



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

JOURNAL OF
Economic
Dynamics
& Control

Journal of Economic Dynamics & Control 28 (2004) 2003–2035

www.elsevier.com/locate/econbase

Limited information capacity as a source of inertia

Giuseppe Moscarini*

Department of Economics, Yale University, P.O. Box 208268, New Haven, CT 06520-8268, USA

Received 30 April 2001; accepted 15 August 2003

Abstract

We derive optimal time-dependent adjustment rules from Shannon's (1948) Information Theory. In a continuous-time LQ prediction problem with a costly rate of information acquisition-processing, the optimal prediction evolves smoothly, but jumps at an optimally chosen frequency, when fresh information accrues. A more volatile and persistent target raises the information rate required to maintain a sampling frequency. This cost-effect moderates and may even reverse the benefit-effect on the value of information, so optimal inertia is unresponsive to and nonmonotonic in the predictability of the environment. Conventional models with a fixed cost per observation imply no such cost-effect, so inertia rises quickly in the predictability of the environment.

© 2003 Elsevier B.V. All rights reserved.

JEL classification: D83; E31

Keywords: Time-dependent rules; Information capacity; Price stickiness; Inertia

1. Introduction

Many important economic decisions are taken infrequently and discretely. Microeconomic empirical evidence shows—virtually without exceptions—that retail prices, wages, investment in structures and equipment, hiring/firing of workers and purchases of consumer durables are ‘sticky’. This inertia does not wash out in aggregation. The typical spectrum of a macroeconomic time series concentrates most power at low frequencies. Sims (1998) estimates on US post-war data an identified VAR including seven classic macroeconomic variables. The impulse responses show that most variables respond immediately to ‘own’ shocks and only with visible delay to shocks

* Tel.: +1-203-432-3596; fax: +1-203-432-5779.

E-mail address: giuseppe.moscarini@yale.edu (G. Moscarini).

URL: <http://www.econ.yale.edu/~mosca/mosca.html>

to other variables. The important macroeconomic consequences of stickiness are well known.

The search for explanations has drawn attention toward different forms of adjustment costs, which generate inertia as a rational behavior. Existing theories can be broadly classified into two main categories: *state-dependent* rules—the agent acts when a state variable affecting payoffs drifts too far away; and *time-dependent* rules—the agent acts on a predetermined time schedule. The former have been investigated at length, both from a decision-theoretic viewpoint and in terms of their aggregation properties (see Sheshinski and Weiss, 1993 for results on pricing). The latter are typically more ad hoc and less rigorously modelled; yet, they have become the overwhelmingly popular avenue to introduce nominal rigidities in equilibrium macroeconomic models. Typically, a behavioral rule incorporating infrequent adjustment of nominal variables is introduced exogenously as a matter of convenience. The best known examples are staggered wage contracts and Calvo (1983) staggered price-setting, which is a staple of the so-called New New-Keynesian Synthesis (Clarida et al., 1999). Any such rule is obviously subject to the Lucas (1976) critique, but it is justified as a reasonable approximation when performing local dynamic analysis.

This paper addresses optimal time-dependent adjustment rules. We investigate new reasons why such rules may be optimal and relatively invariant to changes in the economic environment. We generate optimal inertia from the frictions in the acquisition and/or processing of salient information that have been identified in Information Theory. Beginning with Shannon's (1948) seminal article, this field has afforded great theoretical and practical advances in our understanding of information flows. In particular, this theory provides a measure of the rate of data acquisition and processing which is at once founded on natural axioms, empirically operational, and analytically convenient. As the economics literature explores various modeling possibilities in a similar direction, it appears natural to first turn to the field that has been studying these very issues for the last 50 years.

In the light of his empirical findings of macroeconomic inertia, Sims (1998) first proposed the observation of data relevant to decision-making through channels of limited transmission rate as a possible explanation for inertial behavior adopted by rational economic agents. His key observation is the following: observing a real-valued process either with no noise, or with noise integrated of the same order as the process, implies an infinite rate of information transmission in the sense of Shannon (1948), which is physically unattainable. Therefore, some form of smoothing of the data, observationally similar to inertial behavior, is necessary. Sims (2003) applies this idea to a Permanent Income model of optimal consumption. In this paper, we exploit a similar insight, but in a different direction: the information rate constraint on decision-making is met by observing and processing infrequently new information, rather than by smoothing it on the frequency domain.

The first contribution of this paper is the formulation of a tractable and classic decision problem incorporating the informational frictions emphasized by Shannon. We introduce a costly rate of information acquisition and/or processing into an otherwise standard linear-quadratic optimal prediction problem—a canonical model for a large class of economically relevant situations, such as pricing, hiring/firing, investment and

portfolio allocation. The information flow is measured in terms of Shannon's axiomatic metric. Changing the prediction or the economic 'action' is costless; but, due to the cost of acquiring and processing information, the decision maker optimally samples the source of noisy information about the target process only infrequently. Hence, he obtains fresh information only infrequently, and accordingly adjusts his action in a lumpy way. In between discrete adjustments, he optimally sets a smooth time path for his action, a path that is pre-determined at the time of the last observation and updating. Inertia emerges not as complete inactivity, but rather as a smoothly changing and perfectly predictable behavior, which responds to random news only infrequently, thus with delay.

For illustration, consider the example of a monopolistically competitive firm, continuously setting the price for its product variety based on its estimate of a stochastic state of demand. This is the most important application of time-dependent adjustment in the current macroeconomic literature. The firm can obtain noisy estimates of the current state of demand, which it must then process and incorporate into its new price. But, it can only acquire and process information at a finite rate, to be allocated to a variety of tasks, including price-setting. Due to the (opportunity) cost of the required information capacity per unit time, the firm optimally chooses to obtain and to process this information only infrequently. At such discrete times, say every quarter, the firm optimally formulates a new, radically different price, and a subsequent deterministic pricing plan for the next 3 months. This plan is valid until the next adjustment, and is based entirely on the expected evolution of demand in the absence of new information. For example, price plans formulated in November take into account the holiday season, but do allow for price variation within the quarter. The price plans are reformulated in February, and so forth. The price keeps changing in a predictable manner, only to jump occasionally at every cycle.

The second and main contribution of this paper is a review of the Lucas critique in this context. We study the response of optimal inertia to changes in the predictability of the environment. If demand innovations become more volatile and persistent, then the price-setting task of the firm becomes informationally more demanding. How does the firm's optimal price-adjustment policy change? There are two forces at play, a (marginal) *benefit-effect* and *cost effect*. As expected, the benefit of additional information rises, as the state of demand is less predictable if not tracked closely. But also *the cost of information acquisition and processing rises, for a given sampling frequency*; absorbing new information about the state of demand (say) once a month is informationally more challenging when news are more variable and persistent. The cost-effect moderates and may even reverse the benefit-effect. So the firm's optimal inertia is typically non-monotone and inelastic in the predictability of demand. A more variable demand may induce even more inertia, because devoting more information capacity to track it may not be worth the effort: the firm is 'paralyzed by complexity'. More generally, optimal inertia is relatively unresponsive to changes in the environment, except for extreme values of parameters.

In contrast, the models of optimal time-dependent or state-dependent policies that have appeared to date in the economic literature assume adjustment costs as functions either of the frequency and precision of sampling (e.g. Caballero, 1989), or of the

frequency and size of adjustment (from the Marginal q theory of investment to the entire S,s literature), but not of the randomness of the environment. Only the benefit-effect is present, thus optimal adjustment tends to be more aggressive when the environment is less predictable, in particular when innovations are more persistent. We show that in a model of this type the optimal policy is extremely responsive to changes in process parameters, because the moderating cost effect is missing. Exogenously specified time-dependent rules, such as those of Calvo (1983) or Gabaix and Laibson (2001), once explicitly modelled may react very differently to policy changes, depending on their microfoundations. In a sense, our approach provides a justification for a higher adjustment cost function in a less predictable environment.

Our finding of non-monotone inertia suggests, specifically for the pricing example, that the undesirable real effects of ‘money surprises’ due to the signal extraction problem (Lucas, 1972) may be exacerbated when the information structure is explicitly modelled. Firms have many things to pay attention to, hence they may settle for more inertial price plans when monetary policy becomes less aggressive. An interest rate smoothing term in a Taylor rule may help firms predict the time path of demand, and thus reduce the real effects of monetary policy. On the other hand, we also find that price inertia is very unresponsive to changes in the economic environment, due to the moderating effect on information costs. This suggests that in this world the Lucas critique has not much bite to begin with, and that an exogenous time-dependent price adjustment rule may be a quite accurate approximation.

Our optimal time-dependent rule is observationally different from Calvo (1983), because here adjustment takes place at all times, not only upon observations (when it is discrete, as in Calvo), and takes into account the mean reversion of the underlying stochastic process. In this sense, it is closer to the ad hoc price-setting rule proposed by Mankiw and Reis (2002). They assume that each firm formulates a pricing plan, that is revised only infrequently to incorporate all the new information that emerged *since* the last adjustment, and observed *without* noise. They show that a ‘sticky-information’ Phillips curve emerges, whose predictions for the timing and persistence of the real effects of monetary policy and for inflation inertia match the data much more closely than the New New-Keynesian Phillips curve. Strictly speaking, their pricing rule is inconsistent with our microfoundations, which imply that the (necessarily) *noisy* information obtained only *at* the time of adjustment is incorporated into the new ‘time plan’, valid for the next interval of inertia. But, it is reasonable to expect that the aggregate implications will be similar.

Limited information processing capacity has played a role in the study of organizations. van Zandt (1999), who also surveys the pertinent literature, assumes that each member of an organization can perform at most a given number of arithmetic operations per unit time, and has finite memory. The question is how should organizations be designed to process optimally information toward a centralized final decision. Unlike the present paper, this approach relies on bounded rationality and chooses a different type of processing constraint in designing the informational friction.

Section 2 introduces the prediction problem, Section 3 illustrates the basic results from Information Theory behind our information constraints, Section 4 sets up the two models of optimal sampling, Section 5 presents the comparative statics effects of

changes in parameters of the decision problem on the optimal degree of inertia and Section 6 concludes.

2. A continuous prediction problem with infrequent sampling

2.1. Setup

A decision maker (DM) aims to predict continuously over a time horizon $t \in [0, \infty)$ a ‘target’ Ornstein–Uhlenbeck process $\langle X \rangle$:

$$dX_t = -\alpha X_t dt + \sigma dW_t, \tag{2.1}$$

where $\alpha > 0$, $\sigma > 0$, $X_0 \sim N(x_0, \phi_0^{-1})$ for some $x_0 \in \mathbb{R}$, $\phi_0 > 0$, and $\langle W \rangle$ is a Wiener process. By a result in Karlin and Taylor (1981), the unconditional distribution of X_t is

$$X_t \sim N(x_0 e^{-\alpha t}, h_t^{-1}),$$

where

$$\begin{aligned} \omega &\equiv \frac{2\alpha}{\sigma^2}, \\ h_t &\equiv \frac{\omega}{1 - e^{-2\alpha t}}. \end{aligned} \tag{2.2}$$

This process $\langle X \rangle$ is stationary, with

$$X_t \xrightarrow{d} X_\infty \sim N(0, \omega^{-1}).$$

Hence, ω is a measure of predictability of this process: the lower the variance of innovations σ^2 and the higher the mean-reversion of the process α , the higher the asymptotic precision ω of X_t and the easier to forecast the value of X_t based only on initial conditions $X_0 \sim N(x_0, \phi_0^{-1})$. As $\alpha \downarrow 0$, clearly $\omega \downarrow 0$, so the process becomes non-stationary (a random walk) and unpredictable in the long run without additional information. In the short run, by Hopital $\lim_{\alpha \downarrow 0} h_t = 1/(t\sigma^2)$.

