THE DEMAND FOR INFORMATION: MORE HEAT THAN LIGHT

By

Jussi Keppo, Giuseppe Moscarini and Lones Smith

January 2005

COWLES FOUNDATION DISCUSSION PAPER NO. 1498

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

The Demand for Information: More Heat than Light[*]

Jussi Keppo
University of Michigan
IOE Department

Giuseppe Moscarini
Yale University
Economics Department
and Cowles Foundation

Lones Smith
University of Michigan
Economics Department

This version: January 24, 2005

Abstract

This paper produces a comprehensive theory of the value of Bayesian information and its static demand. Our key insight is to assume 'natural units' corresponding to the sample size of conditionally i.i.d. signals — focusing on the smooth nearby model of the precision of an observation of a Brownian motion with uncertain drift. In a two state world, this produces the heat equation from physics, and leads to a tractable theory. We derive explicit formulas that harmonize the known small and large sample properties of information, and reveal some fundamental properties of demand:

- Value 'non-concavity': The marginal value of information is initially zero.
- The marginal value is convex/rising, concave/peaking, then convex/falling.
- 'Lumpiness': As prices rise, demand suddenly suddenly chokes off (drops to 0)
- The minimum information costs on average exceed 2.5% of the payoff stakes
- Information demand is hill-shaped in beliefs, highest when most uncertain
- Information demand is initially elastic at interior beliefs
- Demand elasticity is globally falling in price, and approaches 0 as prices vanish.
- The marginal value vanishes exponentially fast in price, yielding log demand.

Our results are exact for the Brownian case, and approximately true for weak discrete informative signals. We prove this with a new Bayesian approximation result.

*We acknowledge useful suggestions of Paavo Salminen and Xu Meng, and the comments from the theory seminar at the University of Toronto and Georgetown University. Lones thanks the NSF for financial support.

1 Introduction

Information acquisition is an irreversible process. One cannot return to the pristine state of ignorance once apprised any given fact. Heat dissipation also obeys the arrow of time: The heat equation in physics describing its transition is not time symmetric. This paper begins with an observation that this link is not merely philosophical. In static models of Bayesian information acquisition, the value function of beliefs and the quantity of information acquired obeys an inhomogeneous form of the heat equation. We show that a nonlinear transformation of the value function and beliefs exactly obeys the heat equation. This paper exploits this and a parallel insight and crafts the first global theory of the value of information and its demand. For a binary state world, we derive explicit formulas that provide the bigger picture on the famous nonconcavity of information, and the unique demand curve that it induces: We characterize the "choke-off demand" level, and also make many novel findings — eg., demand elasticity is monotonically falling to zero.

Information is an important good, and lies at the heart of most innovative research in decision-making, game theory, and general equilibrium analysis. Yet information is also a poorly understood good. This first of all owes to a lack of natural units. Blackwell's Theorem only considers one signal, for instance. We thus start by measuring information in its 'natural units' corresponding to signal sample sizes, or equivalently, the precision of a normally distributed signal. This is the foundation of our entire theory.

Our first insight is really a technical one that opens the door to our theory. We produce in Lemmas 1–4 a transformation jointly of time and beliefs yielding a detrended log likelihood ratio process. This process is the unique one sharing two critical characteristics: First, it has unit diffusion coefficient (variance). Second — and much more subtly — it allows us to change measure to produce a Wiener process without adding a new stochastic process. This immediately delivers to us in Lemma 5 a simple transition law for beliefs. This density has the key property that it is proportional to its time derivatives (Lemma 6).

Underlying everything is the standard $\vee$-shaped payoff function of the belief in the high state: A decision maker takes the low action left of a cross-over belief, and the high action right of it. It so turns out that the ex post value of information is a multiple of the payoff to a call option whose strike price is the cross-over belief. With this insight, it is not

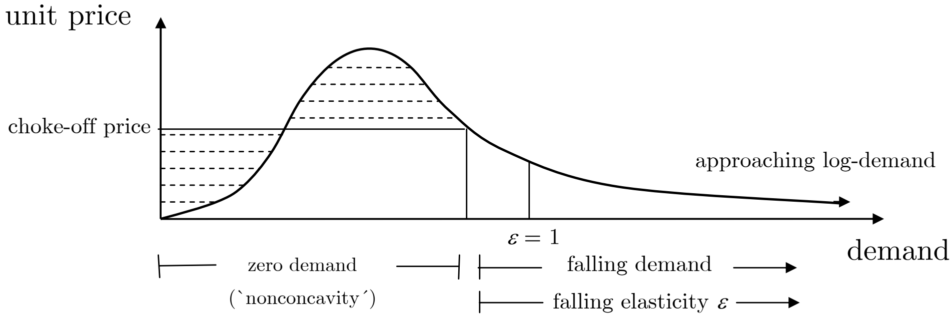

Figure 1: **Marginal Value of Information.**

surprising that our analysis traces the two routes used to price options: the heat equation approach of Black and Scholes (1973) and the martingale measure tack of Harrison and Kreps (1979). Yet the belief process is less tractable than the geometric Brownian motion assumed for asset prices, and our transformation therefore more complex.

Using the original belief process, the payoff function of beliefs and information demand obeys an inhomogeneous heat equation (Lemma 7). In our critical innovation, we perform both a change of variables, blending time and log-likelihood ratios, and a nonlinear transformation of payoffs to produce the standard heat equation. We then exploit the solution of the heat equation, and then in Lemma 8 via martingale methods. Either way, Theorem 1 explicitly expresses the value of information in terms of the normal distribution.

We then turn to our substantive findings. Theorem 2 expresses the marginal value of information in terms of the derivatives of the transition belief density. This reduces analysis of the value derivative to a polynomial in the reciprocal demand. Using this, Corollary 1 finds that the marginal value is initially zero — the nonconcavity — as found in Radner-Stiglitz (1984) [RS], and rigorously formalized in Chade-Schlee (2002) [CS]. The sufficient conditions in CS for this result do not encompass our model. By Theorem 3, the marginal value convexly rises, is then concave and hill-shaped, and finally convex and falling — as in Figure 1. We compute demand at the peak marginal value of information.

So with linear prices, information demand vanishes at high prices, before jumping to a positive level (Theorem 4) strictly below the peak marginal value. The Law of Demand then kicks in, and the demand schedule thereafter coincides with the marginal value schedule. One never buys just a little information. Theorem 5 quantifies the nonconcavity, determining the minimal purchase on average to be 2.5% of the expected payoff stakes.

In Theorem 6, we find that information demand is hill-shaped in beliefs, unlike the dynamic model with impatience of Moscarini and Smith (2001) [MS1]. Furthermore, it jumps down to 0 near beliefs 0 and 1, when the choke-off demand binds. Here we discover an interesting contrast with sequential learning models, because our thresholds owe to global optimality considerations. Completing this picture, Theorem 7 finds that information demand is hill-shaped in beliefs (quasi-concave, precisely) — opposite to MS1.

A novel topic we explore is the demand elasticity. Theorem 8 asserts that information demand is initially elastic at interior beliefs; the elasticity is globally falling in price, and is vanishing in the price. This analysis exploits all of our analytic structure.

We finally revisit in Theorem 9 the large demand analysis of Moscarini and Smith (2002) [MS2] now quickly via our explicit formulas rather than large deviation theory. MS2 also measure information by sample size, but assumed cheap discrete information, and not our continuous or weak discrete signals. The marginal value of information eventually vanishes exponentially fast, producing the logarithmic demand of MS2 at small prices. We sharpen the demand approximation, and find that it is monotonically approached.

Our Gaussian information is generated by the time that one observes a Brownian motion with state-dependent drift, but state-independent diffusion coefficient. Consider a discrete model where the decision-maker chooses how many conditional iid signals to draw. Theorem 10 shows that Bayes' updating weakly approaches the continuous time Brownian filter as the signal strength vanishes, and quantity is rescaled downwards. We show that garbled signals have precisely this property. We next apply the Theorem of the Maximum, finding in Theorem 11 that the demand formulas and value of information approximate the discrete formulas for small bits of information (weak signals).

The experimentation literature aside, we know of one related information demand paper. In Kihlstrom (1974), a consumer faces a linear price for precision of a Gaussian signal given a (conjugate) Gaussian prior. For a specific hyperbolic utility function, he can write utility as a function of signals and avoid computing the density of posteriors.

We next lay out and develop the model, and very rapidly progress through the results on beliefs, the value and marginal value of information, demand, and weak discrete signals.

2 The Model

A. The Decision Problem. Assume a one-shot decision model, where a decision maker (DM) chooses how much information to buy, and then acts. For simplicity, we assume two actions A, B, whose payoffs $\pi_A^\theta, \pi_B^\theta$ depend on the state $\theta = L, H$. Action B is best in state H and action A is best in state L: $0 \le \pi_A^H < \pi_B^H$ and $\pi_A^L \ge \pi_B^L \ge 0$.[1] Since the DM has prior beliefs $q \in (0,1)$, the convex $\vee$-shaped expected payoff function is

$$u(q) = \max\langle q\pi_A^H + (1-q)\pi_A^L, q\pi_B^H + (1-q)\pi_B^L\rangle \equiv \max\langle \pi_A^L + mq, \pi_B^L + Mq\rangle \qquad (1)$$

thereby defining $M = \pi_A^H - \pi_A^L$ and $m = \pi_B^H - \pi_B^L$. We assume no dominated actions, so that payoffs have a kink at a *cross-over belief* $\hat{q} = (\pi_A^L - \pi_B^L)/(M-m) \in (0,1)$.

Notice that the maximum *payoff stakes* here are $(\pi_B^H - \pi_A^H) + (\pi_A^L - \pi_B^L) = M - m > 0$. The DM never incurs a payoff loss greater than $M - m$ from making a wrong action choice; this bound is tight when either difference $(\pi_B^H - \pi_A^H) \ge 0$ or $(\pi_A^L - \pi_B^L) \ge 0$ vanishes.

B. The Standard Information Acquisition Model. Given is a probability space $(\Omega, \mathcal{F}, P)$, were Ω is a set, $\mathcal{F}$ a σ-algebra of subsets of Ω, and P a probability measure on $\mathcal{F}$. This space captures all uncertainty, including the state of the world $\theta = L, H$.

Before making a decision, the DM can obtain information of any level $t \ge 0$ about the state θ. While more information could plausibly connote better quality information, we specifically mean that the DM with information level t_2 knows strictly more about the state of the world than does the DM with information level $t_1 \le t_2$. So assume a filtration $\{\mathcal{F}_t : t \in [0,\infty)\}$, so that the σ-algebras are nested $\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2} \subseteq \mathcal{F}$ when $t_1 < t_2$. The DM observes $\mathcal{F}_t$, updates her beliefs to $q(t) = P(H|\mathcal{F}_t)$ from the prior $q(0) = q$.

The *ex ante expected payoff* (prior to seeing $\mathcal{F}_t$) is $u(t,q) = E[u(q(t))|q(0) = q]$, and the *value of information* is $v(t,q) = u(t,q) - u(q)$ — namely, the expected increase in utility from observing $\mathcal{F}_t$. Faced with a constant marginal cost $c > 0$ of information, the DM can choose the observation 'time' t (namely, the demand level) at cost ct. The net payoff given the level t is $v(t,q) - ct$. This is maximized by choosing the information level $\tau(c,q) > 0$, which is our demand schedule. Finally, the DM chooses the best action.

[1] Further, it is without loss of generality to assume for simplicity that $\pi_A^H = 0$, since the decision must be made. An analogous choice of $\pi_B^H = 1$ is not allowed later on, without also scaling the cost function.

C. The Natural Continuous Units of Information. We actually have a specific filtration in mind. Let the DM observe the time-t realization of a process $X(\cdot)$ with drift $\pm\mu$ in states H, L, respectively, and constant diffusion coefficient $\sigma > 0$. Thus, the signal is twice as informative when t doubles — just as is true for the sample size of conditionally iid signals. We show in Section 7 that this approximates discrete bit sampling models.

By Theorem 9.1 in Liptser and Shirayev (2001), when observing the realizations of the Brownian Motion $X(t)$ in continuous time, the belief $\tilde{q}(t) = P(H|\mathcal{F}_t)$ obeys the Bayes filter $d\tilde{q}(t) = \zeta\tilde{q}(t)(1-\tilde{q}(t))dW(t)$, where $\zeta \equiv 2\mu/\sigma$ is the *signal/noise ratio*, and $W(\cdot)$ is a standard Wiener process w.r.t. the measure P (i.e. unconditional on $\theta = H, L$). Notice that if we define $q(t) = \tilde{q}(t/\zeta^2)$, then[2] $(dq(t))^2 = (d\tilde{q}(t))^2/\zeta^2 = q(t)^2(1-q(t))^2 dt$, and thus

$$dq(t) = q(t)(1-q(t))dW(t). \tag{2}$$

We henceforth set $\zeta = 1$ and compute the time (demand) $\hat{t}$ with any $\hat{\zeta} > 0$ from $\hat{t} = t/\hat{\zeta}^2$.

3 Beliefs and Log Likelihood Ratios

We begin by describing the limit behavior of beliefs $q(\cdot)$.

Lemma 1 (Long Run Beliefs) *The belief process in* (2) *satisfies*

$$\begin{aligned} &(a) \qquad P\left[\inf\{t \geq 0 \mid q(t) = 0 \text{ or } 1\} = \infty\right] = 1 \\ &(b) \qquad P\left[\lim_{t\to\infty} q(t) = 0\right] = 1 - P\left[\lim_{t\to\infty} q(t) = 1\right] = 1 - q. \end{aligned}$$

So the probability that $q(t) \notin (0,1)$ in finite time is zero and $q(\infty) = 0$ or 1.

The proof of (a) is in the appendix: $q(\cdot)$ avoids the boundary as the diffusion coefficient in (2) vanishes quickly near $q(\cdot) = 0, 1$. Part (b) owes to the martingale property.

