $\overline{m}_\ell'' = 0$ and $\overline{m}_r'' = 0$, and their graphs lie entirely in the sub-regions of \mathcal{A} associated with interior solutions. A slightly more complicated variant of the algebra in the first part of the previous proof shows that the right-hand side of (G.4) is positive for these functions. ■

Lemma G.4 *If \hat{q} does not lie in the interior of Q^m , then the right-hand side of the ODE (G.4) is regular with respect to m and \overline{m} on each closed interval contained in $]0, 1[$. Otherwise, the right-hand side of the ODE is regular with respect to m and \overline{m}_ℓ on each closed interval contained in $]0, \hat{\pi}[$, and regular with respect to m and \overline{m}_r on each closed interval contained in $]\hat{\pi}, 1[$.*

PROOF: This follows directly from the fact that in the regions associated with interior solutions, the ODE (G.4) is equivalent to the equation $\tau(\pi)v''(\pi)/2 = F[\pi, v(\pi), v'(\pi)]$ where

$$F[\pi, v_0, v_1] = \beta(\pi) \left\{ \left(r + \Lambda \left[f(\pi) + (\pi - \hat{\pi}) \frac{\Delta\beta}{\beta(\pi)} \right] \right) H(\pi, v_0) + \Lambda (\pi - \hat{\pi}) H_1[\pi, v_0, v_1] \right\}$$

with

$$H(\pi, v_0) = \frac{v_0 - m(\pi)}{v_0 - \hat{m}}, \quad H_1[\pi, v_0, v_1] = \frac{m(\pi) - \hat{m}}{(v_0 - \hat{m})^2} v_1 - \frac{m'(\pi)}{v_0 - \hat{m}}.$$

In particular, v_1 enters linearly. ■

Proposition G.2 *Suppose that \hat{q} does not lie in the interior of Q^m . Then there is a continuous function $v : [0, 1] \rightarrow \mathbb{R}$ which solves (G.4) on $]0, 1[$ with $v(0) = m(0)$, $v(1) = m(1)$, and $m < v < \overline{m}$ on $]0, 1[$.*

PROOF: This follows from Lemmas G.1, G.2 and G.4 and Corollary G.1 applied with $\pi_\ell = 0$, $\pi_r = 1$ and $F_\ell = F_r = F$ as given in the proof of Lemma G.4. \blacksquare

Proposition G.3 *Suppose that \hat{q} lies in the interior of Q^m , and fix $\tilde{\pi} \in]0, 1[-\{\hat{\pi}\}$. Then there are positive constants c_1 and c_2 such that for all $r \geq 0$, $\Lambda \geq 0$ and $\sigma > 0$ satisfying $r + c_1\Lambda \geq c_2/\sigma^2$, there exists a continuous function $v : [0, 1] \rightarrow \mathbb{R}$ which solves (G.4) on $]0, 1[-\{\hat{\pi}\}$ with the following properties: $m < v < \overline{m}_\ell$ on $]0, \hat{\pi}[$; $m < v < \overline{m}_r$ on $]\hat{\pi}, 1[$; $v - \hat{m} \leq 2(m - \hat{m})$ in a neighbourhood of $\hat{\pi}$. In particular, v is differentiable with $v(\hat{\pi}) = \hat{m}$ and $v'(\hat{\pi}) = 0$.*

PROOF: Consider the function \check{m} defined by $\check{m}(\pi) = 2m(\pi) - \hat{m}$. Let $\check{\pi}_\ell$ be the belief where the graph of \check{m} intersects the graph of \overline{m}_ℓ , and $\check{\pi}_r$ the belief where the graph of \check{m} intersects the graph of \overline{m}_r . Define

$$c_1 = \min_{\check{\pi}_\ell \leq \pi \leq \check{\pi}_r} [f(\pi) + (\pi - \tilde{\pi})\Delta\beta/\beta(\pi)], \quad c_2 = 2\Delta\beta^2 \max_{\check{\pi}_\ell \leq \pi \leq \check{\pi}_r} \frac{\pi^2(1-\pi)^2 m''(\pi)}{\beta(\pi)}.$$

While c_2 is clearly positive, the positivity of c_1 follows from the identity

$$f(\pi) + (\pi - \tilde{\pi})\frac{\Delta\beta}{\beta(\pi)} = \frac{\tilde{\pi}(1-\pi)^2(\beta_0 + \beta(\pi)) + (1-\tilde{\pi})\pi^2(\beta_1 + \beta(\pi))}{\pi(1-\pi)\beta(\pi)}.$$

Now let $r + c_1\Lambda \geq c_2/\sigma^2$, implying that \check{m} is a supersolution of (G.4).