The DM observes the realizations of the target $\langle X \rangle$ with noise. Specifically, for each t , he can observe a signal Y_t such that

$$Y_t = X_t + Z_t,$$

where Z_t is a draw from an $N(0, \gamma^{-1})$ random variable. Every draw is i.i.d. and independent of the process $\langle W \rangle$. The DM observes $\langle Y \rangle$ every $\Delta \geq 0$ periods. For $\Delta > 0$, observations are countable and the sample number is denoted by $\tau = 1, 2, \dots$, so that the τ th sample $Y_{\tau\Delta}$ occurs at calendar time $t = \tau\Delta$. For the time being, we take $\Delta > 0$ as given; later, we will analyze the choice of Δ . The two extreme cases are $\Delta \rightarrow 0$, equivalent to continuous observations, and $\Delta \rightarrow \infty$, equivalent to no observation at all.

The information flow from observing $Y_\Delta, Y_{2\Delta}, \dots$ generates a filtration $\{\mathcal{F}_t^{(A)}\}$ on the underlying probability space. Here $\mathcal{F}_t^{(A)}$ is the σ -algebra generated by $\{Y_\Delta, Y_{2\Delta}, \dots, Y_{\tau\Delta}\}$

where τ is the largest integer such that $\tau\Delta \leq t$.¹ The DM then chooses an $\{\mathcal{F}_t^{(\Delta)}\}$ -adapted action process $\langle a \rangle$ to minimize the average MSE per unit time:

$$L(\Delta) = \inf_{\langle a \rangle} \lim_{T \rightarrow \infty} \mathbb{E} \left[\frac{\int_0^T (a_t - X_t)^2 dt}{T} \middle| \mathcal{F}_0^{(\Delta)} \right],$$

where for simplicity $\mathcal{F}_0^{(\Delta)}$ is the null σ -algebra, so the first observation arrives at time Δ (not at time 0). As standard in these problems, the realization of the loss is either unobserved or observed with sufficient delay that the information about $\langle X \rangle$ that it contains is not useful to the current or future prediction.

2.2. Optimal prediction

Given the quadratic loss function, the best prediction a_t^* is the conditional expectation of the target and the resulting loss is the conditional variance. Let $t = (\tau + s)\Delta$ for $s \in [0, 1)$. Then the optimal prediction at time t is $a_{(\tau+s)\Delta}^* = \mathbb{E}[X_{(\tau+s)\Delta} | \mathcal{F}_{(\tau+s)\Delta}^{(\Delta)}]$, which gives an expected ‘flow’ loss $\mathbb{V}[X_{(\tau+s)\Delta} | \mathcal{F}_{(\tau+s)\Delta}^{(\Delta)}]$. Note that at almost all times $(\tau + s)\Delta$, except when observing Y , the DM must formulate his conjecture on the observations up to time $\tau\Delta$ and on the time elapsed since, $s\Delta$. Intuitively, in between sampling times, spaced Δ time apart, the unobserved process $\langle X \rangle$ moves away undetected.

By a standard result in Kalman filtering and the independence of the target innovations W , for all $s \in [0, 1)$:

$$\begin{aligned} \hat{X}_{(\tau+s)\Delta} &\equiv \mathbb{E}[X_{(\tau+s)\Delta} | \mathcal{F}_{(\tau+s)\Delta}^{(\Delta)}] \\ &= e^{-zs\Delta} \mathbb{E}[X_{\tau\Delta} | \mathcal{F}_t^{(\Delta)}] \\ \frac{1}{\phi_{(\tau+s)\Delta}} &\equiv \mathbb{V}[X_{(\tau+s)\Delta} | \mathcal{F}_{(\tau+s)\Delta}^{(\Delta)}] \\ &= e^{-2zs\Delta} \mathbb{V}[X_{\tau\Delta} | \mathcal{F}_t^{(\Delta)}] + h_{s\Delta}^{-1} \\ &= e^{-2zs\Delta} \mathbb{V}[X_{\tau\Delta} | \mathcal{F}_t^{(\Delta)}] + (1 - e^{-2zs\Delta}) \mathbb{V}[X_\infty | \mathcal{F}_t^{(\Delta)}]. \end{aligned} \tag{2.3}$$

Hence, the variance of the target in between samples (for $s \in (0, 1)$) conditional on observations up to $Y_{\tau\Delta}$, is a weighted average of the variance $\mathbb{V}[X_{\tau\Delta} | \mathcal{F}_{\tau\Delta}^{(\Delta)}]$ at the moment of last observation, and of the asymptotic variance $\mathbb{V}[X_\infty | \mathcal{F}_{\tau\Delta}^{(\Delta)}] = 1/\omega$ that would be estimated in the long run, absent from new observations. The time $s\Delta$ elapsed since the last received observation $Y_{\tau\Delta}$ only affects the weights.

An important implication of (2.3) is that, in order to characterize the optimal prediction and the resulting flow loss at every point in time, we only need do so at sampling

¹ Note that the notation \mathcal{F}_t , without superscript (Δ) for the sampling policy, would be inappropriate: by time $t = \tau\Delta$ one could have observed τ observations every Δ periods, or 2τ observations every $\Delta/2$ periods, etc., and the resulting filtrations would differ.

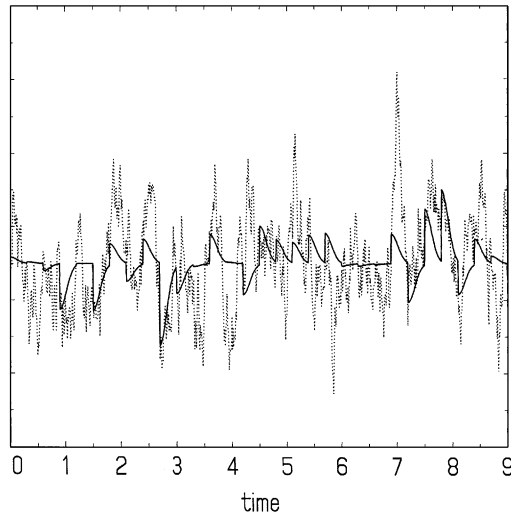


Fig. 1. A sample path of the target X_t and of the optimal prediction \hat{X}_t for $\gamma = 1$, $\sigma = 5$, $\alpha = \ln 2$ ($\rho = 0.5$). The sampling interval is $\Delta = 0.3$, the optimal one with quadratic costs of the information rate implied by the chosen parameter values.

times $\tau\Delta = \Delta, 2\Delta, \dots$. The resulting average loss is

$$L(\Delta) = \lim_{T \rightarrow \infty} \frac{\int_0^T \mathbb{V}[X_t | \mathcal{F}_0^{(\Delta)}] dt}{T}.$$

This model differs from both the continuous time and the discrete time version of the Kalman filter: while the state evolution and the prediction (action a_t) occur continuously, observations $Y_{\tau\Delta}$ accrue infrequently. So at almost all times the DM bases his prediction of X_t on information that is partly outdated. The DM exhibits inertia, in that his optimal action evolves smoothly in the absence of new information, according to a pre-determined *adjustment plan*, only to receive a random discrete innovation when fresh information arrives at times $\Delta, 2\Delta, \dots$. Fig. 1 illustrates a typical sample path of the target and of the corresponding optimal prediction, for a particular parameterization of the model: $\gamma = 1$, $\sigma = 5$, $\alpha = \ln 2$ (corresponding to $e^{-\alpha} = 0.5$), and $\Delta = 0.3$, which is the optimal sampling interval for these parameters in the Costly Information rate model of Sections 4 and 5.

This is a canonical model of many situations of economic interest. The quadratic loss has the familiar interpretation of a second-order Taylor expansion of a full-fledged objective function. The cost of gathering and processing information (to be modelled later) dominates the ‘menu’ cost of taking the action (here assumed to be zero), so the DM settles on a time-dependent optimal adjustment policy. The sampling interval Δ is the natural *measure of inertia*. We can think of the DM as setting a new course of action $a_{\tau\Delta}^*$ at every sampling time $\tau\Delta$, and starting from there a deterministic and smooth action plan, valid until the next observation.

Our leading example is particularly relevant for macroeconomic models where ad hoc staggered price adjustment makes money non-neutral. A monopolistically competitive firm faces demand $q_t = q(p_t, X_t)$ for the variety of the goods it produces with linear technology, where p_t is the price of output in units of labor and X_t is a demand shock. The firm chooses in continuous time a price process p_t that maximizes average expected profits $\lim_{T \rightarrow \infty} \{T^{-1} \mathbb{E}[\int_0^T q(p_t, X_t)(p_t - 1) dt]\}$, where the expectation is taken over the firm’s current beliefs about the state of demand X_t . The firm pays no menu cost to change its price. Every Δ periods, the firm observes an estimate of demand $Y_{\tau\Delta} = X_{\tau\Delta} + Z_{\tau\Delta}$, updates its beliefs, and chooses the optimal price $p^*(\hat{X}_{\tau\Delta}, \phi_{\tau\Delta})$, which depends only on mean $\hat{X}_{\tau\Delta}$ and variance $\phi_{\tau\Delta}^{-1}$ because of the Gaussian structure. Henceforth, the firm sets a smooth price plan $p^*(e^{-\alpha s\Delta} \hat{X}_{\tau\Delta}, \phi_{(\tau+s)\Delta})$ for the next period of time of length Δ , and implements it mechanically until the new cycle, when the price is revised to $p^*(\hat{X}_{(\tau+1)\Delta}, \phi_{(\tau+1)\Delta}) \neq \lim_{s \uparrow 1} p^*(\hat{X}_{(\tau+s)\Delta}, \phi_{(\tau+s)\Delta})$. The price is never fixed, but responds almost always with delay to changes in fundamentals X .

2.3. Filtering

In order to formulate the optimal prediction a_t^* at every point in time, the DM must compute the posterior probability distribution of X_t at times $\tau\Delta = \Delta, 2\Delta, \dots$, conditional on the history of observations $\{\mathcal{F}_t^{(A)}\}$. Extrapolating from Karlin and Taylor (1981, pp. 345–346), we may write the discrete time process at sampling times $t = \tau\Delta$, as an AR(1):

$$X_{\tau\Delta} = e^{-\alpha\Delta} X_{(\tau-1)\Delta} + \varepsilon_{\tau\Delta}, \tag{2.4}$$

where $\varepsilon_{\tau\Delta} = \sigma \int_{(\tau-1)\Delta}^{\tau\Delta} e^{\alpha(s-\tau\Delta)} dW_s \sim N(0, h_\Delta^{-1})$. $\{\varepsilon_{\tau\Delta}\}$ is clearly an i.i.d. sequence. Similarly,

$$Y_{\tau\Delta} = X_{\tau\Delta} + Z_{\tau\Delta}, \tag{2.5}$$

where $Z_{\tau\Delta} \sim N(0, \gamma^{-1})$ is i.i.d. and independent of $\varepsilon_{\tau\Delta}$.