Our objective is to derive from posterior beliefs a tractable process that contains the same information. In particular, we aim for a simple standard Wiener process. First, in Lemma 2, we find a monotone transformation of posterior beliefs which has a unit diffusion coefficient (variance); it turns out that this is unique. Second, in Lemma 4, we change

[2] To justify the first inequality, $E[W^2(t/\zeta^2)] = t/\zeta^2$ and thus $dW^2(t/\zeta^2) = dt/\zeta^2$.

probability measure so that this transformation retains the martingale property. There is a degree of freedom here, which we resolve in Lemma 3 on grounds of tractability.

Lemma 2 (Likelihood Ratios) *Let $z(t) = \lambda(t, q(t))$, where $\lambda \in C^2$. If the diffusion coefficient of $z(\cdot)$ is one then $\lambda(t, q(t)) = A(t) + \log\left(\frac{q(t)}{1-q(t)}\right)$, where $|A(t)| < \infty$ for all t.*

Proof: Using Ito's lemma we get

$$dz(t) = \left(\lambda_t + \tfrac{1}{2}\lambda_{qq}q(t)^2(1-q(t))^2\right)dt + \lambda_q q(t)(1-q(t))dW. \tag{3}$$

Solving $\lambda_q q(t)(1-q(t)) = 1$ yields $\lambda(t, q(t)) = A(t) + \log\left(\frac{q(t)}{1-q(t)}\right)$. □

This lemma is intuitive, since Bayes rule is multiplicative in likelihood ratios, and therefore additive in log-likelihood ratios. Substituting from (3), we then find

$$dz(t) = \left(A'(t) - \frac{1-2q(t)}{2}\right)dt + dW(t) \equiv \nu(t)dt + dW(t), \tag{4}$$

where we have implicitly defined $\nu(t)$. Next, define the probability measure Q on $(\Omega, \mathcal{F}_t)$ by the Radon-Nikodym derivative:

$$\frac{dQ}{dP} = R(t) = \exp\left(-\frac{1}{2}\int_0^t \nu^2(s)ds - \int_0^t \nu(s)dW(s)\right). \tag{5}$$

Lemma 3 (Radon-Nikodym Derivative) *$R(t) = q(t)/q$ iff $A(t) = -t/2$.*

This result is important as it does not introduce a new stochastic process. With a different R-N derivative, we would have two imperfectly correlated stochastic processes $q(\cdot)$ and $R(\cdot)$, and derivation of our results would become exceedingly difficult. A unique change of measure maintains the uni-dimensionality of the stochastic process.

To prove Lemma 3, observe that writing $Y(t) = -\frac{1}{2}\int_0^t(1-q(s))^2ds + \int_0^t(1-q(s))dW(s)$ and guessing $q(t) = qe^{Y(t)}$ correctly yields, using Ito's Lemma, $dq(t) = q(t)dY(t) + \frac{1}{2}q(t)(dY(t))^2 = q(t)(1-q(t))dW(t)$ — namely, our belief filter (2). In other words,

$$q(t) = q\exp\left(-\tfrac{1}{2}\textstyle\int_0^t(1-q(s))^2ds + \int_0^t(1-q(s))dW(s)\right).$$

So motivated, if we set $\nu(s) = q(s) - 1$ in (5), then we get Lemma 3.

Equation (4) implies that $A(t) = -t/2$, which we henceforth assume. Thus,

$$z(t) = \lambda(t, q(t)) = \log\left(\frac{q(t)}{1-q(t)}\right) - \frac{1}{2}t \qquad \Leftrightarrow \qquad q(t) = \ell(t, z(t)) = \frac{1}{e^{-\frac{1}{2}t - z(t)} + 1}.$$

Observe that $z(\cdot)$ is a partially de-trended log-likelihood ratio (logLR). While the logLR of state L to state H is well-known to be a martingale conditional on state H, this is not useful because we do not know state H. We instead *first* subtract the deterministic portion of the drift from the logLR, and *then* change measure from P to the conditional measure Q given $q(t)$. This yields a martingale such that the R-N derivative of Q is $q(t)/q$.

Lemma 4 (Detrended LogLR) *The process $z(\cdot)$ obeys $dz(t) = d\hat{W}(t)$ for all $t \geq 0$, where*

$$\hat{W}(t) = \int_0^t (q(s) - 1)ds + W(t)$$

is a Wiener process under the probability measure Q. Hence, $z(\cdot)$ is a Q-martingale and the pdf for transitions $z \mapsto y$ in time $t > 0$ is given as follows:

$$\frac{1}{\sqrt{t}}\phi\left(\frac{y-z}{\sqrt{t}}\right) = \frac{1}{\sqrt{2\pi t}} e^{\frac{-(y-z)^2}{2t}} \tag{6}$$

for all $(t, z, y) \in (0, \infty) \times \mathbb{R}^2$, where $\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$ is the standard normal pdf.

Equation (6) follows from the facts that (4) can be written as $dz(t) = d\hat{W}(t)$ and $\hat{W}(\cdot)$ is a Q-Wiener process by Girsanov's theorem (see e.g. Øksendal (1998, pp. 155–6)).

We now derive the belief transition pdf $\xi(t, q, r) = \frac{\partial}{\partial r} P(q(t) \leq r | q(0) = q)$, using the above normal transition pdf for transformed log-likelihood ratios $z(t)$ (see Figure 2).

Lemma 5 (Beliefs) *The transition probability density function of beliefs $q(t)$ is given by*

$$\xi(t, q, r) = \frac{q\phi\left(-\frac{1}{2}\sqrt{t} + \frac{1}{\sqrt{t}}L(r, q)\right)}{r^2(1-r)\sqrt{t}} = \sqrt{\frac{q(1-q)}{r^3(1-r)^3 2\pi t}} e^{-\frac{1}{8}t - \frac{1}{2t}L^2(r,q)} \tag{7}$$

for all $(t, q, r) \in (0, \infty) \times (0, 1) \times (0, 1)$, where $L(r, q) = \log\left(\frac{r(1-q)}{(1-r)q}\right)$.

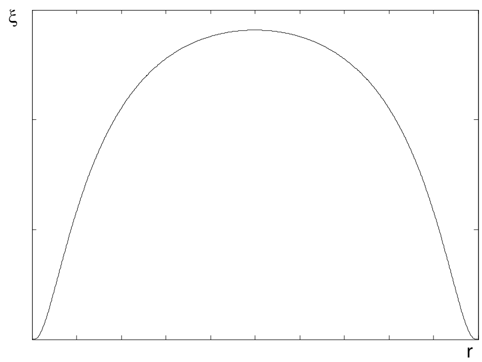

Figure 2: **Transition Probability Function.** We plot the symmetric pdf $\xi(t,q,r)$ for transitions from $q = 0.5$ to any belief r after an elapse time $t = 1$.

Proof: Fix an arbitrary measurable real function $\eta(\cdot)$. Then $\int_0^1 \xi(t,q,r)\eta(r)dr$ equals[3]

$$\begin{aligned} E_q[\eta(q(t))] &= qE_q\left[\frac{q(t)}{q}\frac{\eta(q(t))}{q(t)}\right] = qE_q^Q\left[\frac{\eta(q(t))}{q(t)}\right] = qE^Q_{\lambda(0,q)}\left[\frac{\eta(\ell(t,z(t)))}{\ell(t,z(t))}\right] \\ &= \frac{q}{\sqrt{t}}\int_{-\infty}^{\infty}\phi\left(\frac{z(t)-\lambda(0,q)}{\sqrt{t}}\right)\left(e^{-\frac{1}{2}t-z(t)}+1\right)\eta\left(\frac{1}{e^{-\frac{1}{2}t-z(t)}+1}\right)dz(t) \\ &= \frac{q}{\sqrt{t}}\int_0^1\phi\left(\frac{\lambda(t,r)-\lambda(0,q)}{\sqrt{t}}\right)\frac{\eta(r)}{r}\frac{\partial\lambda(t,r)}{\partial r}dr \\ &= \frac{q}{\sqrt{t}}\int_0^1\frac{1}{r^2(1-r)}\phi\left(\frac{\lambda(t,r)-\lambda(0,q)}{\sqrt{t}}\right)\eta(r)dr. \end{aligned}$$

The paper hereafter repeatedly exploits this belief transition pdf to derive the formula for the value and marginal value of information, as well as the properties of the demand function. In particular, this yields a critical time derivative:

Lemma 6 (t-Derivative) *The belief transition pdf $q(t)$ satisfies, for $0 < q, r < 1$ and $t > 0$:*

$$\xi_t(t,q,r) = \xi(t,q,r)\left[-\frac{1}{8}+\frac{L^2(r,q)}{2t^2}-\frac{1}{2t}\right].$$

4 The Value of Information

4.1 Option Pricing Analogy

Before deriving our value function, it helps to motivate this with a related exercise done in finance. To this end, simplify matters by positing that action A is a safe action yielding

[3] We write $E_q[\cdot] = E[\cdot|q(0) = q]$ and $E_z[\cdot] = E[\cdot|z(0) = z]$.

zero payoff, so that $\pi_A^H = \pi_A^L = 0$. Equation (1) can then be written as follows

$$u(q) = M \max\langle 0, q - \hat{q} \rangle.$$

Here we get immediately the following interpretation for $u(q)/M$: It equals the payoff of a European call option with strike price $\hat{q}$ and underlying asset price q.

Black and Scholes (1973) derive the option pricing formula when the underlying asset follows a geometric Brownian motion. They use an arbitrage argument to deduce a parabolic PDE that reduces to the heat equation after a change of variable. In this transformation, time is understood as time to maturity. But geometric Brownian motion is still far more tractable than our nonlinear belief diffusion (2), and thus only a time rescaling of the range variable is needed. By contrast, we use a more complicated transformation.

Harrison and Kreps (1979) and Harrison and Pliska (1981) later derived the option pricing formula using martingale methods. In our paper, the z-process is a martingale under the measure Q just as the discounted asset price is a martingale under the 'pricing measure' in Harrison and Kreps (1979). Our first approach follows this line of thought, but again, its execution requires a simultaneous range and domain transformation.

4.2 Value Function Derivation

Applying the backward equation to $u(t, q) = E_q[u(q(t))]$ for the driftless belief diffusion process $q(\cdot)$ in (2) yields our key insight into payoffs:[4]

Lemma 7 (Inhomogeneous Heat Equation) *Expected payoffs $u(t, q)$ satisfy*

$$u_t(t, q) = \frac{1}{2} q^2 (1 - q)^2 u_{qq}(t, q) \tag{8}$$

for all $(t, q) \in (0, \infty) \times (0, 1)$ with the initial condition: $u(0, q) = u(q)$.

Notice that (8) implies that the marginal value of information is positive ($u_t(t, q) \geq 0$) iff the value of information is convex in beliefs ($u_{qq}(t, q) \geq 0$). This further motivates the link to the heat equation, as the temperature gradient within a finite bar obeys a

[4]In fact, *all* time derivatives u_{t^n} in Theorem 2 satisfy the inhomogeneous heat equation – an curious truth that we do not exploit.

qualitatively similar law where convexity is critical, as Figure 3 depicts. By the same token, the option value is convex in price, just as expected payoffs are convex in beliefs.

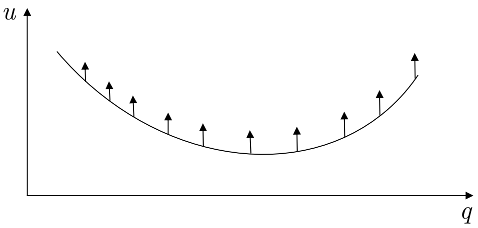

Figure 3: **Analogy with Fourier's Law.** This illustrates Fourier's Law of Heat Conduction — in particular, how the heat flow is locally positive exactly when the heat distribution is locally convex on the bar. Specifically, if $u(t,q)$ is the temperature at position q on a bar, with ends held constant at temperatures $u(0)$ and $u(1)$ respectively, then the temperature is increasing $u_t > 0$ iff $u_{qq} > 0$. This is exactly analogous to the behavior of expected payoffs as we acquire more information.

As the inhomogeneous heat equation is not directly soluble, we proceed as follows. Let us define $z = \log\left(\frac{q}{1-q}\right)$ and transform expected payoffs as $h(t,z) = u\left(t, \frac{1}{1+e^{-z}}\right)(1+e^{-z})$.

Lemma 8 (Stochastic Representation) *Transformed expected payoffs are represented as*

$$h(t,z) = E_z^Q\left[\left(e^{-\frac{1}{2}t-z(t)}+1\right)u\left(\frac{1}{e^{-\frac{1}{2}t-z(t)}+1}\right)\right]. \tag{9}$$

Proof 1: The Martingale Method. We first follow the approach of Harrison and Kreps (1973), and exploit our martingale $z(\cdot)$. Write the posterior expected payoff as

$$u(t,q) = qE_q^P\left[\frac{q(t)}{q}\frac{u(q(t))}{q(t)}\right] = qE_q^Q\left[\frac{u(q(t))}{q(t)}\right] = qE_{\lambda(0,q)}^Q\left[\left(e^{-\frac{1}{2}t-z(t)}+1\right)u\left(\frac{1}{e^{-\frac{1}{2}t-z(t)}+1}\right)\right]$$

using Lemma 4. Now $u(t,q) = qh\left(t, \lambda(0,q)\right)$ gives (9).

Proof 2: The Heat Equation. We now adapt the approach of Black and Scholes (1973). Change the variables in (8) from beliefs q to $Z = \log[q/(1-q)] + t/2 = \lambda(-t,q)$ — where t is understood as the elapse time. Next, define $H(t,Z) = h(t, Z - t/2) = u\left(t, \frac{1}{e^{t/2-Z}+1}\right)\left(e^{t/2-Z}+1\right)$. Then the heat equation obtains:[5] $H_t(t,Z) = \frac{1}{2}H_{ZZ}(t,Z)$.