Clearly, the right-hand side of (G.4) is regular with respect to m and \check{m} on each closed subinterval of $[\check{\pi}_\ell, \check{\pi}_r] - \{\hat{\pi}\}$. Since $\check{m}'(\check{\pi}_\ell) < \overline{m}'_\ell(\check{\pi}_\ell)$ and $\check{m}'(\check{\pi}_r) > \overline{m}'_r(\check{\pi}_r)$, Lemma G.3 and Corollary G.1, applied separately to the left and right of $\hat{\pi}$, yield a continuous function $v :]0, 1[\rightarrow \mathbb{R}$ which solves (G.4) on $]0, 1[-\{\hat{\pi}\}$ with $m \leq v \leq \overline{m}_\ell$ on $]0, \hat{\pi}[$, $m \leq v \leq \overline{m}_r$ on $]\hat{\pi}, 1[$, and $m \leq v \leq \check{m}$ on $[\check{\pi}_\ell, \check{\pi}_r]$. This function extends continuously to the boundaries of $[0, 1]$, and the same argument as in the proof of Proposition G.1 shows that the first and second of these inequalities are strict on $]0, 1[-\{\hat{\pi}\}$. \blacksquare

Proposition G.4 *Suppose that \hat{q} lies in the interior of Q^m and is equal to q_c . Then there are positive constants c_3 , c_4 and c_5 such that for all $r \geq 0$, $\Lambda \geq 0$ and $\sigma > 0$ satisfying $c_3r + c_4\Lambda \leq c_5/\sigma^2$, there exists a continuous function $v : [0, 1] \rightarrow \mathbb{R}$ with the following properties: v is once continuously differentiable and satisfies $m < v < \overline{m}$ on $]0, 1[$; $v(0) = m(0)$ and $v(1) = m(1)$; and v solves (G.4) on $]0, 1[-\{\hat{\pi}\}$.*

PROOF: Choose a strictly convex function $\underline{m} : [0, 1] \rightarrow \mathbb{R}$ with the following properties: $\underline{m}(0) = m(0)$ and $\underline{m}(1) = m(1)$; $\underline{m} = m$ on some intervals $[0, \pi_\ell]$ and $[\pi_r, 1]$ with $0 < \pi_\ell < \hat{\pi} < \pi_r < 1$; $\underline{m} > m$ on $]\pi_\ell, \pi_r[$; \underline{m} has a continuous first derivative on $[0, 1]$ and a continuous second derivative on $[1, 0] - \{\pi_\ell, \pi_r\}$.³⁰ Define $\underline{G}(\pi) = G(\pi, \underline{m}(\pi))$. Next, set

$$c_3 = \max_{\pi} \underline{G}(\pi), \quad c_4 = \sup_{\pi \in]\pi_\ell, \pi_r[-\{\hat{\pi}\}} [f(\pi)\underline{G}(\pi) + (\pi - \tilde{\pi})\underline{G}'(\pi)]$$

and

$$c_5 = \Delta\beta^2 \min_{\pi_\ell \leq \pi \leq \pi_r} \frac{\pi^2(1-\pi)^2 \underline{m}''(\pi)}{2\beta(\pi)}.$$

The constants c_3 and c_5 are clearly positive. As to c_4 , there is at least one belief π^\dagger in $]\pi_\ell, \pi_r[$ such that $(\pi^\dagger - \tilde{\pi})\underline{G}'(\pi^\dagger) \geq 0$, hence $c_4 \geq f(\pi^\dagger)\underline{G}(\pi^\dagger) > 0$. (The positivity of the function f follows from an equation given in the previous proof by setting $\Delta\beta = 0$.) Moreover, c_4 is finite since \underline{G} has finite one-sided derivatives at $\hat{\pi}$.

Now let $r \geq 0$, $\Lambda \geq 0$ and $\sigma > 0$ be such that $c_3r + c_4\Lambda \leq c_5/\sigma^2$. By construction, this implies that \underline{m} is a subsolution of (G.4) both to the left and to the right of the central ray \mathcal{R}_c . As $\hat{q} = q_c$, this ray is vertical at $\pi = \hat{\pi}$. Arguing as in the proof of Lemma G.4, we see that the right-hand side of the ODE is regular with respect to \underline{m} and \overline{m} on each closed interval contained in $]0, \hat{\pi}[$ or $]\hat{\pi}, 1[$, where it is understood that the appropriate one-sided limit is used to calculate the right-hand side of the ODE at $\hat{\pi}$. The result thus follows from Lemma G.2 and Corollary G.1. \blacksquare

³⁰For example, define $\phi(\pi) = (\pi - \pi_\ell)^2(\pi - \pi_r)^2$ on $]\pi_\ell, \pi_r[$ and $\phi(\pi) = 0$ everywhere else; then $\underline{m}(\pi) = [1 + \delta\phi(\pi)]m(\pi)$ will have the desired properties for $\delta > 0$ sufficiently small.

H Numerical Simulations

The adjusted value function can be calculated approximately as a numerical solution to a two-point boundary value problem, namely the ODE (20) subject to the boundary conditions $v^*(0) = m(0)$ and $v^*(1) = m(1)$. We used the method of *relaxation*³¹ to do this. Beliefs were discretised with a step size of 10^{-3} , decreasing to 10^{-5} around the confounding belief. The iterative procedure was deemed to have converged when the maximum pointwise difference between successive approximations to the value function and its first derivative were less than 0.0001%. Convergence was quite rapid, varying from 5 iterations for a high discount rate without switching, to 18 iterations for a low discount rate with an intermediate switching intensity close to the critical level. The procedure was implemented on a VAX minicomputer under VMS v5.4. Each iteration took approximately 19 seconds of CPU time, so the numerical solutions each took between only 1.5 and 6 minutes to calculate.