For fixed $\Delta > 0$, system (2.4)–(2.5) is a pair of linear stochastic difference equations driven by Gaussian white noise, a state-space representation of target and observations, which can be analyzed via the Kalman filter. By the conjugate property of the normal, the posterior $X_{\tau\Delta} | \mathcal{F}_{\tau\Delta}^{(A)}$ is normal with mean $\hat{X}_{\tau\Delta}$ and precision $\phi_{\tau\Delta}$. At sampling times, $\tau\Delta = \Delta, 2\Delta, \dots$ these two moments evolve as follows. First, the new precision is given by

$$\begin{aligned} \phi_{\tau\Delta} &= h_\Delta - \frac{h_\Delta^2 e^{-2\alpha\Delta}}{e^{-2\alpha\Delta} h_\Delta + \phi_{(\tau-1)\Delta}} + \gamma \\ &= \frac{\omega \phi_{(\tau-1)\Delta}}{\omega e^{-2\alpha\Delta} + \phi_{(\tau-1)\Delta} (1 - e^{-2\alpha\Delta})} + \gamma, \end{aligned} \tag{2.6}$$

where the first equality follows from the so-called *information filter version* of the Kalman filter (Elliott et al., 1995, Chapter 4, Theorem 5.1), the second uses simple algebraic manipulations and the definition of h_Δ . Note that the first equality can be

rewritten as

$$\begin{aligned} \frac{1}{\mathbb{V}[X_{\tau\Delta}|\mathcal{F}_{\tau\Delta}^{(A)}]} &= \frac{1}{\frac{e^{-2\alpha\Delta}}{\phi_{(\tau-1)\Delta}} + \frac{1}{h_\Delta}} + \gamma \\ &= \frac{1}{\mathbb{V}[e^{-\alpha\Delta}X_{(\tau-1)\Delta} + \varepsilon_{\tau\Delta}|\mathcal{F}_{(\tau-1)\Delta}^{(A)}]} + \frac{1}{\mathbb{V}[Z_t]} \\ &= \frac{1}{\mathbb{V}[X_{\tau\Delta}|\mathcal{F}_{(\tau-1)\Delta}^{(A)}]} + \frac{1}{\mathbb{V}[Z_t|\mathcal{F}_{(\tau-1)\Delta}^{(A)}]}. \end{aligned}$$

The new posterior precision of $X_{\tau\Delta}$ is the precision of the *same* value $X_{\tau\Delta}$ based on past information, plus the precision of the noise in observing the new noisy signal $Y_{\tau\Delta} = X_{\tau\Delta} + Z_{\tau\Delta}$.

Given this updated precision, the new posterior expectation is

$$\hat{X}_{\tau\Delta} = e^{-\alpha\Delta}\hat{X}_{(\tau-1)\Delta} + \frac{\gamma}{\phi_{\tau\Delta}}(Y_{\tau\Delta} - e^{-\alpha\Delta}\hat{X}_{(\tau-1)\Delta}), \tag{2.7}$$

namely, the last updated expectation discounted by the time change, $e^{-\alpha\Delta}\hat{X}_{(\tau-1)\Delta} = \mathbb{E}[X_{\tau\Delta}|\mathcal{F}_{(\tau-1)\Delta}^{(A)}]$, plus the deviation between the new observation $Y_{\tau\Delta}$ and this discounted expectation $e^{-\alpha\Delta}\hat{X}_{(\tau-1)\Delta}$, multiplied by a gain $\gamma/\phi_{\tau\Delta}$. The initial conditions are ϕ_0 and $\hat{X}_0 = X_0 + Z_0$. The corresponding mean and precision at all other times $t = (\tau + s)\Delta$, $s \in (0, 1)$ are obtained by applying (2.3).

Since the objective function is an average expected loss, and an average of the posterior variances, it is of particular interest to characterize the long-run behavior of the posterior precision and its dependence on the sampling interval. As $\Delta \rightarrow 0$, the process $\langle X \rangle$ is observed continuously. The path of posterior precision is discontinuous at time 0, where it exhibits a jump of magnitude γ : from Eq. (2.6), $\lim_{\Delta \rightarrow 0} \phi_\Delta = \phi_0 + \gamma > \phi_0$. Intuitively, an observation of fixed precision γ is obtained in any infinitesimal time interval. Thus, the posterior precision of the target undergoes a discrete jump γ for every instant, and is infinite at any $t > 0$: by *large numbers*, the process can be estimated perfectly with an infinite sequence of informative observations obtained in $[0, t]$. Only if γ were proportional to Δ we would obtain the familiar updating equation for the continuous time normal learning model, $\dot{\phi}_t = \gamma$, where posterior precision explodes linearly in time.

As $\Delta \rightarrow \infty$, the process X is never observed, and posterior precision converges to a constant: again from (2.6), $\lim_{\Delta \rightarrow \infty} \phi_{\tau\Delta} = \omega + \gamma$ for every τ .

For every other interval length $\Delta \in (0, \infty)$, we have the following general characterization.

Lemma 1 (Dynamics of the posterior precision). *For every fixed non-degenerate sampling interval $\Delta \in (0, \infty) < \infty$, and every $\phi_0 \geq 0$, $\alpha > 0$, ω , and γ , as the number of observations τ grows unbounded, the posterior precision $\phi_{\tau\Delta}$ at sampling times $\tau\Delta = \Delta, 2\Delta, \dots$ converges monotonically to*

$$\lim_{\tau \rightarrow \infty} \phi_{\tau\Delta} = \Phi(\alpha\Delta|\gamma, \omega) \equiv \frac{\omega + \gamma}{2} + \sqrt{\left(\frac{\omega + \gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}}.$$

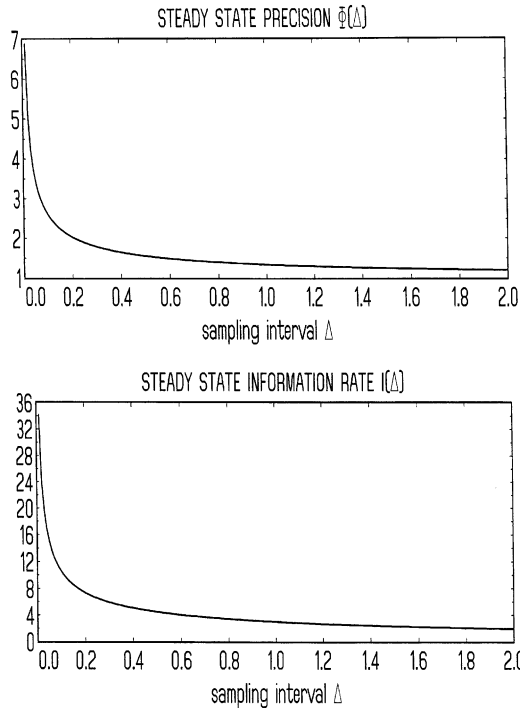


Fig. 2. Steady-state asymptotic posterior precision $\Phi(\alpha\Delta|\gamma, \omega)$ and information rate $\hat{I}(\Delta)$ (in bits per unit time) as a function of Δ , for $\gamma = 1$, $\sigma = 5$, $\alpha = \ln 2$ ($\rho = 0.5$).

This global attractor satisfies $\Phi(0+|\gamma, \omega) = \infty$, $\Phi(\infty|\gamma, \omega) = \omega + \gamma > 0$, $\Phi'(\alpha\Delta|\gamma, \omega) < 0 < \Phi''(\alpha\Delta|\gamma, \omega)$, $\Phi'(0+|\gamma, \omega) = -\infty$, $\Phi'(\infty|\gamma, \omega) = 0$. Finally, $\Phi(\alpha\Delta|\gamma, \omega)$ increases with ω and γ , and decreases in α given ω , that is as α and σ^2 rise proportionally. In between sampling times, for any $s \in (0, 1)$, precision equals:

$$\phi_{(\tau+s)\Delta} = \left[\frac{e^{-2zs\Delta}}{\phi_{\tau\Delta}} + \frac{1 - e^{-2zs\Delta}}{\omega} \right]^{-1}$$

and therefore converges to $[(e^{-2zs\Delta}/\Phi(\alpha\Delta|\gamma, \omega)) + (1 - e^{-2zs\Delta}/\omega)]^{-1}$ as τ grows unbounded.

The upper panel of Fig. 2 illustrates the shape of the fixed point Φ as a function of the sampling interval Δ , for $\gamma = 1$, $\sigma = 5$, $\alpha = 0.34$, as in Fig. 1 (but now Δ varies between 0 and 2).

2.4. Average prediction loss from a given sampling policy

Using the asymptotic properties of the posterior precision, we can obtain a simple closed-form expression for the loss function.

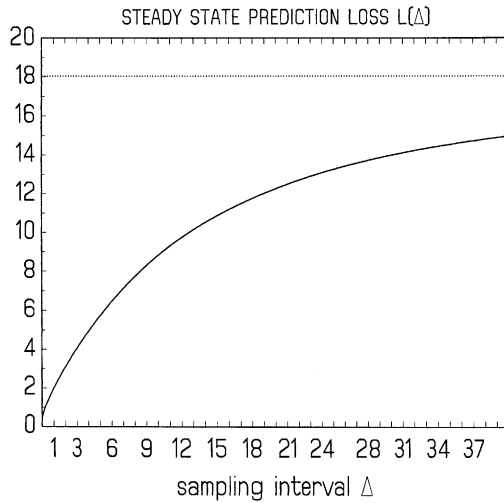


Fig. 3. Average expected loss $L(\alpha\Delta|\gamma, \omega)$ as a function of Δ , for $\gamma = 1$, $\sigma = 5$, $\alpha = \ln 2$ ($\rho = 0.5$). The upper horizontal line equals $L(\infty|\gamma, \omega) = \omega^{-1}$.

Proposition 1 (The average prediction loss). *For every fixed non-degenerate sampling interval $\Delta \in (0, \infty) < \infty$, and every $\phi_0 \geq 0$, $\alpha > 0$, ω , and γ , the time-0 expected average loss from the optimal prediction policy $\langle a^* \rangle$ equals*

$$L(\alpha\Delta|\gamma, \omega) = \frac{1}{\omega} \left(1 - \frac{1 - e^{-2\alpha\Delta}}{2\alpha\Delta} \right) + \frac{1 - e^{-2\alpha\Delta}}{2\alpha\Delta} \frac{1}{\Phi(\alpha\Delta|\gamma, \omega)}, \tag{2.8}$$

which is strictly increasing and concave in Δ , with $L(0|\gamma, \omega) = 0$ and $L(\infty|\gamma, \omega) = 1/\omega$. For given Δ , $L(\alpha\Delta|\gamma, \omega)$ is strictly decreasing in ω and γ given α , and increasing in α given ω .

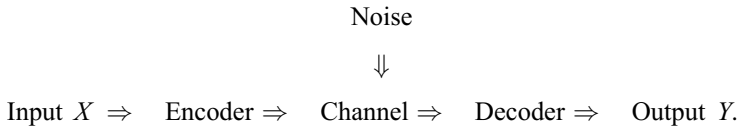
The expected average loss is a weighted average of the asymptotic variance $1/\omega$ and of the posterior variance $1/\Phi$ at sampling times of the target, with weight on the former increasing in Δ as sampling becomes less and less frequent. The concavity of the indirect loss function implies increasing returns to sampling. Fig. 3 illustrates the shape of the prediction loss L as a function of the sampling interval Δ , in our previous baseline parameterization $\gamma = 1$, $\sigma = 5$, $\alpha = \ln 2$.