[5] Simply take the derivatives below and apply $u_t = \frac{1}{2}q^2(1-q)^2u_{qq}$:

$$H_t(t,Z) = \frac{e^{t/2-Z}}{2}u\left(t, \frac{1}{e^{t/2-Z}+1}\right) - \frac{e^{t/2-Z}}{2(e^{t/2-Z}+1)}u_q\left(t, \frac{1}{e^{t/2-Z}+1}\right) + \left(e^{t/2-Z}+1\right)u_t\left(t, \frac{1}{e^{t/2-Z}+1}\right)$$

$$H_{ZZ}(t,Z) = e^{t/2-Z}u\left(t, \frac{1}{e^{t/2-Z}+1}\right) - \frac{e^{t/2-Z}}{e^{t/2-Z}+1}u_q\left(t, \frac{1}{e^{t/2-Z}+1}\right) + \frac{\left(e^{t/2-Z}\right)^2}{(e^{t/2-Z}+1)^3}u_{qq}\left(t, \frac{1}{e^{t/2-Z}+1}\right).$$

We thank Robert Israel of UBC for first pointing out a related transformation.

The solution of the heat equation (e.g. Karatzas and Shreve (1991, page 254)) yields:

$$H(t,Z)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{t}}\phi\left(\frac{y}{\sqrt{t}}\right)\left(e^{-Z-y}+1\right)u\left(\frac{1}{e^{-Z-y}+1}\right)dy.$$

Finally, using $H(t,Z+t/2)=h(t,Z)$ gives the representation (9). □

We now exploit the above stochastic representation to derive the value of information. It follows from Lemma 1 that the long-run limit of the expected payoff is given by

$$\lim_{t\uparrow\infty}u(t,q)=qu(1)+(1-q)u(0)\equiv\bar{u}(q), \tag{10}$$

i.e., we can write $u(\infty,q)=\bar{u}(q)$. Let us define the *full information gap* as follows

$$\mathrm{FIG}(t,q)=u(\infty,q)-u(t,q)=\bar{u}(q)-u(t,q). \tag{11}$$

Thus, FIG is the difference between the expected payoffs with full information and time t information. We now explore the behavior of the value function $v(t,q)$ for finite $t>0$.

Theorem 1 (Value Formula) *The expected payoff is* $u(t,q)=q\pi_B^H+(1-q)\pi_A^L-FIG(q,t)$, *where the full information gap satisfies*

$$FIG(t,q)=q\left(\pi_B^H-\pi_A^H\right)\Phi\left(-\tfrac{1}{2}\sqrt{t}+\tfrac{1}{\sqrt{t}}L(\hat{q},q)\right)-(1-q)\left(\pi_A^L-\pi_B^L\right)\Phi\left(-\tfrac{1}{2}\sqrt{t}-\tfrac{1}{\sqrt{t}}L(\hat{q},q)\right)$$

where $\Phi(\cdot)$ *is the standard normal cdf. The value of information* $v(t,q)=u(t,q)-u(q)$ *is*

$$v(t,q)=\begin{cases} q\left(\pi_B^H-\pi_A^H\right)\Phi\left(\tfrac{1}{2}\sqrt{t}-\tfrac{1}{\sqrt{t}}L(\hat{q},q)\right) \\ \quad -(1-q)\left(\pi_A^L-\pi_B^L\right)\Phi\left(-\tfrac{1}{2}\sqrt{t}-\tfrac{1}{\sqrt{t}}L(\hat{q},q)\right) \quad \forall \quad q\le\hat{q} \\ \\ -q\left(\pi_B^H-\pi_A^H\right)\Phi\left(-\tfrac{1}{2}\sqrt{t}+\tfrac{1}{\sqrt{t}}L(\hat{q},q)\right) \\ \quad +(1-q)\left(\pi_A^L-\pi_B^L\right)\Phi\left(\tfrac{1}{2}\sqrt{t}+\tfrac{1}{\sqrt{t}}L(\hat{q},q)\right) \quad \forall \quad q\ge\hat{q}. \end{cases} \tag{12}$$

The appendix proof uses the fact that in Lemma 8, $e^{-t/2-z(t)}$ is log-normally distributed. Figure 4 illustrates the posterior expected payoff with different information levels.

From Figure 4, we see that the value of information $v(t,q)$ is high when q is close to the cross-over belief $\hat{q}$, and is low when q is close to 0 or 1.

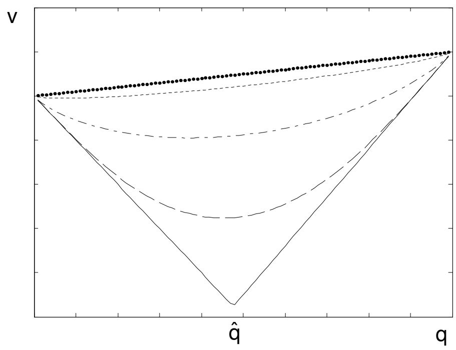

Figure 4: **Posterior expected payoff with different information levels.** The parameter values are: $\hat{q} = 0.48$, $\pi_H^A = 1$, $\pi_L^A = 2$, $\pi_H^B = 2.1$, and $\pi_L^B = 1$. The bottom solid line graphs $\bar{u}(q)$, while the top dotted line is $u(q)$. Between we find $u(1,q), u(5,q), u(15,q)$, respectively the graphs —, – ·,–

The results in our setting are analogous to the properties of call options. The value of information in our setting is like the time value of the option. Informational value is increasing in t just as the time value of the option is increasing in the time to maturity.

5 The Marginal Value of Information

Loosely, information only has positive marginal value if we hit the cross-over belief $\hat{q}$ at time t; its value is then increasing in the payoff slope difference $M - m$ and belief variance $\hat{q}^2(1-\hat{q})^2$. Since $v(t,q) = u(t,q) - u(q)$, the time derivatives of the value of information and expected payoffs coincide, $v_{t^n}(t,q) = u_{t^n}(t,q)$, for all $n = 1, 2, \ldots$. All these derivatives admit a similar expression, as we now see:

Theorem 2 (Time Derivatives) *The t-derivatives of $v(t,q)$, for $n = 1, 2, \ldots$, are*

$$v_{t^n}(t,q) = \frac{1}{2}\hat{q}^2(1-\hat{q})^2(M-m)\left(\frac{\partial}{\partial t}\right)^{n-1}\xi(t,q,\hat{q}) \tag{13}$$

for all $(t,q) \in (0,\infty) \times (0,1)$, where $m \equiv \pi_A^H - \pi_A^L < \pi_B^H - \pi_B^L \equiv M$. In particular, the marginal value of information is given by a scaled standard normal density:

$$v_t(t,q) = \frac{(M-m)q(1-\hat{q})}{2\sqrt{t}}\phi\left(-\frac{1}{2}\sqrt{t} + \frac{1}{\sqrt{t}}L(\hat{q},q)\right). \tag{14}$$

Theorem 2 is proven in the appendix by just differentiating Theorem 1. We now give an *incomplete* but suggestive development of the marginal value of information in terms of the derivative of the belief transition density. Since the backward equation $u_t = \frac{1}{2}q^2(1-q)^2 u_{qq}$ holds at any time $\varepsilon > 0$, and since $u(t,q) = E_q[u(0,q(t))] \approx E_q[u(\varepsilon, q(t))]$,

$$u_t(t,q) \approx E_q[u_t(\varepsilon, q(t))] = \int_0^1 \tfrac{1}{2}r^2(1-r)^2 u_{qq}(\varepsilon, r)\xi(t,q,r)dr \approx \tfrac{M-m}{4\varepsilon}\int_{\hat{q}-\varepsilon}^{\hat{q}+\varepsilon} r^2(1-r)^2\xi(t,q,r)dr$$

where we approximate u_{qq} near the kink with $u_{qq}(\varepsilon,q) \approx \frac{M-m}{2\varepsilon}$ for all $q \in (\hat{q}-\varepsilon, \hat{q}+\varepsilon)$, and otherwise $u_{qq}(\varepsilon,q) = 0$. Taking the limit $\varepsilon \to 0$, we get $u_{qq}(0,\hat{q}) = \infty$ and otherwise $u_{qq}(0,q) = 0$. Thus, $u_t(t,q) \approx \frac{1}{2}(M-m)\hat{q}^2(1-\hat{q})^2\xi(t,q,\hat{q})$.

Comparing the marginal value schedule to the likewise left-skew log-normal density, the tail of v_t is asymptotically much thinner. Further notice that while the value of information behaves continuously as beliefs q converge upon the cross-over belief $\hat{q}$, the marginal value explodes for small t. This discontinuity is reflected in the next result.

Corollary 1 (Derivatives) *The marginal value of information obeys*[6]

(a) For all $t \in (0,\infty)$, $v_t(t,q) \in (0,\infty)$ for all $q \in (0,1)$, while $v_t(\infty,q) = 0$ for all q

(b) ***[Radner-Stiglitz (1984)]*** *$v_t(0+,q) = 0$ for all $q \neq \hat{q}$, while $v_t(0+,\hat{q}) = \infty$*

(c) ***[and beyond...]*** *Finally, $v_{t^n}(0+,q) = 0$ for all q and $n = 2, 3, \ldots$.*

The proof is in the appendix. Part (b) is the 'nonconcavity in the value of information' conclusion of RS and CS, since a marginal value that starts at zero cannot globally fall. (See Figure 5.) We go somewhat beyond this conclusion in part (c), finding that all higher order derivatives also initially vanish for our informational framework.

Note that the Inada condition of RS or of CS (their assumption A1) simply does not apply.[7] As natural as our Gaussian model is — it is the limit of the discrete sampling models, as we see in §8 — it escapes known sufficient conditions. And yet $v_t(0+,q) = 0$.

We now globally describe the marginal value of information. To this end, let us define an in<u>**fl**</u>ection point in the value of information, where the marginal value peaks:

$$T_{FL}(q) = 2\left[\sqrt{1 + L^2(\hat{q},q)} - 1\right]. \tag{15}$$

[6] As usual, $v_t(0+,q) = \lim_{s\downarrow 0} v_t(s,q)$, and $v_{t^n}(0+,q) = \lim_{s\downarrow 0} v_{t^n}(s,q)$.

[7] In our case, the signal X has mean $\mu_\theta t$ and variance t. The RS Inada condition fails, and assumption A1 in CS fails, because the signal density derivative in t explodes at $t = 0$ — as CS acknowledge for Kihlstrom (1974). But the critical difference is not the signal distribution, as CS suggest: Had we assumed the convex continuous action model of Kihlstrom, there would be no easily deduceable nonconcavity: Indeed, $u_t(0+,q) = \lim_{t\to 0} u_t(t,q) = \lim_{t\to 0}\frac{1}{2}q^2(1-q)^2 u_{qq}(t,q) = \frac{1}{2}q^2(1-q)^2 u_{qq}(q) > 0$ by assumption.

We remark that this inflection point demand is surprisingly independent of the payoff levels except insofar as it affects the cross-over belief $\hat{q}$. This holds in spite of the fact that the marginal value of information (14) is indeed increasing in the payoff stakes $M - m$.

Theorem 3 (Value and Marginal Value of Information) *Fix $q \in (0,1)$ with $q \neq \hat{q}$.* *(a) The value of information is initially convex until $t = T_{\mathrm{FL}}(q)$, after which it is concave.* *(b) The marginal value is rising until $T_{\mathrm{FL}}(q)$, and later falling. It is convex in $[0, T_1(q)]$, concave in $(T_1(q), T_2(q))$, and convex for $[T_2(q), \infty)$, where $T_{\mathrm{FL}}(q) \in (T_1(q), T_2(q))$.*

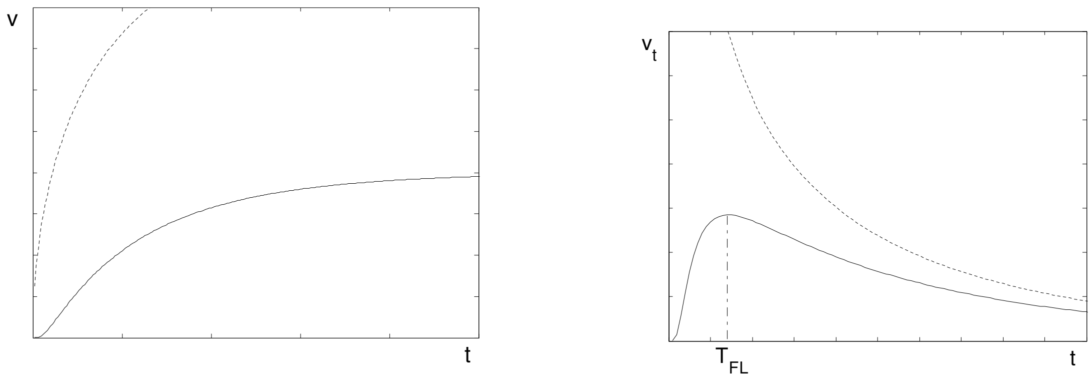

Figure 5: **Value and Marginal Value of Information.** At left (right) is the graph of the value (marginal value) of information for the parameter values $\hat{q} = 0.5$, $\pi_H^A = 1$, $\pi_L^A = 2$, $\pi_H^B = 2$, and $\pi_L^B = 1$. The solid lines are the values with $q = 0.2$ or $q = 0.8$, while the dotted lines with the cross-over belief $q = 0.5$. Observe that the nonconcavity arises when we do not start at the cross-over belief.