Given a numerical approximation to the adjusted value function, the optimal policy correspondence immediately yields an approximately optimal policy function. To generate sample paths of posterior beliefs and optimal quantities, we first chose an initial state and an initial belief. One iteration then consisted of the following steps: (a) calculate the optimal quantity given the current belief (using the above numerical results); (b) introduce a shock; (c) update the belief using equation (4) in its discrete form, namely

$$\begin{aligned} \delta\pi_t = \lambda(\pi_t) \delta t &+ \sigma^{-2} \pi_t (1 - \pi_t) (k_t - \pi_t) (\Delta\alpha - \Delta\beta q_t)^2 \delta t \\ &+ \sigma^{-1} \pi_t (1 - \pi_t) (\Delta\alpha - \Delta\beta q_t) \delta Z_t; \end{aligned}$$

(d) update the state if required (depending on the transition probabilities λ_0 and λ_1). These four steps are then repeated to generate a succession of beliefs and quantities. State switching was implemented by repeatedly drawing a number from the uniform distribution on the unit interval (all the examples reported in the paper have $\hat{\pi} = 0.5$, that is $\lambda_0 = \lambda_1 = \Lambda/2$). If the number drawn is less than $1 - \exp(-\Lambda/2)$, then the state remains unchanged, else it switches. Over a time interval of 100, we expect to see 10 switches for $\Lambda = 0.2$. For other values of Λ , the time interval is ‘stretched’ accordingly, so for $\Lambda = 0.05$, for example, we expect these 10 switches to occur by the time $t = 400$.

The shocks were generated by repeated draws from the standard normal distribution. For given time increment δt , the shock δZ was taken to be $\sqrt{\delta t}$ times the draw from the standard normal distribution.

In order to maintain a reasonable approximation to the continuous case that we are modelling, we must ensure that each $\delta\pi$ is not so large that the agent’s belief can jump to (or past) 0, 1, or $\hat{\pi}$. To achieve this, the time variable was incremented by 0.05 in each discrete period, i.e. $\delta t = 0.05$. (This means that in the graphs illustrating the cases without state switching there are several hundred iterations, and in those with state switching there are a few thousand.)

References

- AGHION, P., P. BOLTON, C. HARRIS, and B. JULLIEN (1991): “Optimal Learning by Experimentation,” *Review of Economic Studies*, 58, 621-654.
- BALVERS, R.J., and T.F. COSIMANO (1990): “Actively Learning about Demand and the Dynamics of Price Adjustments,” *Economic Journal*, 100, 882-898.
- BALVERS, R.J., and T.F. COSIMANO (1993): “Periodic Learning about a Hidden State Variable,” *Journal of Economic Dynamics and Control*, 17, 805-827.
- BALVERS, R.J., and T.F. COSIMANO (1994): “Inflation Variability and Gradualist Monetary Policy,” *Review of Economic Studies*, 61, 721-738.
- BERGEMANN, D., and J. VÄLIMÄKI (1996): “Market Diffusion with Two-Sided Learning,” working paper, Yale University and Northwestern University.

³¹See Press *et al.* (1986), Chapter 16.

- BERNFELD, S.R., and V. LAKSHMIKANTHAM (1974): *An Introduction to Nonlinear Boundary Value Problems*. New York and London: Academic Press.
- BOLTON, P., and C. HARRIS (1993): "Strategic Experimentation," STICERD Discussion Paper No. TE/93/261, London School of Economics.
- DUTTA, P.K. (1991): "What Do Discounted Optima Converge to? A Theory of Discount Rate Asymptotics in Economic Models," *Journal of Economic Theory*, 55, 64-94.
- EASLEY, D., and N.M. KIEFER (1988): "Controlling a Stochastic Process with Unknown Parameters," *Econometrica*, 56, 1045-1064.
- FELLI, L., and C. HARRIS (1996): "Learning, Wage Dynamics and Firm-Specific Human Capital," *Journal of Political Economy*, 104, 838-868.
- FLEMING, W.H., and R.W. RISHEL (1975): *Deterministic and Stochastic Control*. New York: Springer-Verlag.
- HARRIS, C. (1988): "Dynamic Competition for Market Share: An Undiscounted Model," working paper, Nuffield College, Oxford.
- KARLIN, S., and H.M. TAYLOR (1981): *A Second Course in Stochastic Processes*. New York: Academic Press.
- KIEFER, N.M. (1989a): "A Value Function Arising in the Economics of Information," *Journal of Economic Dynamics and Control*, 13, 201-223.
- KIEFER, N.M. (1989b): "A Dynamic Model of Optimal Learning with Obsolescence of Information," CAE working paper 89-14, Cornell University; revised 1991.
- KRYLOV, N.V. (1980): *Controlled Diffusion Processes*. New York: Springer-Verlag.
- LIPTSER, R.S., and A.N. SHIRYAYEV (1977): *Statistics of Random Processes I*. New York: Springer-Verlag.
- MCLENNAN, A. (1984): "Price Dispersion and Incomplete Learning in the Long Run," *Journal of Economic Dynamics and Control*, 7, 331-347.
- MIRMAN, L.J., L. SAMUELSON, and A. URBANO (1993): "Monopoly Experimentation," *International Economic Review*, 34, 549-564.
- PRESS, W.H., B.P. FLANNERY, S.A. TEUKOLSKY, and W.T. VETTERLING (1986): *Numerical Recipes: the Art of Scientific Computing*. Cambridge and New York: CUP.
- REVUZ, D., and M. YOR (1991): *Continuous Martingales and Brownian Motion*. New York: Springer-Verlag.
- ROGERS, L.C.G., and D. WILLIAMS (1987): *Diffusions, Markov Processes and Martingales, Vol.2: Itô Calculus*. Chichester: Wiley.
- ROTHSCHILD, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9, 185-202.
- RUSTICHINI, A., and A. WOLINSKY (1995): "Learning about Variable Demand in the Long Run," *Journal of Economic Dynamics and Control*, 19, 1283-1292.