In order to close the model, we need to describe the optimal sampling decision. The DM would like to sample the observation process continuously ($\Delta = 0$), as its expected average total loss would fall to a minimum ($L(0) = 0$). The following section motivates infrequent sampling via a limited information capacity constraint. We then compare the implications of this model to those derived from imposing an exogenous cost function for sampling.

3. Information rate in information theory

3.1. Definition

The novel element in this paper is the specification of the technological constraint on the rate at which payoff-relevant information may be obtained and processed. The constraint is specified in terms of a physical measure, available for any source of (statistical) information at some cost to the DM. The definition of such a reasonable notion of quantity of information, independent of preferences and depending only on the means of communication and processing available, is a major achievement of Information Theory, an applied mathematical discipline that has secured great practical advances in communication and information technology. It is useful to first recall a few basic notions from this field (see Ash, 1965). The basic model of communication is represented as follows:



Formally, the *channel* is a conditional probability distribution $P(Y|X)$, yielding the chance of output Y given input X . The ex ante occurrence of possible inputs from the point of view of the final observer is also described by a distribution $P(X)$. Information transmission is noiseless when $P(Y|X)$ is degenerate, so that the input is transmitted exactly, otherwise it is noisy and always entails some error.

The *measure of uncertainty* is a function over probability measures derived from four natural axioms. Suppose X may take one of M values with chances p_1, \dots, p_M . Denote the uncertainty associated to X by $H(p_1, \dots, p_M)$. The axioms are: (A1) $H(1/M, 1/M \dots 1/M) \equiv f(M)$ is increasing in M ; (A2) $f(ML) = f(M) + f(L)$; (A3) The grouping axiom: let $R \equiv \sum_{i=1}^r p_i$, $Q = 1 - R$ for some $r < M$, and require

$$H(p_1, \dots, p_M) = H(R, Q) + R \cdot H(p_1/R, \dots, p_r/R) + Q \cdot H(p_{r+1}/Q, \dots, p_M/Q);$$

and finally (A4) $H(p, 1 - p)$ is continuous in p . Shannon (1948) proved that the unique function satisfying all four axioms is the Entropy function $H(\vec{p}) = H(X) = -k \sum p_i \log_b p_i$ for any $k > 0$, $b > 1$. Axioms (A1)–(A3) and the entropy function have several natural interpretations, and both can be extended to continuous and conditional distributions.

Based on this metric of uncertainty, the *measure of information transmitted* by the channel is defined as the reduction in entropy attained by observing the input process: $I(X|Y) = H(X) - H(X|Y)$. If the logarithm is base 2, the information is measured in *bits*.

Finally, the *channel capacity* is the supremum of $I(X|Y)$ over prior distributions of the input X , i.e. the maximum amount of information per unit time the channel may transmit. The Fundamental Theorem of Information Theory states that there exists a code (encoder and decoder, namely a ‘language’) that achieves the transmission of

information at full capacity with an arbitrarily small chance of error. Conversely, any transmission at rate exceeding capacity entails a non-negligible error.²

3.2. Measuring information processing capacity

The frictions that human beings face in communication and information acquisition have to do with understanding information as much as with transmitting or receiving it correctly. A good proxy for understanding received information is information *processing*. In this respect, Information Theory appears to be a promising avenue. Although this theory is concerned with information transmission, the measure of information that it introduces can also be used to gauge the speed of some types of information processing.

To illustrate this point, consider the following example, relative to how the reader can measure his or her own capacity to process and understand information written in English, such as the content of this paper. Given any word, say ‘inertia’, one can think of the word following it in the text as a random variable w , taking values $\{w_i\}$ in a Dictionary and in punctuation signs. The probability p_i of each realization w_i is the frequency with which w_i follows ‘inertia’ in existing English language. Therefore, the entropy of the word following ‘inertia’ in English is $I(\text{inertia}) = -\sum_i p_i \log p_i$, with the convention $0 \times \log 0 = 0$ standard in this Theory. Reading and understanding the word w after ‘inertia’ amounts to completely resolving the uncertainty associated with w , thereby processing an amount of information equal to $I(\text{inertia})$. The same procedure can be used to compute the entropy of the very first word in a text. Just take the relative frequency with which each word appears first in an English text. Once the total amount of information contained in (say) a page of text has been computed by summing the entropy of all words in the page, $I = \sum_w I(w)$, the time T taken by the reader to read and understand that page provides an estimate of his or her information processing rate in bits/second, namely I/T . This is one interpretation of the constraint on decision-making introduced in this paper. From a formal viewpoint, it is immaterial whether our agent faces a cost of increasing his rate of information acquisition or processing.

3.3. Information rate in Kalman filtering

In a continuous Gaussian setting like the one analyzed here, the Fundamental Theorem requires an exogenous limitation on the variability of the input (the ‘power’ of the source), of the type $\Pr(\int_0^1 X_t^2 dt \leq M) = 1$ for some $M < \infty$. Only in this case can channel capacity be computed and be shown to be positively related to M . Without

²The idea is to transform the input X through a language (encoder), using short ‘words’ for frequent inputs, and reserve longer, time-consuming words for infrequent inputs. The longer the ‘word’ attached to each input, the more likely it is to identify it when transmitted with error. For instance, ‘exceedingly’ may be transmitted as ‘exceedinghy’ with little possibility of confusion, while ‘too’ (which has a similar meaning as ‘exceedingly’) can become ‘toe’, or ‘boo’, which are much harder to interpret. But longer words take more time to transmit. The cost of precise encoding is delay. Real-world transmission of information (TV, books, etc.) is governed by this theory but always occurs well below channel capacity.

such a bound the information transmitted without error may achieve an infinite rate, which is physically impossible. This is the standard assumption made in all engineering applications of Information Theory.

Since this type of almost-sure constraint cannot be justified for economic time series, and is violated by any unbounded variation process such as a diffusion, we abstract from coding issues and exploit this fundamental insight of Information Theory. If the input (X) variability is not exogenously constrained at the source and the rate of (noisy) information transmission and or/processing is to remain finite, then the source of information must be sampled infrequently. We then envision the total rate of information acquisition/processing as a scarce resource, that the DM may either acquire through an upfront investment, or own as an individual skill to be allocated optimally among various informational tasks.

The entropy of a Gaussian kernel is proportional to the log of its standard error. Therefore, in the Gaussian context of this Kalman Filter, the information conveyed by (or contained in) the noisy observation $Y_{\tau\Delta}$ of $X_{\tau\Delta}$ is one half of the reduction in the log-variance of beliefs about the target. The additional information transmitted by observation $Y_{\tau\Delta}$ is therefore

$$I_{\tau\Delta} = \frac{1}{2\Delta} \log \frac{\mathbb{V}[X_{\tau\Delta} | \mathcal{F}_{(\tau-1)\Delta}^{(d)}]}{\mathbb{V}[X_{\tau\Delta} | \mathcal{F}_{\tau\Delta}^{(d)}]} = \frac{1}{2\Delta} \log \frac{\phi_{\tau\Delta}}{\phi_{\tau\Delta} - \gamma} \tag{3.1}$$

The second equality derives from the *measurement update equations* of the Kalman filter (Elliott et al., 1995, 5.13), which express the posterior variance Δ time after the latest observation $\mathbb{V}[X_{\tau\Delta} | \mathcal{F}_{\tau\Delta}^{(d)}] = \phi_{\tau\Delta}^{-1}$ as a function of the variance conditional only on previous observations $\mathbb{V}[X_{\tau\Delta} | \mathcal{F}_{(\tau-1)\Delta}^{(d)}]$.

Using (2.6) to replace for $\phi_{\tau\Delta}$ we obtain

$$\begin{aligned} I_{\tau\Delta} &= \frac{1}{2} \log \frac{\frac{\omega\phi_{(\tau-1)\Delta}}{\phi_{(\tau-1)\Delta}(1-e^{-2\alpha\Delta})+\omega e^{-2\alpha\Delta}} + \gamma}{\frac{\omega\phi_{(\tau-1)\Delta}}{\phi_{(\tau-1)\Delta}(1-e^{-2\alpha\Delta})+\omega e^{-2\alpha\Delta}}} \\ &= \frac{1}{2} \log \frac{\omega\phi_{(\tau-1)\Delta} + \gamma[\phi_{(\tau-1)\Delta}(1 - e^{-2\alpha\Delta}) + \omega e^{-2\alpha\Delta}]}{\omega\phi_{(\tau-1)\Delta}} \\ &= \frac{1}{2} \log \left(1 + \gamma \frac{\phi_{(\tau-1)\Delta}(1 - e^{-2\alpha\Delta}) + \omega e^{-2\alpha\Delta}}{\omega\phi_{(\tau-1)\Delta}} \right) \\ &= \frac{1}{2} \log \left(1 + \gamma \frac{1 - e^{-2\alpha\Delta}}{\omega} + \gamma \frac{e^{-2\alpha\Delta}}{\phi_{(\tau-1)\Delta}} \right). \end{aligned}$$

This quantity has several intuitive properties. It is always positive for any finite posterior precision $\phi_{(\tau-1)\Delta}$ held before observing $Y_{\tau\Delta}$. It is decreasing in $\phi_{(\tau-1)\Delta}$, as less information is contained in each $N(0, \gamma^{-1})$ -error ridden observation $Y_{\tau\Delta}$ the more precisely the target is predicted to begin with. In fact,

$$I_{\tau\Delta} \rightarrow \frac{1}{2} \log \left(1 + \gamma \frac{1 - e^{-2\alpha\Delta}}{\omega} \right) > 0 \quad \text{as } \phi_{(\tau-1)\Delta} \rightarrow \infty.$$

Even if the target is known exactly at time $(\tau - 1)\Delta$, it does drift away in the Δ -interval before one obtains a new observation at time $\tau\Delta$, which then always contains some information. By the same token, $I_{\tau\Delta}$ grows without bound as $\phi_{(\tau-1)\Delta} \rightarrow 0$. Also, $I_{\tau\Delta}$ is zero for $\gamma = 0$, as an observation ridden by noise of infinite variance is useless; it grows without bound in γ , because a perfect observation of a real number X_t (when $\gamma = \infty$) contains an infinite amount of information. Finally, given an initial precision ϕ_0 , the quantity of information contained in the first observation after Δ time, as Δ vanishes, converges to $\lim_{\Delta \rightarrow 0} I_{\Delta} = \frac{1}{2} \log(1 + (\gamma/\phi_0))$.

The most important lesson is that the information contained in each observation depends not only on the precision of the observation γ , but also on the predictability of the process, through ω and α , and on the prior probability distribution, through $\phi_{(\tau-1)\Delta}$. A given observation contains more information, the less predictable is the starting point and the subsequent evolution of what we are being informed about.

Information capacity is defined by Information Theory in terms of information rate, per unit time. For τ large enough, the posterior precision $\phi_{(\tau-1)\Delta}$ of the target is arbitrarily close to its steady state $\Phi(\alpha\Delta|\gamma, \omega)$. So for a given sampling policy, by continuity of the information measure $I_{\tau\Delta}$ in $\phi_{(\tau-1)\Delta}$, the average information transmitted per unit time by the last observation converges.