Proof: First note that from (15) we get that $T_{\mathrm{FL}}(q) \in [0, \infty)$ and that $T_{\mathrm{FL}}(q) = 0$ if and only if $q = \hat{q}$. From (8) and Theorem 2 we get

$$v_{tt}(t,q) = v_t(t,q) \left[-\frac{1}{8} + \frac{L^2(\hat{q},q)}{2t^2} - \frac{1}{2t} \right], \tag{16}$$

where $v_t(t,q) > 0$, according to Corollary 1. Now $v_{tt}(t,q) = 0$ gives $\Upsilon(t,q) \equiv t^2 + 4t - 4L^2(\hat{q},q) = 0$, which yields (15). Part (a) owes to $\Upsilon(s,q) \lesseqgtr 0$ for all $s \gtreqless T_{\mathrm{FL}}(q)$.

For (b), note that $v_{ttt}(0+, q) = 0$ according to Corollary 1. Second, from Theorem 2:

$$v_{ttt}(t,q) = v_t(t,q) \left[\left(-\frac{1}{8} + \frac{L^2(\hat{q},q)}{2t^2} - \frac{1}{2t} \right)^2 - \frac{L^2(\hat{q},q)}{t^3} + \frac{1}{2t^2} \right] \tag{17}$$

and hence $v_{ttt}(t,q) = 0$ if $t^4 + 8t^3 + (48 - 8L^2)t^2 - 96L^2 t + 16L^2 = 0$. Clearly, $v_{ttt} > 0$ near $t = 0$. So if there is only one positive root $T_1(q)$ then $v_{ttt}(t,q) < 0$ for all $t > T_1(q)$.

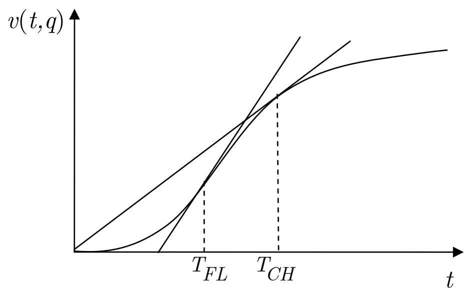

Figure 6: **Consequences of the Information Non-concavity.** The choke-off demand $T_{\mathrm{CH}}(q)$ exceeds the peak marginal value demand $T_{\mathrm{FL}}(q)$ due to the non-concavity of information. Thus, the demand curve for information is not simply the falling portion of the marginal value of information.

Since $v_{ttt}(t,q) > 0$ for large t, if there exists a strictly positive root then there must be two strictly positive roots $T_1(q)$ and $T_2(q)$. If there are no positive roots then $v_{ttt}(t,q) > 0$ for all $t > 0$. Along with (c) in Corollary 1, this gives $v_{tt}(t,q) > 0$ for all $t > 0$ — contradicting part (a) of this Lemma. So there are two positive roots. We have $T_1(q) < T_{\mathrm{FL}} < T_2(q)$ since $v_{tt}(T_{\mathrm{FL}}, q) = 0$ and $v_{tt}(t,q) > 0$ for all $t < T_{\mathrm{FL}}$. This gives (b). □

In light of Lemma 6, the convexity before $T_{\mathrm{FL}}(q)$ owes to the increasing transition pdf $(\xi_t(t,q,\hat{q}) \geq 0)$ and the concavity after $T_{\mathrm{FL}}(q)$ owes to the decreasing pdf $(\xi_t(t,q,\hat{q}) \leq 0)$.

6 The Demand for Information

A. The Demand Curve. We now consider linear pricing of information $c(t) = ct$, where c is a strictly positive constant. Let *demand* $\tau(c,q)$ maximize consumer surplus

$$\Pi(t,q) = u(t,q) - ct = u(q) + v(t,q) - ct. \tag{18}$$

We hereafter fix $q \neq \hat{q}$, ignoring the cross-over belief $\hat{q}$, since it is a single point (i.e. nongeneric); we can thus avoid hedging our theorems. Because of the non-concavity near quantity 0, and since the marginal value finitely peaks, there exists a *choke-off cost* $c_{\mathrm{CH}}(q) > 0$, above which demand is zero, and an implied minimum *choke-off demand*, $T_{\mathrm{CH}}(q) > 0$. At the cost $c_{\mathrm{CH}}(q) > 0$, demand is $T_{\mathrm{CH}}(q)$ and consumer surplus is zero. Thus, marginal value is falling, and so $T_{\mathrm{CH}}(q) \geq T_{\mathrm{FL}}(q)$, as in Figure 6. Summarizing:

Cost 'choke-off': $$c_{\mathrm{CH}}(q) = v_t(T_{\mathrm{CH}}(q), q) = \frac{v(T_{\mathrm{CH}}(q), q)}{T_{\mathrm{CH}}(q)}. \tag{19}$$

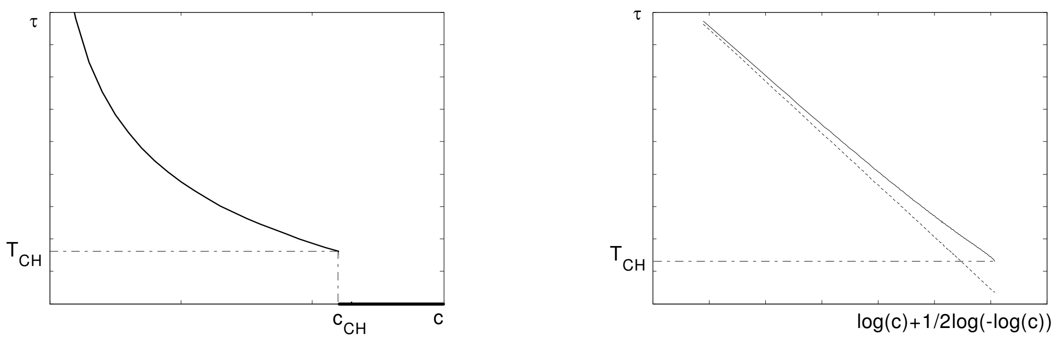

Figure 7: **Optimal and Approximate Large Demand.** The true demand curve is depicted in the left figure. In the right figure, true demand is the solid line and the approximation is the dotted line. Parameter values: $q = 0.2$ or 0.8, $\hat{q} = 0.5$, $\pi_H^A = 1$, $\pi_L^A = 2$, $\pi_H^B = 2$, and $\pi_L^B = 1$.

Let us define $T_{\mathrm{FOC}}(c, q) < T_{\mathrm{FL}}(q)$ by the FOC $v_t(T_{\mathrm{FOC}}(c, q), q) = c$. This is well-defined iff $c \leq v_t(T_{\mathrm{FL}}(q), q)$, since $v_{tt}(T_{\mathrm{FL}}(q), q) < 0$ on $(T_{\mathrm{FL}}(q), \infty)$ (Theorem 3-(b)). Demand is captured by the FOC precisely when the cost falls below the choke-off cost.

Theorem 4 (Demand) $\tau(c, q) = 0$ *if* $c > c_{\mathrm{CH}}(q)$ *and* $\tau(c, q) = T_{\mathrm{FOC}}(c, q)$ *if* $c \leq c_{\mathrm{CH}}(q)$.

Proof: Observe that by (18), $\Pi(t, q) = \int_0^t [v_t(s, q) - c] ds$. The integrand is first negative, since $v_t(0+, q) = 0$, and eventually negative, since $v_t(\infty, q) = 0$. If $c > c_{\mathrm{CH}}(q)$, then the integral (consumer surplus) is always negative, and so the optimal demand is $t = 0$. Otherwise, if $c \leq c_{\mathrm{CH}}(q) < v_t(T_{\mathrm{FL}}(q), q)$, then $T_{\mathrm{FOC}}(c, q)$ exists, and by Theorem 3, the integrand is positive for an interior interval ending at $T_{\mathrm{FOC}}(c, q)$, where $v_t(T_{\mathrm{FOC}}(c, q), q) - c = 0$. Thus, the integral is maximized at $t = T_{\mathrm{FL}}(q)$, as needed. □

Corollary 2 (Law of Demand) *Demand is falling in the price c, for* $c < c_{\mathrm{CH}}(q)$.

Indeed, simply apply $v_{tt}(t, q) < 0$ for all $t > T_{\mathrm{FL}}(q)$ (true by Theorem 3). Notice that the law of demand applies to information too, but only after the price drops below the choke-off level $c_{\mathrm{CH}}(q)$, so that positive demand is warranted. Figure 7 illustrates these results: the jump in information demand as costs drop, as well as the Law of Demand.

B. Quantifying the Lumpiness. We now wish to explore the size of the nonconcavity in the demand for information. The most direct approach here is to quantify the minimum expenditure $Tv_t(T, q)$ on information that the *DM* must incur. Of course, this amount should increase in the maximum payoff stakes, simply because the marginal value of information does, by Lemma 2. Additionally, if beliefs are near 0 or 1, then information demand vanishes. Seeking an appropriate normalization, let the *expected payoff stakes* denote the maximum expected payoff loss from choosing a wrong action. We evaluate these using the worst case scenario, which occurs at the cross-over belief $q = \hat{q}$:

$$\hat{q}[\text{max loss if } \theta = H] + (1-\hat{q})[\text{max loss if } \theta = L] = \hat{q}[\pi_B^H - \pi_A^H] + (1-\hat{q})[\pi_A^L - \pi_B^L] = \hat{q}(1-\hat{q})(M-m)$$

This clearly vanishes when $\hat{q} = 0, 1$, and increases in the maximum payoff stakes $M - m$.

Theorem 5 (Least Positive Information Costs) *The average lower bound on information expenditures normalized by the payoff stakes exceeds* 0.025*, or*

$$\int_0^1 \frac{T_{CH}(r) v_t(T_{CH}(r), r)}{\hat{q}(1-\hat{q})(M-m)} dr > 0.025. \tag{20}$$

Proof: Suppressing the arguments of $L = L(\hat{q}, q)$ and $T_{CH}(q)$, we have

$$\frac{v(T_{CH}, q)}{\hat{q}(1-\hat{q})(M-m)} = \frac{\int_0^{T_{CH}} v_t(s,q) ds}{\hat{q}(1-\hat{q})(M-m)} = \frac{1}{2}\sqrt{\frac{q(1-q)}{2\pi\hat{q}(1-\hat{q})}} \int_0^{T_{CH}} \frac{1}{\sqrt{s}} \exp\left(-\frac{s}{8} - \frac{L^2}{2s}\right) ds.$$

Using this equation, a lower bound on (20) is 0.025, as we show in the appendix. □

One reason why we take an average here is that the threshold choke-off cost $c_{\mathrm{CH}}(q)$ vanishes as beliefs q approach $\hat{q}$ or the extremes $0, 1$. Thus, the minimum information purchase likewise vanishes nearing those three beliefs, and only an average makes sense.

C. Demand as a Function of Beliefs. A classic question asked of Bayesian sequential learning models is the range of beliefs with a positive experimentation level.

Theorem 6 (Interval Demand) *Demand* $\tau(c, q) > 0$ *iff beliefs* q *belong to an interior interval* $\big(\underline{q}(c), \bar{q}(c)\big)$*, where the thresholds* $0 < \underline{q}(c) < \hat{q} < \bar{q}(c) < 1$ *obey*

$$v\left(\tau\left(c, \underline{q}(c)\right), \underline{q}(c)\right) = \tau\left(c, \underline{q}(c)\right) c \qquad \text{and} \qquad v\left(\tau(c, \bar{q}(c))\right) = \tau\left(c, \bar{q}(c)\right) c. \tag{21}$$

Furthermore, the choke-off demands are $T_{\mathrm{CH}}(\bar{q}(c)) = \tau(c, \bar{q}(c))$ *and* $T_{\mathrm{CH}}(\underline{q}(c)) = \tau(c, \underline{q}(c))$*.*

Proof: First of all, demand is clearly positive at belief $\hat{q}$, since $v_t(0+, \hat{q}) = \infty$, by (b) in Corollary 1. Further, demand vanishes at $q = 0, 1$, since $v_t(t, 0) = u_t(t, 0) = 0$ and $v_t(t, 1) = u_t(t, 1) = 0$ for all t. Thus, any interval structure obeys $0 < \underline{q}(c) < \hat{q} < \bar{q}(c) < 1$.

Next, when positive, demand must satisfy the FOC $v_t(\tau(q), q) = c$. Thus, it suffices to prove that $v_t(t, q)$ is strictly quasi-concave in q, but this fails. Instead, we show local quasi-concavity at the optimal demand $\tau(c, q)$, since $v_t(t, q)$ is continuous. Specifically, we show that $v_{tq}(\tau(c, q), q) = 0$ implies $v_{tqq}(\tau(c, q), q) < 0$.

Differentiating the FOC yields $\tau_q(c, q) = -v_{tq}/v_{tt}$, if we suppress arguments. Hence,

$$\tau_{qq}(c, q) = -\frac{1}{v_{tt}}\left(v_{tqq} + v_{ttq}\tau_q\right) + \frac{v_{tq}}{v_{tt}^2}\left(v_{ttq} + v_{ttt}\tau_q\right) = -\frac{1}{v_{tt}} v_{tqq}. \tag{22}$$

because our premise $v_{tq}(\tau(c, q), q) = 0$ is clearly equivalent to $\tau_q(c, q) = 0$, by the FOC. By Theorem 3-(b), we have $v_{tt}(\tau(c, q), q) < 0$, and so $v_{tqq}(\tau(c, q), q)$ and $\tau_{qq}(c, q)$ share the same sign. Since $u_{tt}(\tau(c, q), q) = \frac{1}{2}q^2(1-q)^2 u_{tqq}(\tau(c, q), q)$ by Lemma 7, we complete the proof with

$$-\frac{v_{tqq}(\tau(c, q), q)}{v_{tt}(\tau(c, q), q)} = -\frac{u_{tqq}(\tau(c, q), q)}{u_{tt}(\tau(c, q), q)} = -\frac{2}{q^2(1-q)^2} < 0. \tag{23}$$

Finally, demand vanishes when the *DM* is indifferent between buying and not buying at all — namely, at the choke-off level. So (21) follows, and $T_{\mathrm{CH}}(q)$ is as described. □

These results strikingly differ from their analogues in a dynamic setting. That information demand is positive precisely on an interior interval is completely in harmony with the standard sequential experimentation result (see MS1). However, it holds for an entirely unrelated reason! In sequential experimentation, the *DM* stops when his Bellman equation indicates that *marginal* costs and benefits of further experimentation balance. In our static demand setting, the *DM* buys no information when *total* costs and benefits of any information purchase balance. Namely, given the nonconcavity in the value of information, this demand choke-off decision turns on considerations of *global optimality.*

The following theorem gives the relationship between positive demand and beliefs.