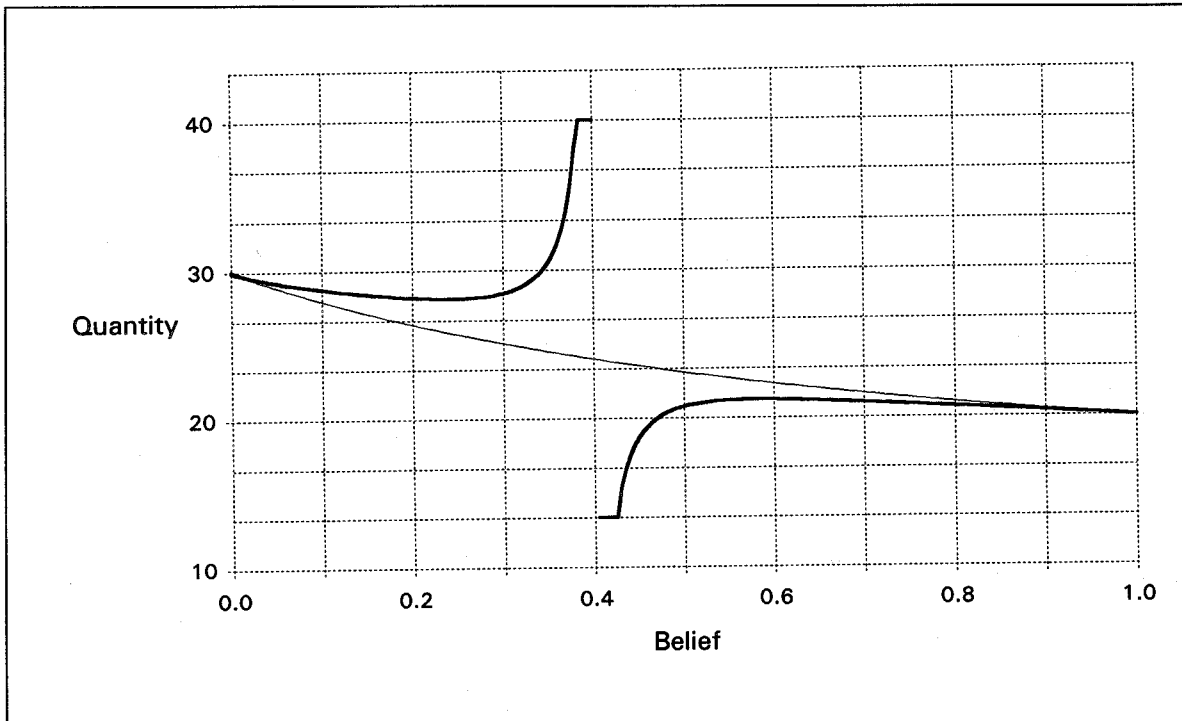
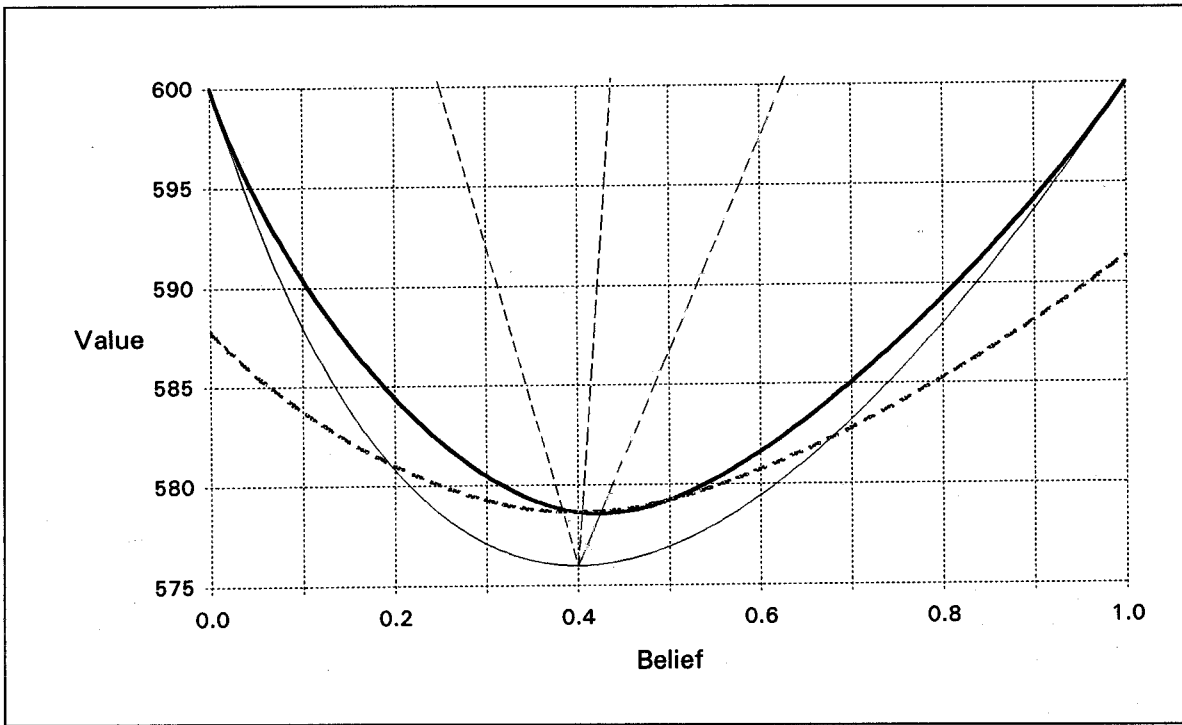