Proposition 2 (The asymptotic information rate). *As the number of observations grows unbounded, the information transmitted by each observation (the information rate in between observations) converges to*

$$\begin{aligned} \lim_{\tau \rightarrow \infty} \frac{I_{\tau\Delta}}{\Delta} &= \frac{1}{2\Delta} \log\left(\frac{\Phi(\alpha\Delta|\gamma, \omega)}{\Phi(\alpha\Delta|\gamma, \omega) - \gamma}\right) \\ &= \frac{1}{2\Delta} \log\left(1 + \frac{\gamma}{\omega}(1 - e^{-2\alpha\Delta}) + \frac{e^{-2\alpha\Delta}\gamma/\omega}{\frac{1+\gamma/\omega}{2} + \sqrt{\left(\frac{1+\gamma/\omega}{2}\right)^2 + \frac{\gamma/\omega}{e^{2\alpha\Delta}-1}}}\right) \\ &\equiv \hat{I}(\Delta|\alpha, \gamma/\omega). \end{aligned}$$

The asymptotic information rate $\hat{I}(\Delta|\alpha, \gamma/\omega)$ is decreasing in Δ , with $\hat{I}(0|\alpha, \gamma/\omega) = \infty$ and $\hat{I}(\infty|\alpha, \gamma/\omega) = 0$. For each $\Delta > 0$, $\hat{I}(\Delta|\alpha, \gamma/\omega)$ is decreasing in ω given α (increasing in $\sigma^2 = 2\alpha/\omega$ given α), decreasing in α given $\sigma^2 = 2\alpha/\omega$, and increasing in $\sigma^2 = 2\alpha/\omega$ given ω (in α and σ keeping $\omega = 2\alpha/\sigma^2$ constant).

The lower panel of Fig. 2 illustrates the shape of the asymptotic information rate $\hat{I}(\Delta|\alpha, \gamma/\omega)$ as a function of the sampling interval Δ for our usual parameterization.

4. The optimal sampling policy

In order to complete the description of the optimal sampling problem, we need to specify the constraint on information acquisition/processing faced by the DM. Continuous sampling is the special case of periodic sampling ($\Delta = 0$) which requires an

infinite information rate, which is physically impossible. Thus, we restrict attention to discrete sampling policies, characterized by a sequence of positive intervals $\{\Delta_\tau\}$. Samples $Y_t = X_t + Z_t$ are observed at times $t = \Delta_1, \Delta_1 + \Delta_2, \dots$ and so forth. In the previous sections we showed that the average expected loss $L(\alpha\Delta|\gamma, \omega)$ and the rate of information $\hat{I}(\Delta|\alpha, \gamma/\omega)$ are known in closed-form when the sampling interval Δ is fixed. From this point on, we restrict attention to periodic sampling policies with fixed $\Delta > 0$. Beyond its simplicity, this case has obvious independent interest. But, in general, periodic sampling at fixed Δ might not be optimal.

It is easy to rule out any variable sampling policy $\{\Delta_\tau\}$ which has a limit Δ_∞ . By similar arguments to those illustrated earlier, if $\lim \Delta_\tau = \Delta_\infty$ then the posterior precision ϕ_t converges to $\Phi(\alpha\Delta_\infty|\gamma/\omega)$ (at each sampling time), so the average prediction loss L would be the same as if one always sampled from the beginning at the same fixed frequency Δ_∞ . If the cost of sampling is weakly convex in Δ , then variable but converging sampling intervals would raise the average cost without reducing the prediction loss L . Hence, such a converging staggering policy $\{\Delta_\tau\}$ cannot be optimal. A similar argument, however, does not apply to non-converging sampling policies, which in fact might be optimal. We leave the characterization of the optimal (possibly aperiodic) sampling policy for future research.³

We impose two alternative types of constraints on the choice of a periodic sampling policy Δ . First, we assume that the information rate $\hat{I}(\Delta|\alpha, \gamma/\omega)$ has a cost. Second, following a fairly standard approach in applied economic models, we specify an ad hoc cost function on the signal precision per unit time. We then show that optimal inertia has quite different comparative statics properties in the two models.

4.1. Optimal sampling with costly information rate

We treat the information rate as a costly input into the prediction of the process $\langle X \rangle$. We will discuss shortly the possible interpretations of this cost. Since the average loss depends only on ‘tail’ events as the sample number τ grows unbounded, and the information rate converges to $\hat{I} = \hat{I}(\Delta|\alpha, \gamma/\omega)$, we state the information rate constraint in terms of \hat{I} . The average cost per unit time of the information rate \hat{I} is given by an increasing and weakly convex function $c(\hat{I})$, with $c(0) = 0$. The DM chooses to sample at the time interval that minimizes the sum of the average prediction loss and

³ One would expect a periodic sampling policy $\Delta > 0$ to dominate a variable non-converging sampling policy in a stationary environment, namely when the initial condition ϕ_0 equals the fixed point $\Phi(\alpha\Delta|\gamma, \omega)$ for posterior precision. But ϕ_0 is a datum of the problem, while $\Phi(\alpha\Delta|\gamma, \omega)$ depends on the endogenous sampling interval Δ , so the optimal periodic policy starting from $\phi_0 = \Phi(\alpha\Delta|\gamma, \omega)$ is generically different than Δ . Alternatively, one may imagine time running from $-\infty$ to $+\infty$, the objective being to minimize the average two-sided loss. In this case one might expect at any finite time the posterior precision to be in steady state $\Phi(\alpha\Delta|\gamma, \omega)$, for any initial conditions $\phi_{-\infty}$. However, convergence of the precision to a stable fixed point depends on the nature of the sampling policy $\{\Delta_\tau\}$ itself. Furthermore, the average prediction loss is concave in the sampling interval Δ for a fixed Δ policy (Proposition 1). If this concavity is stronger than the convexity of the costs in Δ , then by Jensen’s inequality introducing small periodic variability in Δ might reduce the average loss. So a variable non-converging sampling policy may be optimal even in a stationary environment.

of the cost of the information rate:

$$\Delta^* = \arg \min_{\Delta \in \bar{\mathbb{R}}_+} \{L(\alpha\Delta|\gamma, \omega) + c(\hat{I}(\Delta|\alpha, \gamma/\omega))\} \quad (4.1)$$

where $\bar{\mathbb{R}}_+$ is the extended positive real line, a closed set. Here $\Delta^* = \infty$ means that the DM optimally chooses to never sample and to base his predictions only on prior information, counting on the stationarity of the process. In this case the information rate acquired and its cost are zero, and the average loss is $1/\omega$. Conversely, $\Delta^* = 0$ means continuous sampling.

The continuity of the objective function ensures the existence of a minimizer in $\bar{\mathbb{R}}_+$. By Proposition 1, the prediction loss L is concave in Δ . Therefore, a first-order condition is in general not sufficient for problem (4.1). By Proposition 2 the information rate \hat{I} , and thus its cost and the whole objective function in (4.1) explode as $\Delta \rightarrow 0$. Therefore, the corner solution $\Delta^* = 0$ is not feasible.

Proposition 3 (Optimal periodic sampling with costly information). *In the Costly Information (CI) model (4.1), the optimal sampling interval Δ^* always exists and is strictly positive, possibly infinite.*

The cost function $c(\cdot)$ has several possible interpretations. In the first, ‘market’ interpretation, the DM can choose among a range of different media, securing the arrival of information at different frequencies and thus providing different amounts of information. For example, he can subscribe to a monthly newsletter, to a broadband line to navigate the web, to a cable TV station to watch the daily wrap up of the markets, and so forth. The optimal frequency of information arrival, $1/\Delta^*$, is chosen in advance by subscribing to the corresponding medium. Indeed, economic data accrue to households mostly through channels that transmit at fixed frequencies. The acquisition of an information rate is treated as an upfront investment, a choice of technology. The information rate is best thought of as a ‘funnel’, a resource that cannot be reallocated over time but only across tasks, by opening and closing the funnel capacity. The special case of costless but bounded information rate

$$\min_{\Delta \in \bar{\mathbb{R}}_+} L(\alpha\Delta|\gamma, \omega) \text{ s.t. } \hat{I}(\Delta|\alpha, \gamma/\omega) \leq I_{\max}$$

obtains when the cost function $c(\hat{I})$ is flat (zero marginal cost) up to some maximum rate I_{\max} , and then becomes vertical (infinite marginal cost).

In the second, ‘opportunity cost’ interpretation, the DM must allocate his limited information processing capacity to various tasks. For example, the firm must predict demand and formulate a price, but also predict and hedge shocks to energy costs. Implicit is a cost of time, but the efficiency units of time are measured in bits/second.

The main point of this paper is that the cost-of-information function c is *not* just an ad hoc cost function defined only on sampling frequencies. An exogenous change in the parameters of the target dynamics (α and σ) affects both the benefits *and* the cost of a given sampling frequency Δ , by changing the information transmitted per unit time $\hat{I}(\Delta|\alpha, \gamma/\omega)$ for a given Δ . For example, reading the Wall Street Journal every day in recent times of stock market turbulence is more time- and capacity-consuming

because the quantity of information transmitted is higher for the given daily frequency, and less capacity is left for reading novels or thinking about dinner. By contrast, a ‘pure’ sampling cost function is unaffected by these changes in the target parameters. We address this case next.

4.2. Optimal sampling with costly samples

As a benchmark for comparison, consider a cost function defined on the sampling frequency and/or the precision γ of each observation $Y_{\tau\Delta}$. This is the intuitive, albeit ad hoc avenue followed in most economic models of information acquisition (e.g., Moscarini and Smith, 2001, 2002). Sampling n times in a T -time interval (once every $\Delta = T/n$ periods) yields a string of observations $Y_{\Delta}, Y_{2\Delta}, \dots, Y_{n\Delta} = Y_T$. A sufficient statistic for our prediction problem is the sample mean $n^{-1} \sum_{\tau=1}^n Y_{\tau\Delta}$, which is a normal with mean $n^{-1} \sum_{\tau=1}^n X_{\tau\Delta}$ and variance $1/(n\gamma) = \Delta/(\gamma T)$.

Therefore, it is natural to define a cost function on the number of equivalent observations per unit time $n\gamma/T$, namely on the ratio γ/Δ . We assume a cost $\kappa(\gamma/\Delta)$, increasing and convex. Confronted with such a cost, the DM chooses to sample at time interval:

$$\Delta^{**} = \arg \min_{\Delta \in \mathbb{R}_+} \left\{ L(\alpha\Delta | \gamma, \omega) + \kappa\left(\frac{\gamma}{\Delta}\right) \right\}. \tag{4.2}$$

Once again, an unbounded cost function implies that the total loss is unbounded as $\Delta \rightarrow 0$, so continuous sampling is suboptimal and $\Delta^{**} > 0$. As before, it may still be optimal not to sample at all ($\Delta^{**} = \infty$).

Proposition 4 (Optimal periodic sampling with costly sampling). *In the Costly Sampling (CS) model (4.2), the optimal sampling interval Δ^{**} always exists and is strictly positive, possibly infinite.*

5. Comparative statics: predictability and optimal inertia

5.1. The moderating effects of information-rate costs

We evaluate the two competing theories of time-dependent rules laid out in the previous section—the Costly Information model (4.1) (from now on, CI) and the Costly Sampling model (4.2) (from now on, CS)—on the basis of their comparative statics predictions. Specifically, we illustrate the response of the optimal sampling frequencies Δ^* , Δ^{**} in the two models to changes in the *persistence* α and the *volatility* σ of target innovations. The predictability of the target process $\langle X \rangle$ measures in some sense the complexity of the tracking task and its informational demands. We distinguish three cases:

1. When σ rises, for a given α , namely when ω declines for a given α , the target becomes less predictable both in the short and in the long run: both $\mathbb{V}[X_t | \mathcal{F}_0^{(A)}] = h_t^{-1}$ and $\mathbb{V}[X_\infty | \mathcal{F}_0^{(A)}] = 1/\omega = \sigma^2/2\alpha$ rise.