Theorem 7 (Hill-Shaped Demand) *$\tau(c, \cdot)$ is quasi-concave in beliefs $q \in \big(\underline{q}(c), \bar{q}(c)\big)$.*

Proof: By Theorem 6, $\tau(c, q) > 0$ and so $\tau_q(c, q) = 0$ implies $\tau_{qq}(c, q) < 0$ by (22)–(23). □

Again, a comparison with the dynamic case is instructive, and here it is a contrast in the result, and not just the rationale for the result. MS1 assume a convex cost of information in a sequential experimentation model and deduce instead that information demand is U-shaped and convex, and not hill-shaped and concave. The static demand solution is the intuitive one, with demand greatest when the DM is most uncertain.

D. The Elasticity of Demand. The elasticity of the demand equals $|\tau_c(c,q)c/\tau(c,q)|$. When the elasticity equals 1, the demand level is $T_E(q)$, and revenue $v_t(t,q)t$ is maximized. Using this fact, we characterize $T_E(q)$ below, using the belief derivative (16):

$$T_E(q) = -\frac{v_t\left(T_E(q),q\right)}{v_{tt}\left(T_E(q),q\right)} = 2\left[1+\sqrt{1+L^2(\hat{q},q)}\right] = T_{\mathrm{FL}}(q)+4. \tag{24}$$

Like the peak marginal value $T_{\mathrm{FL}}(q)$, the unit elastic demand does not depend on the underlying payoff stakes, apart from the dependence on the cross-over belief $\hat{q}$. Further, the marginal value is clearly falling at $T_E(q)$, since it exceeds $T_{\mathrm{FL}}(q)$. A key question in our setting is whether it lies above the choke-off demand. The answer is yes, provided the belief is sufficiently interior.

Theorem 8 (Elasticity)

(a) *Demand is initially elastic for* $q \in (\acute{q},\grave{q})$*, where* $0<\acute{q}<\hat{q}<\grave{q}<1$.

(b) *Demand elasticity is decreasing in the cost* c*, for all* $c \le c_{\mathrm{CH}}(q)$.

Observe that $(\acute{q},\grave{q}) \subset (\underline{q}(c),\bar{q}(c))$ because demand is positive for $q \in (\acute{q},\grave{q})$. Define $c_{\mathrm{E}}(q) = v_t\left(T_E(q),q\right)$, namely, the cost level where demand elasticity equals 1. Then $c_{\mathrm{E}}(q) < c_{\mathrm{CH}}(q)$ iff $q \in (\acute{q},\grave{q})$. Rephrasing part ($a$), information demand is elastic for $c \in (c_{\mathrm{E}}(q), c_{\mathrm{CH}}(q)]$ and inelastic if $c < c_{\mathrm{E}}(q)$.

To make some sense of part (b), one can reason that the marginal value of information drops off so fast in the tail of the normal pdf that the demand elasticity falls monotonically.

E. Large Demand. We now give a simple asymptotically applicable demand formula. It is consistent with the formula that MS2 derived for conditionally iid samples from any signal distribution, using large deviation theory. Our work here, which follows from our Gaussian framework, is more refined, as we specify an additional error term,[8] and show

[8] In the discrete signal world of MS2, their formula was eventually accurate within one signal.

that it is positive — in other words, the limit demand curve is approached from above.

Theorem 9 (Low Prices) *If c is small then the optimal demand is approximated by:*

$$\tau(c,q)/8 = F(q) - \log(c) - \tfrac{1}{2}\log(-\log(c)) + \frac{\log(-\log(c))}{4\log(c)}(1+o(1)), \qquad (25)$$

where $o(1)$ vanishes in c, and where $F(q) = \frac{1}{2}\log[q(1-q)\hat{q}(1-\hat{q})/64\pi] + \log(M-m)$.

Observe that the approximate difference $2[\log(-\log(c))]/\log(c)$ between demand and $F(q) - \log c - \frac{1}{2}\log(-\log c)$ is negative, and asymptotically vanishing in c (see Figure 7). Therefore, the three c-dependent terms of the demand function (25) provide (in order, adding them from left to right) increasingly accurate approximations. As the cost of information vanishes, demand is essentially logarithmic.

Proof of Theorem 9*:* First, if the cost c is small then $T_{\text{FOC}}(c,q)$ exists and $u(T_{\text{FOC}}(c,q),q) \geq u(0,q)$. So in this case $\tau(c,q) = T_{\text{FOC}}(c,q)$. Second, from the FOC $v_t(\tau(c,q),q) = c$:

$$c = \frac{(M-m)q(1-\hat{q})}{2\sqrt{2\pi\tau(c,q)}}\exp\left\{-\frac{1}{2}\left[\frac{1}{4}\tau(c,q) - L(\hat{q},q) + \frac{L^2(\hat{q},q)}{\tau(c,q)}\right]\right\}.$$

Taking logs, using the definitions of $L(\hat{q},q)$ and $F(q)$ yields the *log inverse demand curve*:

$$\log(c) = F(q) - \frac{1}{2}\log(\tau(c,q)/8) - \frac{L^2(\hat{q},q)}{16\tau(c,q)/8} - \frac{1}{8}\tau(c,q) = F(q) - \psi(\tau(c,q)/8), \qquad (26)$$

where $\psi(x) = x + \frac{1}{2}\log x + B/x$, for $B = L^2(\hat{q},q)/16$. Notice that ψ^{-1} exists when $\psi'(x) = 1 - 1/(2x) - B/x^2 > 0$, which is true for large x.[9] Rearranging (26) produces the demand curve $\tau(c,q) = 8\psi^{-1}(F(q) - \log c)$. The appendix shows that $\psi^{-1}(x) = x - \frac{1}{2}\log x + 2[(\log x)/x](1+o(1))$ and that $\psi^{-1}(x) > x - \frac{1}{2}\log x$. Thus, (25) follows from

$$\tau(c,q)/8 = F(q) - \log c - \tfrac{1}{2}\log(-\log c) - \tfrac{1}{2}\log\left(1 - \frac{F(q)}{\log c}\right) + \frac{1}{4}\frac{\log(F(q) - \log c)}{F(q) - \log c}(1+o(1)).$$

□

Recall that we have normalized $\zeta = 1$. Had we not done so, the demand function in (25) would hold if divided by ζ^2. MS2 analyze the large demand for information as

[9] Specifically, this holds for $x > 1 + \sqrt{1 + L^2(\hat{q},q)/16}$, which is less than $T_E(q)$. In other words, certainly starting when demand is inelastic, our inverse demand curve $\tau(c,q) = 8\psi^{-1}(F(q) - \log c)$ is valid.

the price c vanishes, for any arbitrary signal, not just for weak or gaussian signals as we do here. They define a general information index for a signal, $\rho \in (0,1)$, where 0 means perfectly informative, and 1 uninformative. They show that the demand function for small c has the same logarithmic form as in (25), with same slope when we define $\zeta^2 = -8\log(\rho)$:

$$\frac{\tau(c,q)}{8\log(\rho)} = F(q) - [\log(c) + \frac{1}{2}\log(-\log(c))] + \frac{\log(-\log(c))}{4\log(c)}(1+o(1)).$$

In other words, the demand for any signal with information index ρ at small price is approximately the same as the demand time with $\zeta^2 = -8\log(\rho)$. Notice that ζ rises from 0 to infinity as informativeness rises (ρ falls from 1 to 0).

7 The Natural 'Small Bits' of Information

7.1 Beliefs

Any theory of variable quantity information requires that it be measured in sufficiently small units. While the paper assumed the diffusion process $X(\cdot)$, we now observe that this well-approximates a wide class of discrete choice models with 'small bits' of information. Assume the DM chooses the number n of i.i.d. draws from a signal, not necessarily Gaussian. We show that as the informational content and the cost of each draw jointly vanish, the associated value of information and the optimal sample size suitably converge to their continuous time Brownian Motion counterparts.[10] Therefore, our demand for Gaussian information approximates the optimal number of cheap conventional signals.

Let $\{g(\cdot|\theta,\Delta)\}$ be a simple signal — a family of p.d.f.s, each indexed by the state $\theta \in \{L,H\}$, with support $\mathcal{Z}$ independent of θ. Further, assume that the signal becomes uninformative as Δ vanishes: $\lim_{\Delta\downarrow 0} g(Z|\theta,\Delta) = \bar{g}(Z) > 0$. Here, Δ is the real time elapse length of a time interval in discrete time. In the (continuous) time span $[0,t]$, the DM observes $n = \lfloor t/\Delta \rfloor$ draws from $g(\cdot|\theta,\Delta)$ at times $\Delta, 2\Delta, \ldots, n\Delta \doteq t$, where $\lfloor a \rfloor$ is the largest integer smaller than a. As Δ vanishes, the DM observes an exploding number of increasingly uninformative conditionally i.i.d. signal realizations at very high frequency.

[10] Abreu, Milgrom, Pearce (1991) confront a related informational problem.

Beliefs evolve according to Bayes rule:

$$q^{\Delta}(n\Delta|Z) = \frac{q^{\Delta}((n-1)\Delta|Z)g(Z|H,\Delta)}{q^{\Delta}((n-1)\Delta|Z)g(Z|H,\Delta) + [1-q^{\Delta}((n-1)\Delta|Z)]g(Z|L,\Delta)}. \tag{27}$$

We want this Markov process to converge weakly to the diffusion (2) as Δ vanishes and n explodes.[11] When this obtains, $q^{\Delta}(\lfloor t/\Delta \rfloor \Delta)$ and $q(t)$ have nearly the same finite-dimensional distributions (suppressing Z arguments), as the next result asserts.

Theorem 10 (Small Bits) *For every $t > 0$, the Markov process in* (27) *converges weakly to the diffusion $d\tilde{q}(t) = \zeta\tilde{q}(t)(1-\tilde{q}(t))dW(t)$ as $\Delta \downarrow 0$, or $q^{\Delta}(\lfloor t/\Delta \rfloor \Delta) \Rightarrow \tilde{q}(t)$ if*

$$\lim_{\Delta \downarrow 0} \frac{1}{\Delta} \int \frac{[g(z|H,\Delta) - g(z|L,\Delta)]^2}{\bar{g}(z)} dz = \zeta^2 \tag{28}$$

Example. We propose one natural way to construct the required $g(\cdot|\theta,\Delta)$ from *any* signal $\{f(\cdot|\theta)\}$. Just garble f by supposing that in state θ the signal is drawn from $f(\cdot|\theta)$ with chance $1/2+\sqrt{\Delta}$, and from the other "incorrect" distribution with chance $1/2-\sqrt{\Delta}$. For every $\Delta > 0$, this yields the required state-independent support $\mathcal{Z}$. As Δ vanishes, the signal becomes pure noise, with state-independent limit $\bar{g}(Z) = [f(Z|H)+f(Z|L)]/2$. One can verify for this example, using (28) and $\lim_{\Delta \downarrow 0} g(Z|\theta,\Delta) = \bar{g}(Z) > 0$, that

$$\zeta^2 = 8 \int_{\mathcal{Z}} \frac{[f(z|H) - f(z|L)]^2}{f(z|H) + f(z|L)} dz \tag{29}$$

For example, if $f(z|H) = 2z$ and $f(z|L) = 2(1-z)$ for $z \in [0,1]$, then (29) implies

$$\zeta^2 = 4\int_0^1 (4z-2)^2 dz = 16\int_0^1 (2z-1)^2 dz = 16/3.$$

Thus, the relevant approximation is an arithmetic Wiener process with state-dependent drift $\mu = \pm(2/\sqrt{3})\sigma$ for any diffusion coefficient σ.

[11] Convergence is in the sense of Skorokhod's topology on the space $D([0,t],\mathbb{R})$ of right-continuous functions with left limits.

7.2 Value and Demand Approximation

The ex ante value of a sample of $n = \lfloor t/\Delta \rfloor$ conditionally i.i.d. draws from $\{g(\cdot|\theta,\Delta)\}$ is

$$v^{\Delta}(t,q) \equiv E_q[u(q^{\Delta}(\lfloor t/\Delta \rfloor \Delta))] - u(q),$$

where the expectation is taken w.r.t. the distribution of the discrete time belief process $q^{\Delta}(\lfloor t/\Delta \rfloor \Delta)$. Since the latter converges weakly to $q(t)$, and $u(\cdot)$ is a continuous function:

$$\lim_{\Delta \downarrow 0} \left| v^{\Delta}(t,q) - v(t,q) \right| = \lim_{\Delta \downarrow 0} \left| E_q[u(q^{\Delta}(\lfloor t/\Delta \rfloor \Delta))] - E_q[u(q(t))] \right| = 0 \tag{30}$$

for every prior $q \in [0,1]$. The discrete time value function converges to the continuous one. If we let $v^0(t,q) \equiv v(t,q)$, then $v^{\Delta}(t,q)$ is continuous in Δ at $\Delta = 0$ for every $t \geq 0$.