Figure 3: Value function & optimal policy for $r = 0.1$, $\Lambda = 0.05$, $\tilde{\pi} = 0.5$; $\hat{\pi} = 0.4$

The bold lines are the adjusted value function v^* and the optimal policy function q^* .

The thin lines are the myopic optimum pay-off m and the myopic policy q^m .

The upper panel shows the three rays \mathcal{R}_l , \mathcal{R}_c and \mathcal{R}_r .

In the upper panel, the value function u^* is plotted as the dashed line.

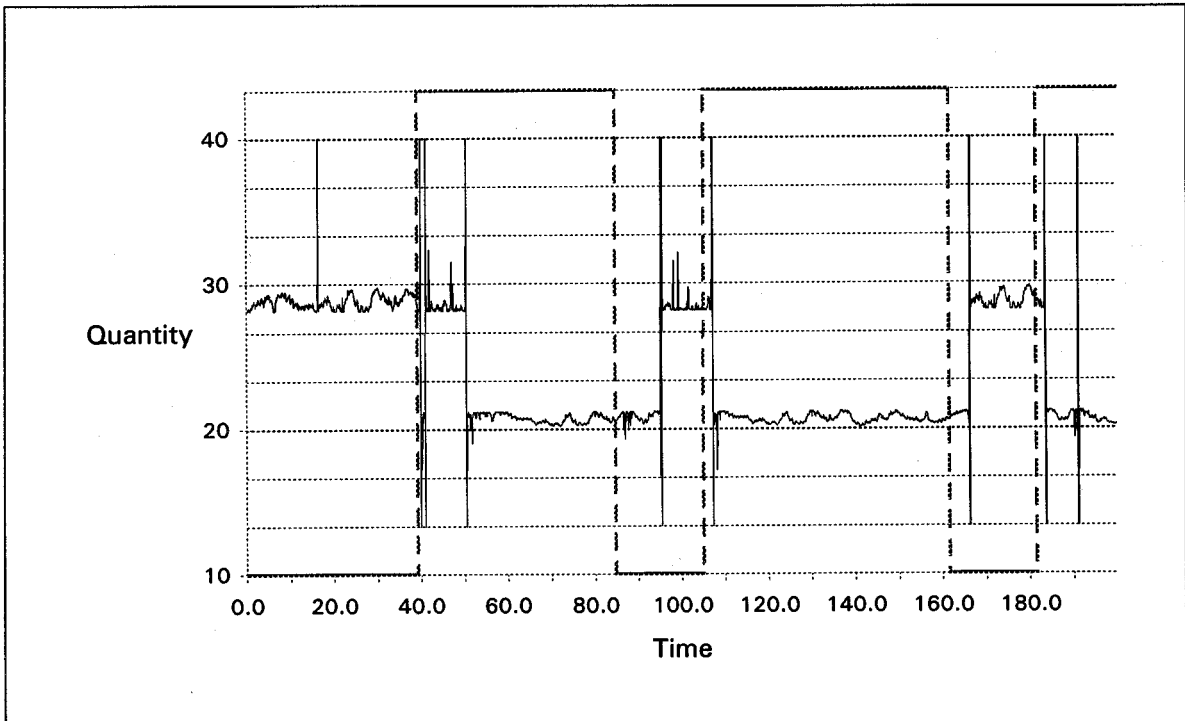
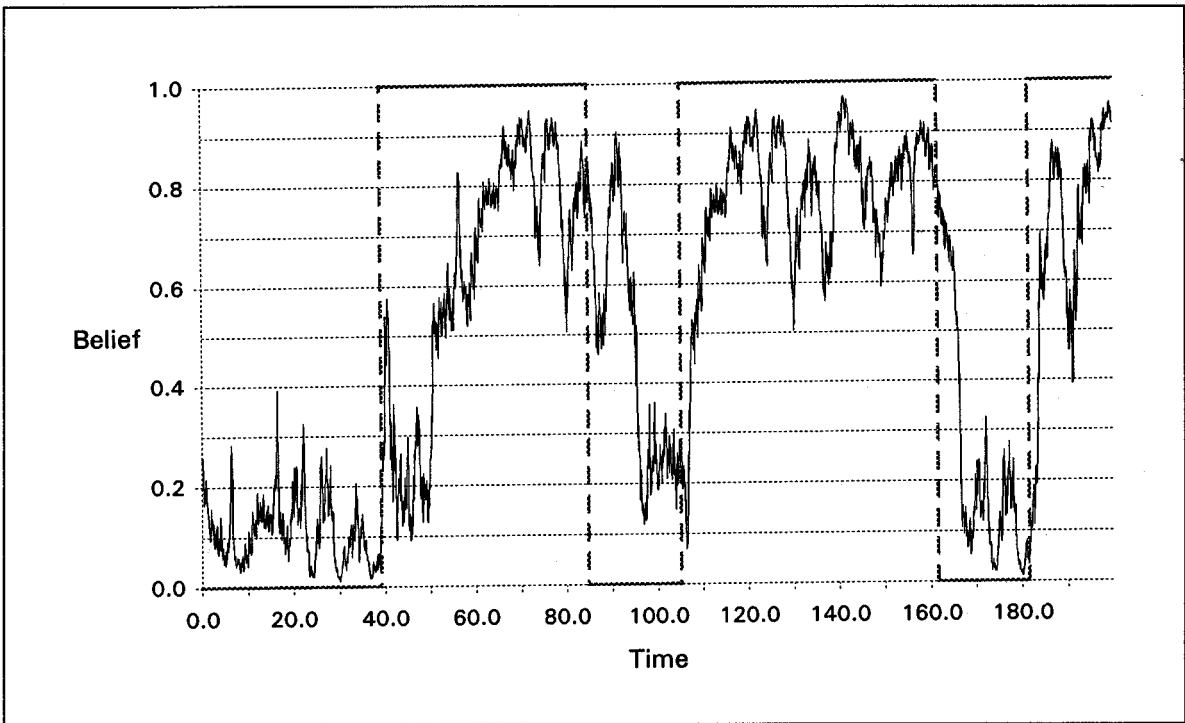


Figure 4: Sample paths for $r = 0.1$, $\Lambda = 0.05$, $\tilde{\pi} = 0.5$; $\hat{\pi} = 0.4$

The dashed line indicates the state switches.