2. When α and $\omega = 2\alpha/\sigma^2$ decline for a given σ , again the target becomes less predictable both in the short and in the long run: both $\mathbb{V}[X_t|\mathcal{F}_0^{(\Delta)}] = h_t^{-1}$ and $\mathbb{V}[X_\infty|\mathcal{F}_0^{(\Delta)}] = 1/\omega = \sigma^2/2\alpha$ rise. Now, $1/h_t$ rises due to both α and ω .
3. Finally, when α and σ rise together, leaving the asymptotic variance $1/\omega$ unchanged, the target becomes less predictable in the short run, as $\mathbb{V}[X_t|\mathcal{F}_0^{(\Delta)}] = 1/h_t$ rises, while it is equally predictable in the long-run by construction. That is, when the persistence and volatility of innovations change so as to maintain the long-run behavior of the target, the effect of volatility dominates in the short run.

The main qualitative difference between the CI model introduced in this paper, (4.1), and the standard CS model, (4.2), is easily identified. The parameters determining the predictability of the target, α and σ (hence ω), affect the prediction loss L and produce the same *marginal benefit effect* in both cases. But these parameters also affect the cost of the information rate—and hence produce a *marginal cost effect*—only in the CI model. Namely, in the standard CS model there is a fundamental disconnection between the difficulty of the decision problem and the cost of observations, which only depend on the quality of the observations, not on those of the information they are about. The CI approach, instead, explicitly recognizes that a given observation has a different informational content, depending not only on the use that one makes of the information (on the loss function), but also on *prior uncertainty* and on the *stochastic properties of the environment*.

Consequently, the two models may have quite different comparative statics predictions. In particular, it is natural to expect sampling in the CS model to be more frequent the less predictable the target, because the marginal benefit of sampling a less predictable target at the same frequency Δ is higher, and the marginal cost is the same. Conversely, sampling a less predictable target in the CI model might be prohibitively expensive, so the higher marginal cost might outweigh the higher marginal benefit. The DM might be ‘paralyzed by complexity’, and the optimal sampling frequency $1/\Delta^*$ may be non-monotone or even globally increasing in the predictability of the target. More generally, *we expect the optimal CI sampling interval Δ^* to be much less sensitive to parametric shifts than the optimal CS sampling interval Δ^{**} , due to the moderating cost-effect.*

5.2. Quantitative effects

In the CS model, the dependence of the optimal sampling interval Δ^{**} on parameters can be investigated and, to some extent, signed analytically. The signing part is not possible in the CI model for Δ^* , due to algebraic complexity, although the relevant expressions are all available in closed form. To contrast the two approaches most clearly, we use numerical examples to illustrate the impact of the *same* changes in α and σ on Δ^* and Δ^{**} , starting from the *same* baseline parametrization.

We first transform α into the serial correlation parameter $\rho = e^{-\alpha}$ corresponding to a unit time interval. This innocuous monotone decreasing transformation makes it easier to both visualize the effects of persistence of innovations, as it bounds the range of ρ to $(0, 1)$, and to interpret them, as ρ corresponds to the persistence parameter

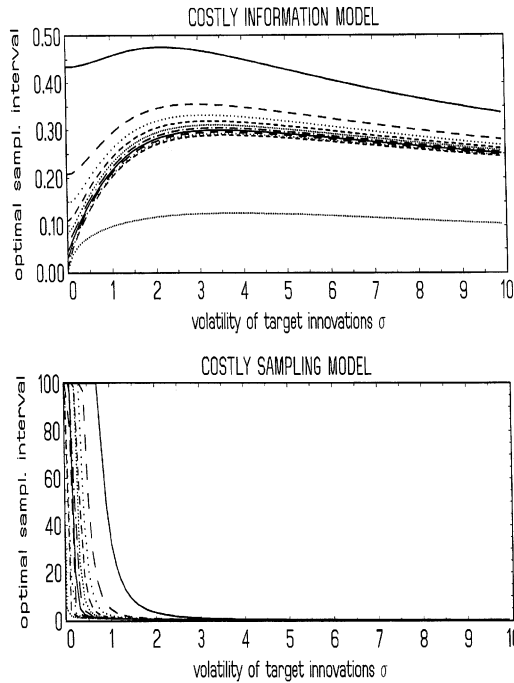


Fig. 4. Optimal sampling interval as a function of the volatility σ of target innovations, given their persistence $\rho = e^{-\alpha}$. Each curve corresponds to a different given level of ρ .

most familiar to economists. In this reparameterization, the prediction problem is more demanding, and the environment less predictable: the higher is σ^2 for a given ρ , the lower is ρ for a given σ , and the higher is ρ for a given $\omega = -\log \rho/\sigma^2$.

The cost function in the baseline parametrization is quadratic $c(z) = \kappa(z)/\kappa_0 = z^2/2$ for both models. We normalize the error precision to $\gamma = 1$, which is equivalent to choosing the units of both processes and the scale of the loss function. Given the obtained values for Δ^* , in each exercise we rescale the cost function in the CS model (through the parameter κ_0) so that Δ^* and Δ^{**} are equal at the mid-point of the parameter range considered: $\rho = 0.5$, $\sigma = 5$. This rescaling makes comparisons between the scales of Δ^* and Δ^{**} visually easier. The effects of these three parametric shifts are illustrated in Figs. 3–7 respectively. The effects of changes in a parameter is illustrated for a wide range of given values of the other parameter held fixed in each exercise.

5.2.1. Changes in the volatility of target innovations

Fig. 4 reports the optimal sampling intervals in the two models as a function of the volatility of target innovations σ for given persistence $\rho = e^{-\alpha}$. The larger σ , the smaller is ω , which imply a more informationally demanding prediction problem. Each curve corresponds to a given value of ρ from 0.0001, 0.01, 0.02... to 0.98, 0.99, 0.9999. The larger the ρ , and the smaller α is, the more persistent are target innovations, and

the lower are the curves. The comparison between the two models reveals two clear differences, which are both consistent with our general qualitative conjectures.

First, the optimal sampling interval Δ^{**} in the CS model (4.2) (lower panel) is monotone as expected: the higher is σ , the lower is Δ^{**} , and more frequently is it optimal to sample. By contrast, the optimal sampling interval Δ^* in the CI model (4.1) (upper panel) is non-monotone and hump-shaped. The intuition is simple, building on Propositions 1 and 2. As target volatility σ rises, the marginal benefit of tracking the target always rises. By Proposition 1, the loss function rises uniformly with σ , and therefore its slope rises in the sampling interval Δ . A more complex prediction problem is addressed more aggressively, as the marginal benefit of a sample is high when the target is hard to predict without observations. Thus, Δ^{**} always declines in volatility σ . But Δ^* increases from zero with σ up to a maximum, to then decline. A ‘simple’ prediction problem is not informationally demanding and can be sampled very frequently. By Proposition 2, as σ rises, the information rate of a given observation frequency Δ rises with it, and so does its marginal cost. When this cost-effect dominates the benefit-effect, optimal inertia Δ^* rises with σ . As σ explodes, the latter effect eventually dominates, and inertia declines to zero once again.

The second clear difference is that, due to the same moderating cost-effect, inertia Δ^{**} in the CS model is much more sensitive to parametric changes than Δ^* in the CI model. The intuition is the same: the cost-effect always works against the benefit-effect, thus its addition can only be moderating.

The second result has important implications for economic applications. Analytical convenience has made the Calvo-type adjustment rule quite popular to generate nominal rigidities in monopolistically competitive macroeconomic models. However, the Lucas critique (among others) applies to this practice, since the frequency of price adjustment is invariant to changes in the stochastic process describing the environment, for example to policy changes. The standard answer is that this invariance is a good approximation when the environment is relatively stable. Our results indicate that the Lucas critique is quite appropriate even for small parameter changes if infrequent adjustment à la Calvo (1983) is motivated by a CS model. The same critique, however, has much less bite in a Costly Information approach, which provides some firm grounds for the standard rebuttal.

5.2.2. Changes in the persistence of target innovations

Fig. 5 reports the results of varying the persistence of target innovations $\rho = e^{-\alpha}$ for a given volatility σ . The higher is ρ , the harder is the prediction problem. Now both α and $\omega = 2\alpha/\sigma^2$ change together. The curves correspond to fixed values of $\sigma = 1, 2, \dots, 10$. Inertia decreases in the persistence of target innovations in both models: a more persistent target is harder to track both in the short and in the long run, and the standard marginal benefit effect prevails. However, once again inertia is widely more sensitive to parameter changes in the CS model (lower panel). Recall that the sampling cost function κ is rescaled so that $\Delta^* = \Delta^{**} = 0.3$ for $\sigma = 5$ and $\rho = 0.5$, so the scale of the curves is the same in the middle of both graphs.

However, the results for the CI model (4.1) change if we make the marginal cost vanish faster at zero. Fig. 6 reports the results with $\kappa(z) = c(z)/\kappa_0 = z^3$. Now, optimal

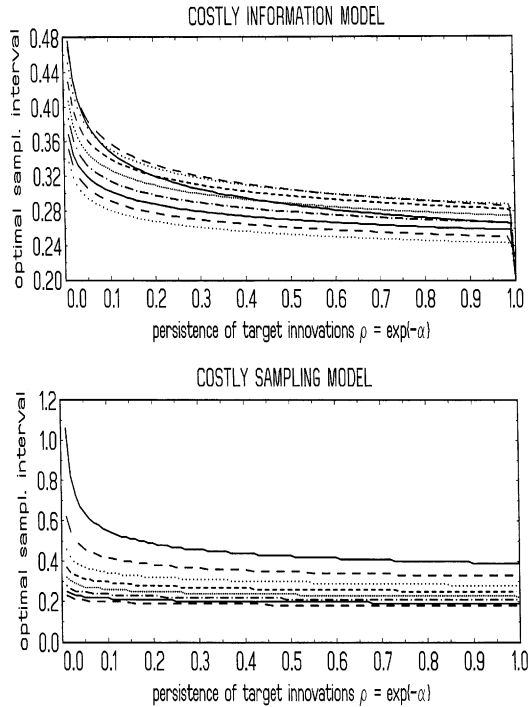


Fig. 5. Optimal sampling interval as a function of the persistence $\rho = e^{-\alpha}$ of target innovations, given their volatility σ . Quadratic information costs. Each curve corresponds to a different given level of σ .

inertia Δ^* is non-monotone in and very inelastic to the persistence of innovations in the CI model, while inertia Δ^{**} in the CS model (4.2) is still quickly decreasing. Indeed, Δ^{**} is decreasing in ρ in any parameterization we tried, as one would expect. The intuition should be familiar by now.