Now consider the following decision problem. Fix $\Delta > 0$. The DM can purchase n conditionally i.i.d. draws of Z, non-sequentially, at unit price $c\Delta$ and total outlays $nc\Delta$, yielding payoff $\Pi^{\Delta}(n\Delta|c,q) = v^{\Delta}(n\Delta,q) - cn\Delta$. The optimization problem is

$$\sup_{n \in \mathbb{N}} \Pi^{\Delta}(n\Delta|c,q). \tag{31}$$

Note that $\Pi^{\Delta}(0|c,q) = v(0,q)$ and $v^{\Delta}(n\Delta,q) \leq \max_{\theta,a} \pi_a^{\theta}$ for any $\Delta > 0$, so that $\lim_{n\to\infty} \Pi^{\Delta}(n\Delta|c,q) = -\infty$. It follows that a finite non-negative maximum of $\Pi^{\Delta}(n\Delta|c,q)$ over integers $n = 0,1,2,\ldots$ exists, and is attained by a correspondence $N^{\Delta}(c,q)$.

The analogous continuous time problem when observing $X(t)$ with state-dependent drift is $\sup_{t\geq 0} \Pi(t|c,q)$, where $\Pi(t|c,q) = v(t,q) - ct$. This problem yields the real-valued demand function $\tau(c,q)$ that we have characterized. Along with (30), we find that

$$\lim_{\Delta \downarrow 0} \Pi^{\Delta}(t|c,q) = \lim_{\Delta \downarrow 0} v^{\Delta}(t,q) - ct = v^0(t,q) - ct = v(t,q) - ct = \Pi(t|c,q).$$

The maximized value in the discrete model is thus near that of the continuous time model.

Theorem 11 (Demand Convergence) *There is a selection $n^{\Delta}(c,q) \in N^{\Delta}(c,q)$ such that for a.e. parameter $\pi_A^{\theta}, \pi_B^{\theta}, q, c > 0$:*

$$\lim_{\Delta \downarrow 0} \left| n^{\Delta}(c,q)\Delta - \tau(c,q) \right| = 0.$$

The proof is in the appendix.

Next, the demand elasticity is also approximated by the discrete approximations:

$$\lim_{\varepsilon\to 0}\lim_{\Delta\downarrow 0}\frac{n^{\Delta}(c+\varepsilon,q)-n^{\Delta}(c,q)}{n^{\Delta}(c+\varepsilon,q)\varepsilon}=\lim_{\varepsilon\to 0}\frac{\tau(c+\varepsilon,q)-\tau(c,q)}{\varepsilon\tau(c+\varepsilon,q)}=\frac{\tau_c\,(c,q)}{\tau(c,q)}c.$$

8 Conclusion

We have measured information in units that correspond to sample sizes of conditionally iid signals, and assumed weak enough signals. In this setting, we have completely and tractably characterized the demand for information for the two state world. We have, in particular, provided the full picture of the famous informational nonconcavity for the first time. Additionally, we have characterized the elasticity of information, the large demand formula, and the dependence of demand on beliefs. Finally, we have shown that our model well approximates all small bit models of information. Our theory extends, with additional complexity, to a model with any finite set of actions, since that merely adds to the number of cross-over beliefs. The restriction to two states, instead, is real.

Kihlstrom (1974) succeeded with off-the-shelf techniques precisely because of the self-conjugate property of the normal, as well as the particular payoff function. Analysis of the learning *per se* is bipassed entirely. But a learning model is generally useful insofar as one knows how posterior beliefs evolve stochastically. Either a finite action space or state space invalidates this approach, and one must treat the learning problem seriously. And indeed, many choices in life are inherently discrete, like whether to change jobs, have a child, or build a prototype. In our two-state model, beliefs are not linear in signal realizations, so that this solving for this density requires new solution methods. We show that this problem is just like the option pricing exercise — only harder.

A Omitted Proofs

A.1 Limit Beliefs: Proof of Lemma 1 *(a)*

For all $q \in (0,1)$, we have $q^2(1-q)^2 > 0$ and $\int_{q-\varepsilon}^{q+\varepsilon} \frac{2dy}{y^2(1-y)^2} < \infty$ for some $\varepsilon > 0$. So Feller's test for explosions (Karatazas and Shreve (1991), Theorem 5.29) implies part (a) because

$$\int_0^c \frac{2dq}{q^2(1-q)^2} \geq 2\int_0^c \frac{1}{q}dq = \infty \quad \forall \quad c \in (0,1).$$

A.2 Value Function Derivation: Proof of Theorem 1

By Lemma 8, the expected payoff can be represented as an integral as follows

$$u(t,q) = \int_{-\infty}^{\infty} \left((1-q)\,e^{-\frac{1}{2}t-\sqrt{t}y} + q\right) u\left(\frac{1}{\left(\frac{1}{q}-1\right)e^{-\frac{1}{2}t-\sqrt{t}y}+1}\right)\phi(y)dy.$$

Let us exploit symmetry $\phi(y) = \phi(-y)$. Since $u(q) = \max\langle \pi_A^L + mq, \pi_B^L + Mq\rangle$,

$$\begin{aligned}
u(t,q) \;=\; & q\int_{\hat{y}(t,q)}^{\infty}\left(\left(\tfrac{1}{q}-1\right)e^{-\frac{1}{2}t+\sqrt{t}y}+1\right)\left(\pi_A^L + \frac{m}{\frac{1}{q}-1\;\;e^{-\frac{1}{2}t+\sqrt{t}y}+1}\right)\phi(y)dy \\
& +q\int_{-\infty}^{\hat{y}(t,q)}\left(\left(\tfrac{1}{q}-1\right)e^{-\frac{1}{2}t+\sqrt{t}y}+1\right)\left(\pi_B^L + \frac{M}{\frac{1}{q}-1\;\;e^{-\frac{1}{2}t+\sqrt{t}y}+1}\right)\phi(y)dy \\
=\; & q\left(\pi_A^L + m\right)\int_{\hat{y}(t,q)}^{\infty}\phi(y)dy + (1-q)\pi_A^L\int_{\hat{y}(t,q)}^{\infty} e^{-\frac{1}{2}t+\sqrt{t}y}\phi(y)dy \qquad (32)\\
& +q\left(\pi_B^L + M\right)\int_{-\infty}^{\hat{y}(t,q)}\phi(y)dy + (1-q)\pi_B^L\int_{-\infty}^{\hat{y}(t,q)} e^{-\frac{1}{2}t+\sqrt{t}y}\phi(y)dy,
\end{aligned}$$

where $\sqrt{t}\hat{y}(t,q) = t/2 + \log\left(\left(\frac{M-m}{\pi_A^L-\pi_B^L}-1\right)\frac{q}{1-q}\right)$ satisfies

$$\pi_A^L + \frac{m}{\left(\frac{1}{q}-1\right)e^{-\frac{1}{2}t+\sqrt{t}\hat{y}(t,q)}+1} = \pi_B^L + \frac{M}{\left(\frac{1}{q}-1\right)e^{-\frac{1}{2}t+\sqrt{t}\hat{y}(t,q)}+1}.$$

The second and the last integrands in (32) can be simplified using

$$e^{-\frac{1}{2}t+\sqrt{t}y}\phi(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t+\sqrt{t}y-\frac{y^2}{2}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{(y-\sqrt{t})^2}{2}},$$

which is the pdf of a normal variable with mean $\sqrt{t}$ and unit variance. This and (32) give

$$\begin{aligned} u(t,q) &= q\left(\pi_A^L + m\right)\int_{\hat{y}(t,q)}^{\infty}\phi(y)dy + (1-q)\pi_A^L\int_{\hat{y}(t,q)-\sqrt{t}}^{\infty}\phi(y)dy \\ &\quad +q\left(\pi_B^L + M\right)\Phi\left(\hat{y}(t,q)\right) + (1-q)\pi_B^L\Phi\left(\hat{y}(t,q)-\sqrt{t}\right). \end{aligned}$$

Symmetry $\phi(y)=\phi(-y)$ and all parametric definitions yield

$$\begin{aligned} u(t,q) &= q\pi_A^H\Phi\left(-\frac{1}{2}\sqrt{t}+\frac{1}{\sqrt{t}}L(\hat{q},q)\right) + (1-q)\pi_A^L\Phi\left(\frac{1}{2}\sqrt{t}+\frac{1}{\sqrt{t}}L(\hat{q},q)\right) \\ &+ q\pi_B^H\Phi\left(\frac{1}{2}\sqrt{t}-\frac{1}{\sqrt{t}}L(\hat{q},q)\right) + (1-q)\pi_B^L\Phi\left(-\frac{1}{2}\sqrt{t}-\frac{1}{\sqrt{t}}L(\hat{q},q)\right). \end{aligned}$$

Using $\Phi\left(y\right)=1-\Phi\left(-y\right)$, we get $FIG(t,q)$; $v(t,q)=u(t,q)-u(q)$ then yields (12). □

A.3 Marginal Value: Proof of Theorem 2

Differentiating Theorem 1 with respect to t, and denoting $L=L(\hat{q},q)$, yields

$$\begin{aligned} v_t(t,q) &= q\pi_A^H\phi\left(-\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right)\left(-\frac{1}{4\sqrt{t}}-\frac{L}{2t^{3/2}}\right) + (1-q)\pi_A^L\phi\left(\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right)\left(\frac{1}{4\sqrt{t}}-\frac{L}{2t^{3/2}}\right) \\ &\quad +q\pi_B^H\phi\left(\frac{\sqrt{t}}{2}-\frac{L}{\sqrt{t}}\right)\left(\frac{1}{4\sqrt{t}}+\frac{L}{2t^{3/2}}\right) + (1-q)\pi_B^L\phi\left(-\frac{\sqrt{t}}{2}-\frac{L}{\sqrt{t}}\right)\left(-\frac{1}{4\sqrt{t}}+\frac{L}{2t^{3/2}}\right) \\ &= \phi\left(-\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right)\left[(\pi_B^H-\pi_A^H)q\left(\frac{1}{4\sqrt{t}}+\frac{L}{2t^{3/2}}\right) + (\pi_A^L-\pi_B^L)\frac{q(1-\hat{q})}{\hat{q}}\left(\frac{1}{4\sqrt{t}}-\frac{L}{2t^{3/2}}\right)\right] \\ &= \frac{(M-m)q(1-\hat{q})}{2\sqrt{t}}\phi\left(-\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right), \end{aligned}$$

where the second equality owes to $\phi\left(-x\right)=\phi\left(x\right)$ and $\phi\left(\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right)=\phi\left(-\frac{\sqrt{t}}{2}+\frac{L}{\sqrt{t}}\right)e^{-L}$, and the last to $M-m=(\pi_B^H-\pi_A^H)+(\pi_A^L-\pi_B^L)$ and $\hat{q}=(\pi_A^L-\pi_B^L)/(M-m)$.

Finally (14) yields (13) by taking further time derivatives and by using Lemma 5. □

A.4 Slopes at Zero: Proof of Corollary 1

Claim 1 *For any $n>1$ we have:*

$$\frac{\partial^n v(t,q)}{\partial t^n} = v_t(t,q)\left[\frac{A_{2(n-1)}(q)}{t^{2(n-1)}}+\cdots+\frac{A_1(q)}{t}+A_0(q)\right],$$

where $A_{2(n-1)}(q), \cdots, A_0(q)$ are bounded functions of q.

Proof: From (8), since $v_{tt}(t,q) = v_t(t,q)\left[\frac{A_2(q)}{t^2} + \frac{A_1(q)}{t} + A_0(q)\right]$, each extra differentiation produces, via the product rule, a polynomial in $1/t$ whose highest power rises by two. □

The first equality of Corollary 1 follows from (14) because $\phi\left(-\frac{1}{2}\sqrt{t} + \frac{1}{\sqrt{t}}L(\hat{q},q)\right) > 0$ for all $(t,q) \in (0,\infty)\times(0,1)$. The second, third, and fourth equalities owe to (7), (8), and (14), by taking the limit $t \to \infty$ or $t \downarrow 0$. [Indeed, $\lim_{t\downarrow 0} \xi(t,q,\hat{q}) = \lim_{s\downarrow 0} \frac{e^{-\frac{1}{s}}}{\sqrt{s}} e^{-s} = 0$ if $q \neq \hat{q}$.] From Lemma 1, Lemma 6, and Theorem 2, we get

$$\frac{\partial^n v(t,q)}{\partial t^n} = \sqrt{\frac{q(1-q)}{\hat{q}^3(1-\hat{q})^3 2\pi}} \frac{1}{e^{\frac{1}{8}t + \frac{1}{2t}L^2(\hat{q},q)}} \frac{1}{\sqrt{t}} \left[\frac{A_{2(n-1)}(q)}{t^{2(n-1)}} + \cdots + \frac{A_1(q)}{t} + A_0(q)\right] = 0.$$

This gives the fifth equality of Corollary 1. □

A.5 Information Lumpiness: Proof of Theorem 5

Because $T_{\mathrm{CH}}(q) \geq T_{FL}(q)$, the last integral is greater than

$$\int_0^{T_{FL}} \frac{e^{-\frac{1}{8}T_{FL} - L^2/2s}}{2\sqrt{s}} ds = \begin{cases} e^{-\frac{1}{8}T_{FL}}\left[\sqrt{T_{FL}}e^{-\frac{L^2}{2T_{FL}}} - \sqrt{2\pi}L\left(1 - \Phi\left(\frac{L}{\sqrt{T_{FL}}}\right)\right)\right] & \forall\ \hat{q} \geq q \\ e^{-\frac{1}{8}T_{FL}}\left[\sqrt{T_{FL}}e^{-\frac{L^2}{2T_{FL}}} + \sqrt{2\pi}L\Phi\left(\frac{L}{\sqrt{T_{FL}}}\right)\right] & \forall\ q \geq \hat{q}. \end{cases} \tag{33}$$