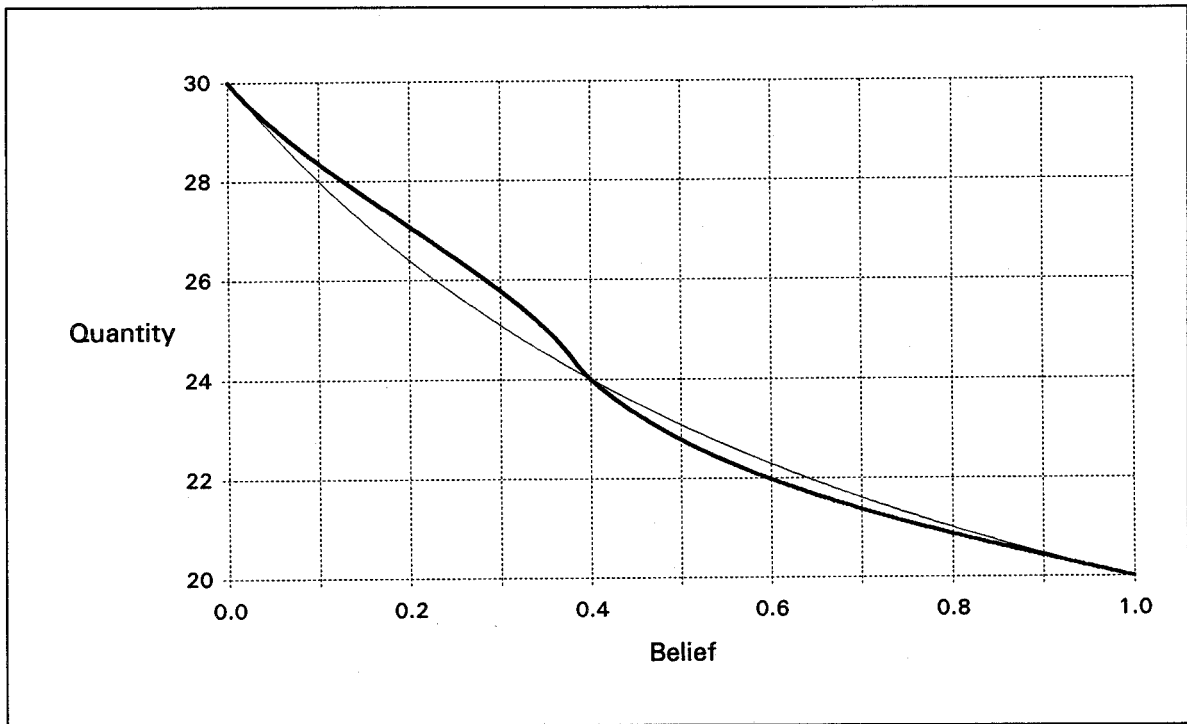
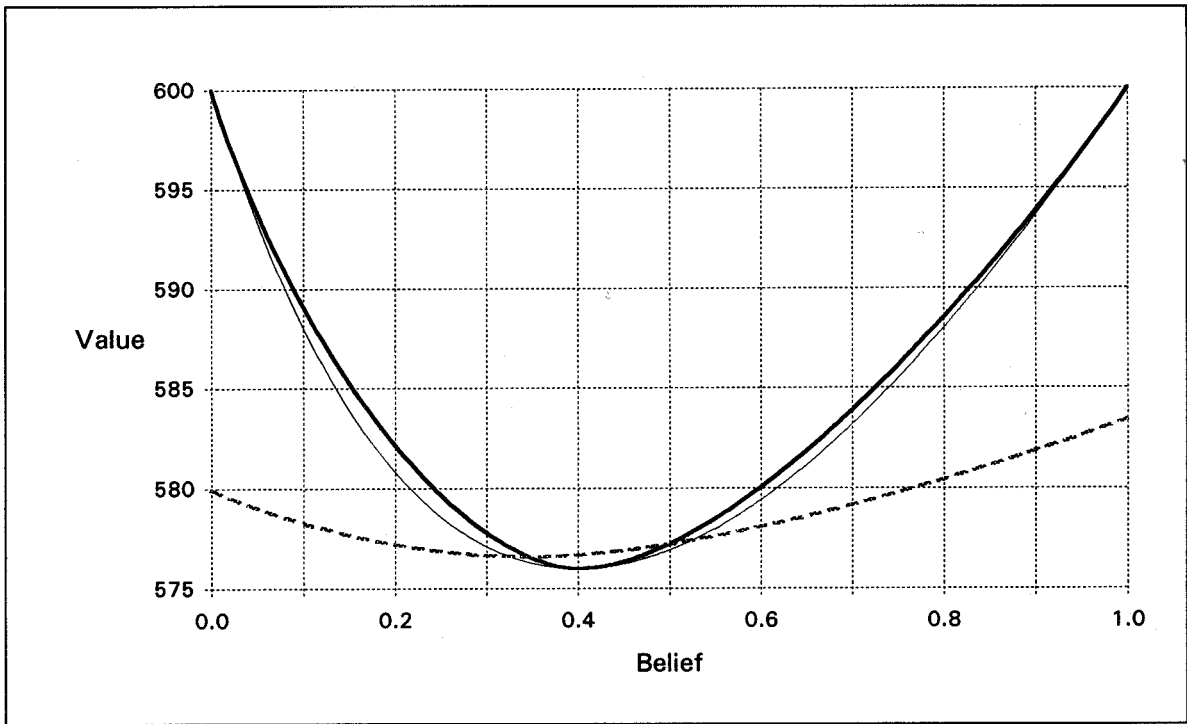


Figure 5: Value function & optimal policy for $r = 0.1$, $\Lambda = 0.2$, $\tilde{\pi} = 0.5$; $\hat{\pi} = 0.4$

The bold lines are the adjusted value function v^* and the optimal policy function q^* .
 The thin lines are the myopic optimum pay-off m and the myopic policy q^m .
 In the upper panel, the value function u^* is plotted as the dashed line.