5.2.3. Changes in short-run volatility of the target

Finally, Fig. 7 reports the response of the optimal sampling intervals as we vary both the persistence $\rho = e^{-\alpha}$ and the volatility of target innovations σ so as to maintain the asymptotic variance of the process $1/\omega = \sigma^2/2\alpha$ constant. The different curves correspond to values of ω ranging from 0.01 to 5. The higher is ω , the higher are the curves in both panels of Fig. 7. In this case, both models predict that as σ rises (and α with it) and the target becomes less predictable in the short run, while remaining equally predictable in the long run, the optimal degree of inertia rises. The marginal gain from tracking a target that is more variable in the short-run but not in the long run is smaller. This result appears robust to the many parameterizations that we experimented with.

This result seems to run counter to the previous intuition. Indeed, by Proposition 1, when σ and α are proportionally higher, the prediction loss is closer to its fixed

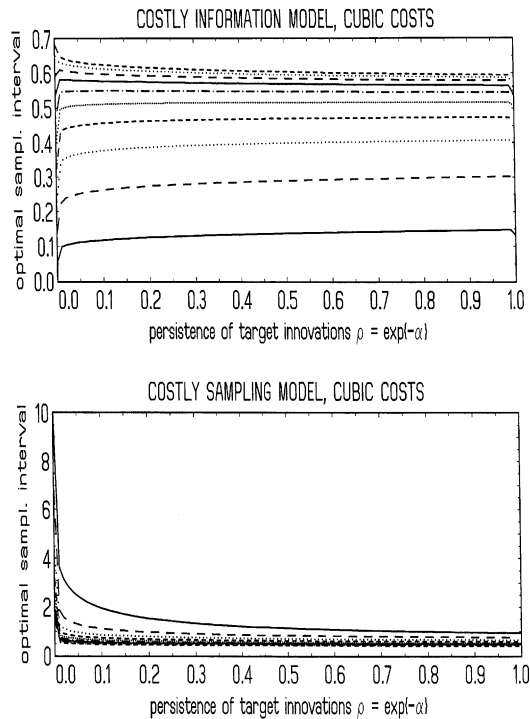


Fig. 6. Optimal sampling interval as a function of the persistence $\rho = e^{-\alpha}$ of target innovations, given their volatility σ . Cubic information costs. Each curve corresponds to a different given level of σ .

upper bound $1/\omega$. Therefore, it is relatively flat at a level near $1/\omega$ and inelastic in the sampling frequency Δ (see Fig. 3). Sampling more or less frequently makes little difference when the target is unpredictable at high frequencies, but relatively under control at low frequencies. It appears that the effects of stronger mean-reversion dominate in the short run. Once again, the moderating cost-effect in the CI model makes inertia much less sensitive to this parametric shift (compare the scales of the two panels).

We summarize our main findings as follows.

- In the CS model (4.2), a target which is less predictable both in the short and in the long run is sampled relatively more frequently. With Costly Information (4.1), instead, inertia Δ^* may be non-monotone or even decreasing in the predictability of the target.
- The optimal degree of inertia is much more sensitive to changes in the environment with Costly Sampling than with Costly Information: the effects of parametric changes on the level and on the cost of the information contained in each observation moderates the effects on the marginal benefit of adjustment.

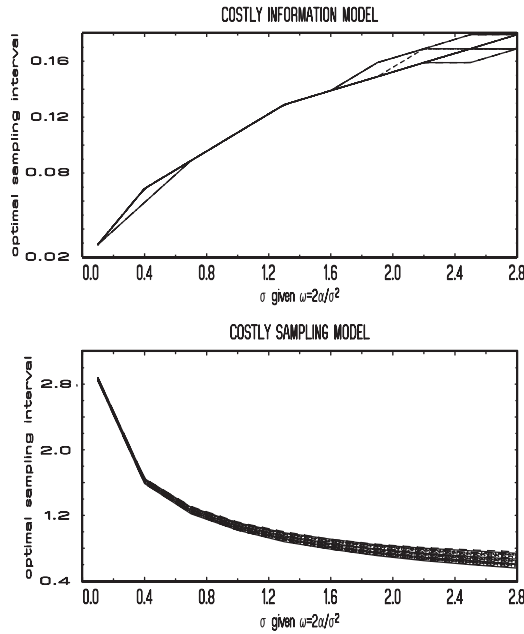


Fig. 7. The optimal sampling interval as a function of the volatility σ and the persistence $\rho = e^{-\alpha}$ of target innovations, keeping the asymptotic variance of the target $\omega = 2\alpha/\sigma^2$ constant. Each curve corresponds to a different given level of ω .

5.3. One more benchmark: state-dependent adjustment

As a further benchmark of comparison for the CI model (4.1), popular in the economics literature, consider a problem in which information accrues freely, continuously, and without noise ($\Delta = 0, \gamma = \infty$) to the DM, but there is a fixed cost to change the prediction a_t . As well known, such adjustment ‘menu’ costs give rise to optimal state-dependent rules: the DM spends the fixed cost to update his action whenever the target has drifted too far given the current action. Inertia has stochastic frequency. Optimal (S, s) rules have been derived in a variety of contexts, but always with non-stationary (standard or geometric Brownian Motion) target processes. The case of a mean-reverting target has not been solved to date. Its complication arises from the fact that the distribution of target increments depends on the current level. It seems plausible, just like in the models solved so far, that a higher volatility of target innovations (σ) would raise the option value of waiting and the resulting optimal inertia. However, it seems equally plausible that a stronger mean-reversion of target innovations (higher α) would have the same effect and increase inertia, as deviations of the target from its long-run value would disappear or be reversed with higher probability, reducing the need for a costly frequent adjustment. This last prediction is similar to that of the traditional CS model, and different from the one of the CI model (4.1) introduced in this paper.

6. Concluding remarks

Over more than 50 years, since Shannon's (1948) seminal article, Information Theory has uncovered general restrictions on the technology for information transmission, acquisition, and processing. As the production, communication, and trade of information services represent an ever-increasing share of post-industrial economies, these insights bear increasing relevance to economic analysis, where such information-related activities and technologies are typically modelled in very ad hoc ways. This paper contains the very first attempt to embed a basic idea from Information Theory into an economically interesting decision problem. Compared to ad hoc adjustment rules proposed in macroeconomic models, both of the time-dependent and state-dependent breeds, this model introduces yet another type of behavior. Adjustment occurs continuously, but almost always according to informationally simple and pre-determined time paths, that get drastically revised infrequently with the incorporation of new information. The comparative statics properties of this adjustment policy are also new. For example, a sudden and unforeseen decline in the predictability of environment may make economic decisions so costly in terms of the required information resources as to 'scare' agents away from these decisions.

In the light of the model illustrated in this paper, we can think of the advent of the Internet as a sudden increase in the information capacity available to economic agents, which should be expected to reduce stickiness in the economy. Investors, rather than being paralyzed by complexity, may now cope fruitfully with the massive amount of data and news that flows to them daily. While the pace of the diffusion of information has certainly risen, the effective amount of additional valuable information transmitted, as measured for example in bits per second, remains to be verified.

Acknowledgements

I gratefully acknowledge conversations with Chris Sims at the onset of this work, comments by William Brainard and the Editor, and the many insightful suggestions of an anonymous referee, which substantially improved the substance and the exposition of this article. The usual disclaimer applies. I also gratefully acknowledge the financial support of the Alfred P. Sloan Foundation and Yale University.

Appendix A

Proof of Lemma 1. The unique fixed point of the recursion (2.6) solves

$$\Phi = \frac{\Phi\omega}{\Phi(1 - e^{-2\alpha\Delta}) + e^{-2\alpha\Delta}\omega} + \gamma, \quad (\text{A.1})$$

$$\Phi^2 = \Phi(\omega + \gamma) + \gamma\omega \frac{e^{-2\alpha\Delta}}{1 - e^{-2\alpha\Delta}},$$

$$\begin{aligned} \Phi &= \Phi(\alpha\Delta|\gamma, \omega) \\ &= \frac{\omega + \gamma + \sqrt{(\omega + \gamma)^2 + 4\omega\gamma \frac{e^{-2\alpha\Delta}}{1-e^{-2\alpha\Delta}}}}{2} \\ &= \frac{\omega + \gamma}{2} + \sqrt{\left(\frac{\omega + \gamma}{2}\right)^2 + \omega\gamma \frac{e^{-2\alpha\Delta}}{1 - e^{-2\alpha\Delta}}} \\ &= \frac{\omega + \gamma}{2} + \sqrt{\left(\frac{\omega + \gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}} \end{aligned}$$

implying $\Phi(0 + |\gamma, \omega) = \infty$, $\Phi(\infty|\gamma, \omega) = \omega + \gamma$.

Next,

$$\begin{aligned} \Phi'(\alpha\Delta|\gamma, \omega) &= -\frac{1}{\sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta}-1}}} \frac{e^{2\alpha\Delta}\omega\gamma}{(e^{2\alpha\Delta} - 1)^2} \\ &= -\frac{e^{2\alpha\Delta}\omega\gamma}{\sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 (e^{2\alpha\Delta} - 1)^4 + \omega\gamma(e^{2\alpha\Delta} - 1)^3}} < 0, \end{aligned}$$

so

$$\begin{aligned} \Phi'(0 + |\alpha, \gamma, \omega) &= -\frac{\omega\gamma}{\sqrt{0}} = -\infty, \\ \Phi'(\infty|\alpha, \gamma, \omega) &= -\frac{1}{\sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + 0}} = 0. \end{aligned}$$

Finally,

$$\begin{aligned} &\frac{\Phi''(\alpha\Delta|\gamma, \omega)}{\Phi'(\alpha\Delta|\gamma, \omega)} \\ &= \frac{d \log [-\Phi'(\alpha\Delta|\gamma, \omega)]}{d(\alpha\Delta)} \\ &= \frac{d}{d(\alpha\Delta)} \log \left(\frac{1}{\sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta}-1}}} \frac{e^{2\alpha\Delta}\omega\gamma}{(e^{2\alpha\Delta} - 1)^2} \right) \\ &= 2 - 2 \frac{2e^{2\alpha\Delta}}{e^{2\alpha\Delta} - 1} - \frac{1}{2} \frac{1}{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta}-1}} \frac{-2e^{2\alpha\Delta}\omega\gamma}{(e^{2\alpha\Delta} - 1)^2} \\ &= -2 \frac{e^{2\alpha\Delta} + 1}{e^{2\alpha\Delta} - 1} + \frac{e^{2\alpha\Delta}\omega\gamma}{\left(\frac{\omega+\gamma}{2}\right)^2 (e^{2\alpha\Delta} - 1)^2 + \omega\gamma(e^{2\alpha\Delta} - 1)} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{e^{2\alpha\Delta} - 1} \left[-2(e^{2\alpha\Delta} + 1) + \frac{e^{2\alpha\Delta} \omega\gamma}{\left(\frac{\omega+\gamma}{2}\right)^2 (e^{2\alpha\Delta} - 1) + \omega\gamma} \right] \\
 &= \frac{1}{e^{2\alpha\Delta} - 1} \frac{-2(e^{2\alpha\Delta} + 1) \left(\frac{\omega+\gamma}{2}\right)^2 (e^{2\alpha\Delta} - 1) - 2(e^{2\alpha\Delta} + 1)\omega\gamma + e^{2\alpha\Delta} \omega\gamma}{\left(\frac{\omega+\gamma}{2}\right)^2 (e^{2\alpha\Delta} - 1) + \omega\gamma} \\
 &= \frac{1}{e^{2\alpha\Delta} - 1} \frac{-2(e^{2\alpha\Delta} + 1) \left(\frac{\omega+\gamma}{2}\right)^2 (e^{2\alpha\Delta} - 1) - (e^{2\alpha\Delta} + 2)\omega\gamma}{\left(\frac{\omega+\gamma}{2}\right)^2 (e^{2\alpha\Delta} - 1) + \omega\gamma} \\
 &< 0,
 \end{aligned}$$

implying

$$\Phi''(\alpha\Delta|\gamma, \omega) > 0.$$

To investigate the stability of the difference equation (2.6), using the steady-state equation (A.1) for $\Phi(\alpha\Delta|\gamma, \omega)$ (as a shorthand, Φ):

$$\begin{aligned}
 \phi_{\tau\Delta} - \Phi &= \frac{\omega\phi_{(\tau-1)\Delta}}{\phi_{(\tau-1)\Delta}(1 - e^{-2\alpha\Delta}) + \omega e^{-2\alpha\Delta}} - \frac{\omega\Phi}{\Phi(1 - e^{-2\alpha\Delta}) + \omega e^{-2\alpha\Delta}} \\
 &= \Gamma(\phi_\tau)[\phi_{(\tau-1)\Delta} - \Phi],
 \end{aligned}$$

where

$$\Gamma(\phi_\tau) = \frac{\omega^2 e^{-2\alpha\Delta}}{[\phi_{(\tau-1)\Delta}(1 - e^{-2\alpha\Delta}) + \omega e^{-2\alpha\Delta}][\Phi(1 - e^{-2\alpha\Delta}) + \omega e^{-2\alpha\Delta}]} > 0.$$