Equation (33) gives that $v(T_{CH}(q),q)/[\hat{q}(1-\hat{q})(M-m)]$ exceeds

$$j(q,\hat{q}) = \begin{cases} \sqrt{\frac{q(1-q)}{2\pi\hat{q}(1-\hat{q})}} e^{-\frac{1}{8}T_{FL}}\left[\sqrt{T_{FL}}e^{-\frac{L^2}{2T_{FL}}} - \sqrt{2\pi}L\left(1 - \Phi\left(\frac{L}{\sqrt{T_{FL}}}\right)\right)\right] & \forall\ \hat{q} \geq q \\ \sqrt{\frac{q(1-q)}{2\pi\hat{q}(1-\hat{q})}} e^{-\frac{1}{8}T_{FL}}\left[\sqrt{T_{FL}}e^{-\frac{L^2}{2T_{FL}}} + \sqrt{2\pi}L\Phi\left(\frac{L}{\sqrt{T_{FL}}}\right)\right] & \forall\ q \geq \hat{q}. \end{cases}$$

It can be shown that $J(\hat{q}) = \int_0^1 j(q,\hat{q})dq$ is convex on $(0,1)$, and is minimized at $\hat{q} = 0.5$. We shall now bound this minimized value. Indeed, $j(\cdot, 0.5)$ is a double-hump shape: It is concave on $(0, 0.5)$ and $(0.5, 1)$, and $\lim_{q\downarrow 0} j(q,0.5) = \lim_{q\to 0.5} j(q,0.5) = \lim_{q\uparrow 1} j(q,0.5) = 0$ because $L(0.5,q) = 0$ and $T_{\mathrm{FL}} = 0$ if $\hat{q} = q = 0.5$, while the limits at $q = 0$ and $q = 1$ require l'Hopital's rule. Further, $j(\cdot, 0.5)$ satisfies $j(r, 0.5) = j(1-r, 0.5)$ where $0 < r < 0.5$. Therefore, we can inscribe between the horizontal $j = 0$ axis and the

$j(q, 0.5)$ curve two equally tall triangles, whose area is a lower bound on $J(0.5)$, namely, $J(0.5) > (0.5)\max_{q\in(0,0.5)} j(q, 0.5)$. Finally, $\max_{q\in(0,0.5)} j(q, 0.5) = j(0.25, 0.5) > 0.05$. □

A.6 Elasticity of Demand: Proof of Theorem 8

Proof of Part (a)**.** Now we show that there exist two points where $T_E(q) = T_{\mathrm{CH}}(q)$ — one when $q < \hat{q}$ and one when $q > \hat{q}$. Let us define the *gross surplus function* γ as follows

$$\gamma(q) \equiv v(T_E(q), q) - v_t(T_E(q), q)T_E(q). \tag{34}$$

It suffices to show that $\gamma(q) > 0$ iff $q \in (\acute{q}, \grave{q})$, where $\hat{q} \in (\acute{q}, \grave{q})$. We will prove that $\gamma(q) > 0$ for $q \in (\acute{q}, \hat{q})$; the case $(\hat{q}, \grave{q})$ follows from symmetry. Differentiating (7) yields:

Claim 2 (q**-Derivatives)** *The belief transition pdf* $q(t)$ *obeys, for* $0<q,r<1$ *and* $t>0$:

$$\begin{aligned} \xi_q(t,q,r) &= \xi(t,q,r)\left[\tfrac{1-2q}{2q(1-q)} + \tfrac{L(r,q)}{tq(1-q)}\right] \\ \xi_{qq}(t,q,r) &= \xi(t,q,r)\tfrac{1}{q^2(1-q)^2}\left[-\tfrac{1}{4} + \tfrac{L^2(r,q)}{t^2} - \tfrac{1}{t}\right]. \end{aligned} \tag{35}$$

Claim 3 *We have* $\gamma(0+) = 0$, $\gamma_q(0+) = -\infty$, *and* $\gamma(\hat{q}) > 0$.

Proof: First we calculate $\gamma(0+)$. Since $\lim_{q\downarrow 0} L(\hat{q}, q) = \lim_{q\downarrow 0} \log\left(\frac{\hat{q}(1-q)}{q(1-\hat{q})}\right) = \infty$, we get $\lim_{q\downarrow 0} T_E(q) = \lim_{q\downarrow 0} L^2(\hat{q}, q) = \infty$ and using (7) and (14), we get

$$\lim_{q\downarrow 0} v_t\left(T_E(q), q\right) T_E(q) = \lim_{q\downarrow 0} \frac{\sqrt{qT_E(q)}}{e^{\frac{T_E(q)}{8}+\frac{1}{2T_E(q)}L^2(\hat{q},q)}} = 0.$$

From (12), the first term of $v(T_E(q), q)$ satisfies

$$\lim_{q\downarrow 0} q(\pi_B^H - \pi_A^H)\Phi\left(\frac{\sqrt{T_E(q)}}{2} - \frac{L(\hat{q}, q)}{\sqrt{T_E(q)}}\right) = 0,$$

because $\Phi(r) \leq 1$ for all r. For the second term of $v(T_E(q), q)$, we have

$$\lim_{q\downarrow 0}(1-q)(\pi_A^L - \pi_A^L)\Phi\left(-\frac{\sqrt{T_E(q)}}{2} - \frac{L(\hat{q}, q)}{\sqrt{T_E(q)}}\right) = 0$$

since $\lim_{q\downarrow 0} L(\hat{q}, q)/\sqrt{T_E(q)} = \lim_{q\downarrow 0} \sqrt{T_E(q)} = \infty$. Thus, $\lim_{q\downarrow 0} \gamma(q) = 0$.

Next we solve $\lim_{r\downarrow 0}\gamma_q(r)$, where

$$\gamma_q(q) = v_q(T_E(q), q) - v_{tt}(T_E(q), q)T_E(q)\tfrac{\partial}{\partial q}T_E(q) - v_{tq}(T_E(q), q)T_E(q). \tag{36}$$

Let us analyze the different terms separately. By Theorem 3, we have $v_{tt}(T_E(q), q) < 0$. Differentiating (24) gives

$$\tfrac{\partial}{\partial q}T_E(q) = -\tfrac{2}{q(1-q)}\tfrac{L(\hat{q},q)}{\sqrt{1+L^2(\hat{q},q)}},$$

and so $\lim_{r\downarrow 0}\frac{\partial}{\partial q}T_E(r) = -\infty$. Because $\lim_{q\downarrow 0}T_E(q) = \infty$, the second term in (36) satisfies $\lim_{q\downarrow 0}v_{tt}(T_E(q), q)T_E(q)\frac{\partial}{\partial q}T_E(q) = \infty$. For the third term of (36):

$$\lim_{r\downarrow 0} v_{tq}(T_E(r), r) = \lim_{q\downarrow 0}\xi(T_E(q), q, \hat{q})\left[\tfrac{1-2q}{2q(1-q)} + \tfrac{L(\hat{q},q)}{T_E(q)q(1-q)}\right] = \infty,$$

because $\lim_{q\downarrow 0}\frac{q}{T_E(q)} = 0$. Since $v_q(T_E, q)$ is bounded, we get $\gamma_q(0+) = -\infty$.

Finally, we calculate the value of $\gamma(\hat{q})$ by using (34). We again divide the expression into different parts. Because $L(\hat{q}, \hat{q}) = 0$, we have $T_E(\hat{q}) = 4$, and so from (14), we get

$$v_t(T_E(\hat{q}), \hat{q})T_E(\hat{q}) = \hat{q}(1-\hat{q})(M-m)\sqrt{\tfrac{1}{2\pi}}e^{-1/2}. \tag{37}$$

Since $T_E(\hat{q}) = 4$, equation (12) gives

$$v(T_E(\hat{q}), \hat{q}) = \hat{q}(1-\hat{q})(M-m)\left(\Phi(1) - \Phi(-1)\right) = \hat{q}(1-\hat{q})(M-m)\sqrt{\tfrac{1}{2\pi}}\int_{-1}^{1} e^{-\frac{x^2}{2}}\,dx. \tag{38}$$

By subtracting (37) from (38), we get $\gamma(\hat{q}) > 0$, since $\int_{-1}^{1} e^{-\frac{x^2}{2}}\,dx > e^{-1/2}$. □

Claim 4 $\gamma(\cdot) : (0, \hat{q}) \to \mathbb{R}$ *is convex.*

Proof: Twice differentiating (34) gives

$$\begin{aligned}-\gamma_{qq}(q) &= v_{ttt}(T_E(q), q)T_E(q)(\tfrac{\partial}{\partial q}T_E(q))^2 + 2v_{ttq}(T_E(q), q)T_E(q)\tfrac{\partial}{\partial q}T_E(q) + v_{tt}(T_E(q), q)(\tfrac{\partial}{\partial q}T_E(q))^2 \\ &\quad + v_{tt}(T_E(q), q)T_E(q)\tfrac{\partial^2}{\partial q^2}T_E(q) + v_{tqq}(T_E(q), q)T_E(q) - v_{qq}(T_E, q).\end{aligned}$$

Substitute from (16), (17), as well as the formulas below from Claim 2 and Theorem 2:

$$\begin{aligned} v_{ttq}(t,q) &= v_t(t,q)\left[\left(-\tfrac{1}{8}+\tfrac{L^2}{2t^2}-\tfrac{1}{2t}\right)\left(\tfrac{1-2q}{2q(1-q)}+\tfrac{L}{tq(1-q)}\right)-\tfrac{L}{t^2q(1-q)}\right] \\ v_{tq}(t,q) &= v_t(t,q)\left[\tfrac{1-2q}{2q(1-q)}+\tfrac{L}{tq(1-q)}\right] \\ v_{tqq}(t,q)/v_{tt}(t,q) &= v_{qq}(t,q)/v_t(t,q) = \tfrac{2}{q^2(1-q)^2} \end{aligned}$$

$$\Longrightarrow \gamma_{qq}(q) = v_t(T_E(q),q)\frac{\left(3+L^2\right)+\left(L^4+4L^2(2+S)+8(1+S)\right)}{(1-q)^2q^2\left(1+L^2\right)^{\frac{3}{2}}S^3} > 0,$$

where $S = 1+\sqrt{1+L^2}$. □

From Claim 3 we get $\gamma(\hat{q}) > 0$ and $\gamma(\varepsilon) < 0$ for $\varepsilon > 0$. Since $\gamma(\cdot)$ is continuous, there exists $\acute{q} \in (0,\hat{q})$ such that $\gamma(\acute{q}) = 0$. The uniqueness follows from Claim 4. Hence, $\gamma(q) > 0$ for $q \in (\acute{q},\hat{q})$. The proof for the existence and uniqueness of $\grave{q} \in (\hat{q},1)$ is symmetric. □

Proof of Part (*b*)**.** Note that when $c \le c_{\mathrm{CH}}(q)$ then $\tau(c,q) > 0$. Let us denote elasticity $E(c) = -c\tau_c(c,q)/\tau(c,q)$. Clearly, $E'(c) > 0$ iff

$$c\tau_c^2(c,q) - \left[\tau_c(c,q)+\tau_{cc}(c,q)c\right]\tau(c,q) > 0. \tag{39}$$

Differentiating $v_t(\tau(c,q),q) = c$ yields $\tau_c(c,q) = 1/v_{tt}$ and $\tau_{cc}(c,q) = -v_{ttt}(\tau,q)/v_{tt}^3(\tau,q)$. Hence, if we substitute from (16) and (17) for v_{tt}/v_t and v_{ttt}/v_{tt}, we get

$$\begin{aligned} c\tau_c^2 - \tau(\tau_c+\tau_{cc}c) &= \frac{v_t}{v_{tt}^2} - \tau\left(\frac{1}{v_{tt}} - \frac{v_{ttt}v_t}{v_{tt}^3}\right) \\ &= -\frac{\tau}{v_{tt}}\left(1-\frac{v_t^2}{v_{tt}^2}\left(\left(\frac{v_{tt}}{v_t}\right)^2 - \frac{L^2}{\tau^3}+\frac{1}{2\tau^2}\right)-\frac{v_t}{\tau v_{tt}}\right) \\ &= -\frac{1}{v_{tt}(v_{tt}/v_t)^2\tau^2}\left(L^2-\frac{1}{2}\tau-\frac{v_{tt}}{v_t}\tau^2\right) \\ &= -\frac{1}{v_{tt}(v_{tt}/v_t)^2\tau^2}\left(\frac{L^2}{2}+\frac{\tau^2}{8}\right) \end{aligned}$$

which is positive because $v_{tt}(\tau,q) < 0$ when $\tau(c,q) > T_{CH}(q)$. Hence, $E(c)$ is rising in the cost c, and thus falling in the quantity τ. □

A.7 Large Demand: Inverting the Inverse Demand Curve in Theorem 9

Claim 5 *Assume that* $\varepsilon(x) > 0$ *is an increasing* C^1 *function of* x*, with* $\varepsilon(x)/x \to 0$*, and* $\varepsilon'(x) = \varsigma/x + O(1/x^2)$*. Then the map* $\psi(x) = x + \varepsilon(x)$ *has inverse* $\beta(x) = x - \delta(x)$ *where* $\delta(x) = \varepsilon(x)(1 - \varsigma/x + O(1/x^2))$*. Furthermore,* $\delta(x) > \varepsilon(x)$ *for all* x*.*

Simply notice that $\varepsilon(x) = \frac{1}{2}\log x + B/x$ obeys the required conditions with $\varsigma = 1/2$.

Proof of Claim: Let $\beta(x) = x - \delta(x)$, for $\delta(x) > 0$ — whose sign is clear, because β and ψ are reflections of each other about the diagonal. Also, since $\psi(x) \to \infty$, so must $\beta(x) \to \infty$, by reflection. By the inverse property, $\psi(x-\delta(x)) \equiv x-\delta(x)+\varepsilon(x-\delta(x)) \equiv x$. Since $x \mapsto \varepsilon(x)$ is increasing, $\delta(x) = \varepsilon(x - \delta(x)) < \varepsilon(x)$.