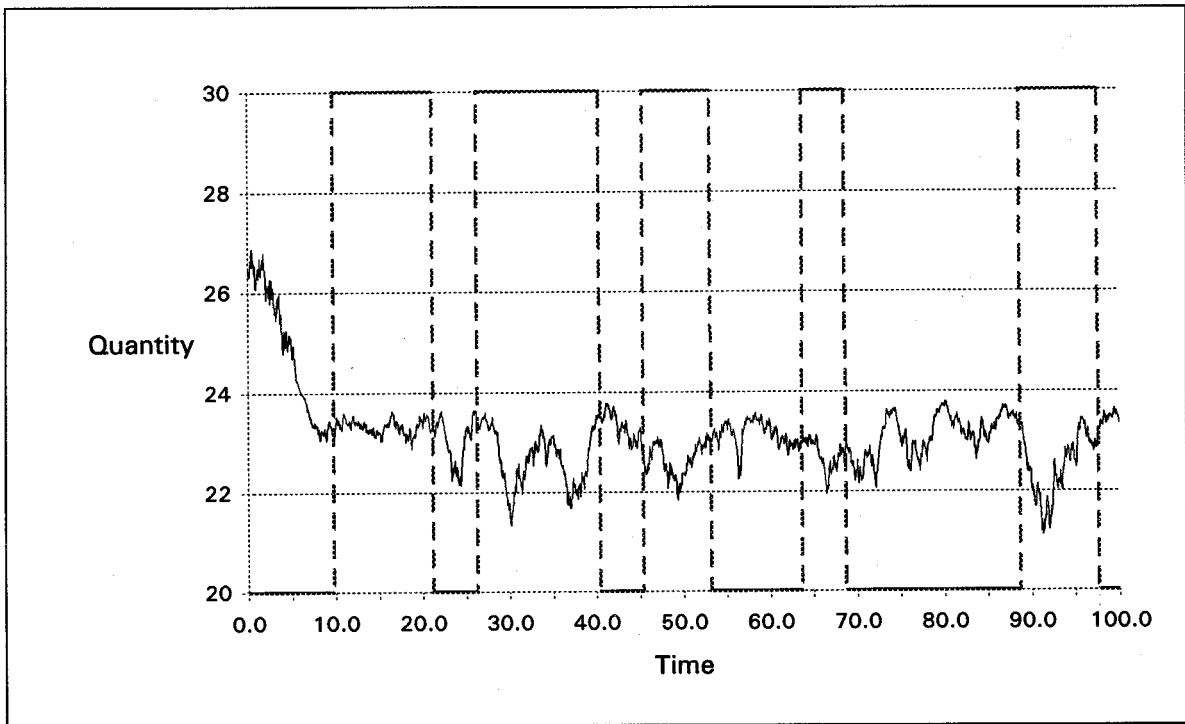
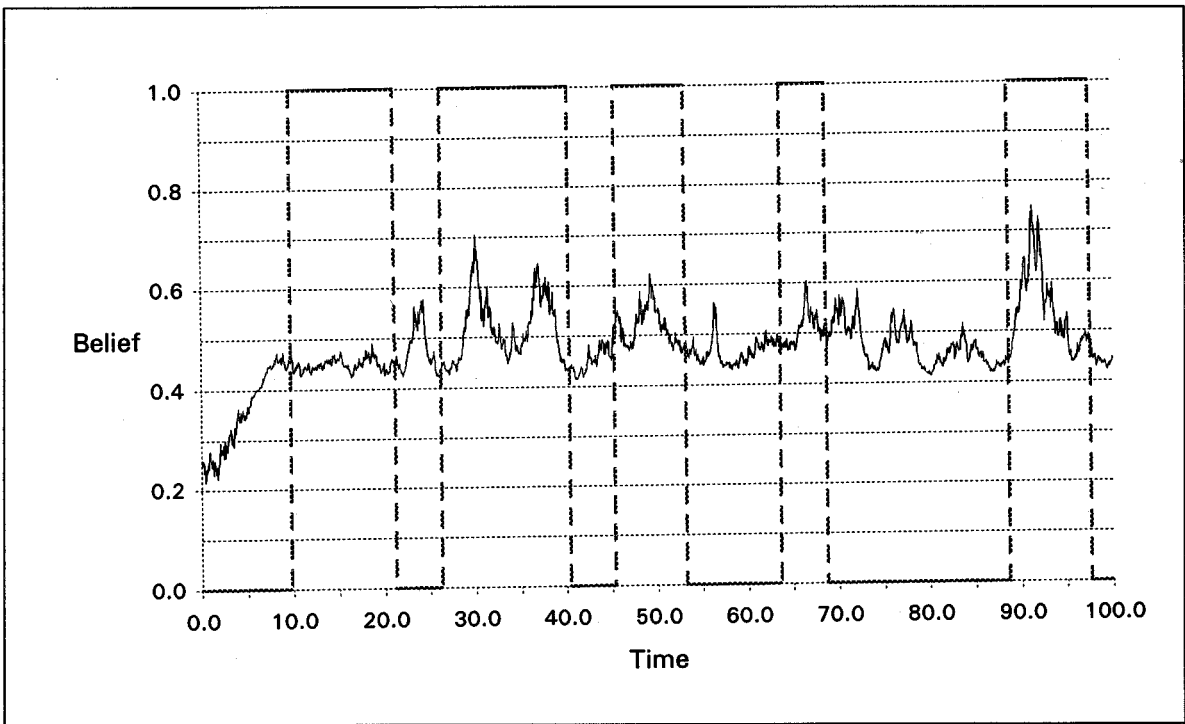


Figure 6: Sample paths for $r = 0.1$, $\Lambda = 0.2$, $\tilde{\pi} = 0.5$; $\hat{\pi} = 0.4$

The dashed line indicates the state switches.