Taking a first order Taylor series of ε about x yields $\delta(x) = \varepsilon(x) - \delta(x)\varepsilon'(\hat{x}) < \varepsilon(x)$ for some intermediate value $\hat{x} \in [x - \delta(x), x]$. Hence, $\hat{x}/x \geq 1 - \delta(x)/x \geq 1 - \varepsilon(x)/x \to 1$.

$$\delta(x) = \frac{\varepsilon(x)}{1+\varepsilon'(\hat{x})} = \varepsilon(x)\left(1 - \frac{\varepsilon'(\hat{x})}{1+\varepsilon'(\hat{x})}\right) = \varepsilon(x)(1-\varsigma/\hat{x}+O(1/\hat{x}^2)) = \varepsilon(x)(1-\varsigma/\hat{x}+O(1/x^2))$$

A.8 Convergent Belief Processes: Proof of Theorem 10

By Durrett (1996, Theorem 8.7.1), three conditions must be met for weak convergence: The discrete belief process has sample paths that are continuous in the limit with probability one, and the first two moments of the changes in the discrete time process converge to those of the continuous time process. We now verify these in succession:

1. Limit Continuity. Formally, for each $\varepsilon > 0$, we have the unconditional limit:

$$\lim_{\Delta \downarrow 0} \sup_{q\in[0,1]} P\left(\left|q^{\Delta}(n\Delta) - q^{\Delta}((n-1)\Delta)\right| \geq \ \varepsilon \,\middle|\, q^{\Delta}((n-1)\Delta) = q\right) = 0.$$

This follows from the continuity of Bayes rule (27) in the likelihoods $g(Z|\theta)$: As the signal becomes non-informative, beliefs move less and less, so that the above probability equals

$$P\left(\frac{q(1-q)\,|g(Z|H,\Delta) - g(Z|L,\Delta)|}{qg(Z|H,\Delta) + (1-q)\,g(Z|L,\Delta)} \geq \varepsilon\right)$$

which vanishes with Δ for every $Z \in \mathcal{Z}$ as $|g(Z|H,\Delta) - g(Z|L,\Delta)| \to 0$.

2. Convergent First Moments.

$$\lim_{\Delta\downarrow 0}\sup_{q\in[0,1]}\left|\frac{1}{\Delta}\mathbb{E}\left[q^{\Delta}(n\Delta)-q^{\Delta}((n-1)\Delta)|q^{\Delta}((n-1)\Delta)=q\right]-\frac{\mathbb{E}\left[dq(t)|q(t)=q\right]}{dt}\right|=0$$

which holds because the terms vanish due to the martingale property of beliefs.

3. Convergent Second Moments.

$$\lim_{\Delta\downarrow 0}\sup_{q\in[0,1]}\left|\frac{1}{\Delta}\mathbb{V}\left[q^{\Delta}(n\Delta)-q^{\Delta}((n-1)\Delta)|q^{\Delta}((n-1)\Delta)=q\right]-\frac{\mathbb{V}\left[dq(t)|q(t)=q\right]}{dt}\right|=0.$$

By the martingale property, we replace the variance with the expected squared increment:

$$\begin{aligned}\mathbb{V}\left[q^{\Delta}(n\Delta)-q^{\Delta}((n-1)\Delta)|q^{\Delta}((n-1)\Delta)=q\right] &= E\left[(q^{\Delta}(n\Delta)-q)^2|q^{\Delta}((n-1)\Delta)=q\right]\\ &= \int\frac{q^2(1-q)^2[g(Z|H,\Delta)-g(Z|L,\Delta)]^2}{qg(Z|H,\Delta)+(1-q)g(Z|L,\Delta)}dZ.\end{aligned}$$

Using (28), the limit as $\Delta\downarrow 0$ equals $q^2(1-q)^2=\mathbb{V}\left[dq(t)|q(t)=q\right]/dt$ for all $q\in[0,1]$, after the time normalization $\zeta=1$.

A.9 Approximate Value Functions: Proof of Theorem 11

For every $y\geq 0$, let

$$\Omega^{\Delta}(y\Delta|c)\equiv\Pi^{\Delta}\left(\lfloor y+1\rfloor\Delta|c\right)\left[1-(y-\lfloor y\rfloor)\right]^{\frac{1}{\Delta}}+\Pi^{\Delta}\left(\lfloor y\rfloor\Delta|c\right)\left(y-\lfloor y\rfloor\right)^{\frac{1}{\Delta}}.$$

Since $(y-\lfloor y\rfloor)^{\frac{1}{\Delta}}\in[0,1]$, this a weighted average of the maximand $\Pi^{\Delta}(y\Delta|c)$ of program (31) at the two adjacent integers to y. Also, $\Omega^{\Delta}(y\Delta|c)$ is continuous in $y\geq 0$ (continuity at integer values of y can be verified directly), in $c\geq 0$, and in $\Delta>0$. The latter holds because $c\lfloor y\rfloor\Delta$ is clearly continuous in Δ, and $v^{\Delta}\left(\lfloor y\rfloor\Delta\right)=\int u(g(x))f_{X_{\lfloor y\rfloor}}\left(x|\Delta\right)dx$ is continuous in Δ for given sample size $\lfloor y\rfloor$ from previous results. Thus, $\Pi^{\Delta}\left(\lfloor y\rfloor\Delta|c\right)=v^{\Delta}\left(\lfloor y\rfloor\Delta\right)-c\lfloor y\rfloor\Delta$ is continuous in Δ, and $\Omega^{\Delta}(y\Delta|c)$ inherits this property.

Next, $\Omega^{\Delta}\left(\lfloor y\rfloor\Delta|c\right)=\Pi^{\Delta}\left(\lfloor y\rfloor\Delta|c\right)$, i.e. $\Omega^{\Delta}(y\Delta|c)$ is defined over a real-valued y but coincides with the discrete maximand at integer values of y (at multiples of Δ). Also, $\Omega^{\Delta}(y\Delta|c)$ is a weighted average of $\Pi^{\Delta}\left(\lfloor y\rfloor\Delta|c\right)$ and $\Pi^{\Delta}\left(\lfloor y+1\rfloor\Delta|c\right)$, and so of $\Omega^{\Delta}\left(\lfloor y\rfloor\Delta|c\right)$ and $\Omega^{\Delta}\left(\lfloor y+1\rfloor\Delta|c\right)$. Then $\Omega^{\Delta}(y\Delta|c)\leq\max\left\langle\Omega^{\Delta}\left(\lfloor y\rfloor\Delta|c\right),\Omega^{\Delta}\left(\lfloor y+1\rfloor\Delta|c\right)\right\rangle$

with strict inequality if and only if: y is not an integer and $\Pi^{\Delta}(\lfloor y+1 \rfloor \Delta|c) \neq \Pi^{\Delta}(\lfloor y \rfloor \Delta|c)$. Since we can always improve weakly over any y by choosing either $\lfloor y \rfloor$ or $\lfloor y+1 \rfloor$, it follows that the correspondence $M^{\Delta}(c) \equiv \arg\max_{y \geq 0} \Omega^{\Delta}(y\Delta|c)$ always contains a non-negative integer. It follows that

$$\max_{y \geq 0} \Omega^{\Delta}(y\Delta|c) = \max_{n=0,1,2,\ldots} \Pi^{\Delta}(n\Delta|c). \tag{40}$$

Finally, $0 \leq y - \lfloor y \rfloor \leq 1$, and so $(y - \lfloor y \rfloor)^{\frac{1}{\Delta}}$ is a positive function of Δ that vanishes with Δ but remains continuous in y for every $\Delta > 0$. So for every given $t > 0$,

$$\lim_{\Delta \downarrow 0} \Omega^{\Delta}(t|c) = \lim_{\Delta \downarrow 0} v^{\Delta}(t) - ct = v^{0}(t) - ct = \Pi(t|c).$$

We are now ready to use the auxiliary problem of maximizing $\Omega^{\Delta}(y\Delta|c)$ over $y \geq 0$. Again, $\Omega^{\Delta}(0|c) = 0$ and $\lim_{y \to \infty} v^{\Delta}(\lfloor y \rfloor \Delta) \leq \max_{\theta,a} \pi_{a}^{\theta} < \infty$. We can thus restrict the choice of y to a compact interval $[0, \bar{y}(\Delta)]$, where $\bar{y}(\Delta)$ is the continuous function defined by the largest solution m to $v^{\Delta}(m\Delta) = cm\Delta$. Therefore, we can rewrite

$$M^{\Delta}(c) = \arg \max_{y \in [0, \bar{y}(\Delta)]} \Omega^{\Delta}(y\Delta|c).$$

We conclude that y maximizes a function $\Omega^{\Delta}(y\Delta|c)$ continuous in y, c, Δ over a compact-valued and continuous correspondence $[0, \bar{y}(\Delta)]$. Notice that by definition of Ω^{Δ}, a non-integer y belongs to $M^{\Delta}(c)$ iff both $\lfloor y \rfloor$ and $\lfloor y+1 \rfloor$ do. This is a non-generic event (w.r.t. Lebesgue measure over the space of $c > 0$); generically, either there exists one (integer) maximizer in $M^{\Delta}(c)$, or the maximizers are non-consecutive integers, so that $M^{\Delta}(c)$ contains only integers a.e. in parameter space. Since Ω^{Δ} maximized at an integer coincides with the value of the discrete-sample maximand $\Pi^{\Delta}(n\Delta|c)$:

$$\max_{y \in [0, \bar{y}(\Delta)]} \Omega^{\Delta}(y\Delta|c) = \max_{y \in [0, \bar{y}(\Delta)]} \Omega^{\Delta}(\lfloor y \rfloor \Delta|c) = \max_{n=0,1,2,\ldots} \Pi^{\Delta}(n\Delta|c)$$

we conclude that $M^{\Delta}(c) = N^{\Delta}(c)$ a.e. in parameter space. The first maximization above can be rewritten as

$$\max_{t \in [0, \bar{y}(\Delta)\Delta]} \Omega^{\Delta}(t|c)$$

the maximization over a compact-valued and continuous correspondence $[0, \bar{y}(\Delta)\Delta]$ of a function $\Omega^{\Delta}(t|c)$ which is continuous in t, c, Δ. This maximization yields a correspondence $T^{\Delta}(c) = M^{\Delta}(c)/\Delta$ (with a slight abuse of notation).

By Berge's Theorem of the Maximum, the correspondence of maximizers $T^{\Delta}(c)$ is u.h.c. in Δ and c. Hence, as $\Delta \downarrow 0$, some selection $\tau^{\Delta}(c,q) \in T^{\Delta}(c)$ converges to the unique maximizer $\tau(c,q)$ of the continuous time problem $\Pi(t|c) = \lim_{\Delta\downarrow 0} \Omega^{\Delta}(t|c)$: namely, $\lim_{\Delta\downarrow 0} \left|\tau^{\Delta}(c,q) - \tau(c,q)\right| = 0$ a.e. in parameter space. Let $y^{\Delta}(c) \equiv \tau^{\Delta}(c,q)/\Delta \in M^{\Delta}(c)$. This selection must be integer-valued and a maximizer of $\Pi^{\Delta}(n\Delta|c)$ a.e. in parameter space. Thus, $y^{\Delta}(c) = n^{\Delta}(c)$ for some $n^{\Delta}(c) \in N^{\Delta}(c,q)$. Hence, for some choice $n^{\Delta}(c) = \tau^{\Delta}(c,q)/\Delta \in N^{\Delta}(c,q)$ among the optimal discrete sample sizes, a.e. in parameter space,

$$0 = \lim_{\Delta\downarrow 0} \left|\tau^{\Delta}(c) - \tau(c,q)\right| = \lim_{\Delta\downarrow 0} \left|y^{\Delta}(c)\Delta - \tau(c,q)\right| = \lim_{\Delta\downarrow 0} \left|n^{\Delta}(c)\Delta - \tau(c,q)\right|. \qquad \square$$

References

Abreu, Dilip, Paul Milgrom, and David Pearce (1991). "Information and Timing in Repeated Partnerships." *Econometrica*, V. 59-#6. pp. 1713–33.

Black, Fischer and Myron S. Scholes (1973). "The pricing of options and corporate liabilities." *Journal of Political Economy*, V. 81, pp. 637–54.

Chade, Hector and Edward Schlee (2002). "Another Look at the Radner-Stiglitz Non-concavity in the Value of Information." *Journal of Economic Theory*, V.107, pp. 421–52.

Durrett, Richard (1996). *Stochastic Calculus: a Practical Introduction.* CRC Press, Boca Raton, FL.

Harrison, J. Michael and David Kreps (1979). "Martingales and Arbitrage in Multi-period Securities Markets." *Journal of Economic Theory*, V. 20, pp. 381–408.

Harrison, J. Michael and Stanley R. Pliska (1981). "Martingales and Stochastic Integrals in the Theory of Continuous Trading." *Stochastic Processes and Their Applications*, V. 11, pp. 215–60.

Karatzas, Ioannis and Steven E. Shreve (1991). *Brownian Motion and Stochastic Calculus* 2nd Edition. Springer-Verlag, New York, NY.

Kihlstrom, Richard (1974). "A Bayesian model of demand for information about product quality". *International Economic Review*, V. 15, pp. 99–118.

Liptser, Robert S. and Albert N. Shiryayev (2001). *Statistics of Random Processes, I.* 2nd Edition. Springer-Verlag, New York, NY.

Moscarini, Giuseppe and Lones Smith (2001). "The Optimal Level of Experimentation." *Econometrica*, V. 69-#6, pp. 1629–44.

Moscarini, Giuseppe and Lones Smith (2002). "The Law of Large Demand for Information." *Econometrica*, V. 70-#6, pp. 2351–66

Øksendal, Bernt (1998). *Stochastic Differential Equations: An Introduction with Applications.* 5th Edition. Springer-Verlag, New York, NY.

Radner, Roy and Joseph Stiglitz (1984), "A Nonconcavity in the Value of Information" in Marcel Boyer and Richard Kihlstrom, eds. *Bayesian Models in Economic Theory.* Amsterdam: North-Holland, pp. 33–52.