Clearly

$$\begin{aligned}
 \Gamma(\phi_\tau) &\leq \frac{\omega^2 e^{-2\alpha\Delta}}{\omega e^{-2\alpha\Delta}[\Phi(1 - e^{-2\alpha\Delta}) + \omega e^{-2\alpha\Delta}]} \\
 &= \frac{\omega/\Phi}{[1 - (1 - \omega/\Phi)e^{-2\alpha\Delta}]} \\
 &< 1
 \end{aligned}$$

because we showed $\Phi = \Phi(\alpha\Delta|\gamma, \omega) > \Phi(\infty|\alpha, \gamma, \omega) = \omega + \gamma > \omega$. Also, $\Gamma(\Phi) > 0$ and Γ is continuous. Therefore, $\Phi(\alpha\Delta|\gamma, \omega)$ is a unique global attractor for the recursion.

The last result follows directly from the optimal prediction formulae. \square

Proof of Proposition 1. At any time t the flow loss is $\mathbb{E}[(\hat{X}_t - X_t)^2 | \mathcal{F}_t^{(A)}] = \mathbb{V}[X_t | \mathcal{F}_t^{(A)}] \equiv \phi_t^{-1}$, where ϕ_t^{-1} is derived from $\phi_{\tau\Delta}^{-1}$ and $\tau\Delta$ is the time of the last observation available at time t : $\tau\Delta \leq t < (\tau+1)\Delta$. The ex ante ‘flow’ loss, using the Law of Iterated Expectations and the fact that the $\langle \phi \rangle$ process is predictable, equals $\mathbb{V}[X_t | \mathcal{F}_0^{(A)}] = \mathbb{E}[\mathbb{V}[X_t | \mathcal{F}_t^{(A)}] | \mathcal{F}_0^{(A)}] = \mathbb{E}[\phi_t^{-1} | \mathcal{F}_0^{(A)}] = \phi_t^{-1}$. The total loss is the average integral of this magnitude.

When sampling is infrequent ($\Delta > 0$), at each observation and updating times $t = \tau\Delta = \Delta, 2\Delta \dots$ the DM computes a new pair of conditional moments $\hat{X}_{\tau\Delta}$, $\phi_{\tau\Delta}$. The

average loss in any interval between samples is therefore the average of the posterior variance over the period in between samples:

$$\begin{aligned} \frac{L_{(\tau+1)\Delta} - L_{\tau\Delta}}{\Delta} &= \frac{1}{\Delta} \int_{\tau\Delta}^{(\tau+1)\Delta} \mathbb{V}[X_{\tau\Delta+s} | \mathcal{F}_{\tau\Delta+s}^{(\Delta)}] ds \\ &= \frac{1}{\Delta} \int_{\tau\Delta}^{(\tau+1)\Delta} \left[\frac{e^{-2\alpha(s-\tau\Delta)}}{\phi_{\tau\Delta}} + \frac{1 - e^{-2\alpha(s-\tau\Delta)}}{\omega} \right] ds \\ &= \frac{1}{\omega} + \left(\frac{1}{\phi_{\tau\Delta}} - \frac{1}{\omega} \right) \frac{1 - e^{-2\alpha\Delta}}{2\alpha\Delta}. \end{aligned}$$

In this notation, the total average loss over the infinite time horizon is

$$L(\Delta) = \lim_{T \rightarrow \infty} \frac{L_T}{T}.$$

This limit, if it exists, is the same along any sequence of exploding times T . Fix any integer $k \geq 0$ and let $S = k\Delta$. Next, take the above limit for a sequence $T = (n + k)\Delta$ as n explodes. Since any initial loss over any finite initial period $[0, S]$ is irrelevant:

$$\begin{aligned} L &= \lim_{T \rightarrow \infty} \frac{L_T - L_S}{T - S} \\ &= \lim_{n \rightarrow \infty} \frac{\sum_{\tau=k}^{\tau=n+k} [L_{(\tau+1)\Delta} - L_{\tau\Delta}]}{n\Delta} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{\tau=n+k} \frac{L_{(\tau+1)\Delta} - L_{\tau\Delta}}{\Delta} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{\tau=n+k} \left\{ \frac{1}{\omega} + \left(\frac{1}{\phi_{\tau\Delta}} - \frac{1}{\omega} \right) \frac{1 - e^{-2\alpha\Delta}}{2\alpha\Delta} \right\} \\ &= \frac{1}{\omega} \left(1 - \frac{1 - e^{-2\alpha\Delta}}{2\alpha\Delta} \right) + \frac{1 - e^{-2\alpha\Delta}}{2\alpha\Delta} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{\tau=n+k} \frac{1}{\phi_{\tau\Delta}}. \end{aligned}$$

Since the limit is independent of k , for every $\varepsilon > 0$ we can take k large enough that $\phi_{\tau\Delta}$ is within ε of its global attractor Φ , so $1/\phi_{\tau\Delta}$ is within $\varepsilon/[\Phi(\Phi - \varepsilon)]$ of $1/\Phi$ and therefore $(1/n) \sum_{\tau=k}^{\tau=n+k} 1/\phi_{\tau\Delta}$ is within $\varepsilon/[\Phi(\Phi - \varepsilon)]$ of $1/\Phi$, i.e. arbitrarily close to $1/\Phi$. Therefore,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\tau=k}^{\tau=n+k} \frac{1}{\phi_{\tau\Delta}} = \frac{1}{\Phi},$$

and the average expected loss of the optimal prediction policy equals the asymptotic average loss per unit time:

$$L(\alpha\Delta | \gamma, \omega) = \frac{1}{\omega} - \frac{1 - e^{-2\alpha\Delta}}{2\alpha\Delta} \left(\frac{1}{\omega} - \frac{1}{\Phi(\alpha\Delta | \gamma, \omega)} \right)$$

which is (2.8).

Now, take any pair of smooth positive functions $\{f_i\}_{i=0,1}$ such that $f'_i(\cdot) < 0 < f''_i(\cdot)$ for $i=0, 1$. Let $f_2(\Delta) = -f_0(\Delta)f_1(\Delta)$. Then $f'_2(\Delta) = -f'_0(\Delta)f_1(\Delta) - f_0(\Delta)f'_1(\Delta) > 0$ and $f''_2(\Delta) = -f''_0(\Delta)f_1(\Delta) - 2f'_0(\Delta)f'_1(\Delta) - f_0(\Delta)f''_1(\Delta) < 0$. Thus, to prove that $L(\alpha\Delta|\gamma, \omega)$ is increasing and concave in Δ , it suffices to show that both $f_0(\Delta) = (2\alpha\Delta)^{-1}(1 - e^{-2\alpha\Delta})$ and $f_1(\Delta) = \omega^{-1} - \Phi^{-1}(\alpha\Delta|\gamma, \omega)$ are positive, decreasing and convex in $\alpha\Delta$.

For $(2\alpha\Delta)^{-1}(1 - e^{-2\alpha\Delta})$, clearly positive, one can verify that it is decreasing and convex by taking two derivatives and using properties of the exponential function. For $\omega^{-1} - \Phi^{-1}(\alpha\Delta|\gamma, \omega)$, we know that it is positive because $\omega < \Phi(\alpha\Delta|\gamma, \omega)$, and decreasing because so is $\Phi(\alpha\Delta|\gamma, \omega)$. For its convexity, we need to show

$$\frac{\Phi''(\alpha\Delta|\gamma, \omega)}{\Phi^2(\alpha\Delta|\gamma, \omega)} - 2 \frac{(\Phi'(\alpha\Delta|\gamma, \omega))^2}{\Phi^3(\alpha\Delta|\gamma, \omega)} > 0$$

multiplying through by $-\Phi^2(\alpha\Delta|\gamma, \omega)/\Phi'(\alpha\Delta|\gamma, \omega) > 0$:

$$-\frac{\Phi''(\alpha\Delta|\gamma, \omega)}{\Phi'(\alpha\Delta|\gamma, \omega)} > 2 \frac{-\Phi'(\alpha\Delta|\gamma, \omega)}{\Phi(\alpha\Delta|\gamma, \omega)}$$

using the expressions derived in an earlier proof

$$\begin{aligned} & \frac{1}{e^{2\alpha\Delta} - 1} \left[2(e^{2\alpha\Delta} + 1) + \frac{\omega\gamma}{\left(\frac{\omega+\gamma}{2}\right)^2 (e^{2\alpha\Delta} - 1) + \omega\gamma} \right] \\ & > \frac{2}{\frac{\omega+\gamma}{2} + \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}}} \frac{1}{\sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}}} \frac{e^{2\alpha\Delta}\omega\gamma}{(e^{2\alpha\Delta} - 1)^2} \\ & = \frac{1}{e^{2\alpha\Delta} - 1} \frac{2e^{2\alpha\Delta}}{\frac{\omega+\gamma}{2} \frac{e^{2\alpha\Delta} - 1}{\omega\gamma} + \frac{e^{2\alpha\Delta} - 1}{\omega\gamma} \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}}} \frac{1}{\sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}}} \\ & = \frac{1}{e^{2\alpha\Delta} - 1} \frac{2e^{2\alpha\Delta}}{\frac{\omega+\gamma}{2} \frac{e^{2\alpha\Delta} - 1}{\omega\gamma} \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}} + \frac{e^{2\alpha\Delta} - 1}{\omega\gamma} \left(\frac{\omega+\gamma}{2}\right)^2 + 1} \end{aligned}$$

simplifying $1/(e^{2\alpha\Delta} - 1)$ and rearranging

$$\begin{aligned} & 2(e^{2\alpha\Delta} + 1) + \frac{\omega\gamma}{\left(\frac{\omega+\gamma}{2}\right)^2 (e^{2\alpha\Delta} - 1) + \omega\gamma} \\ & > \frac{2e^{2\alpha\Delta}}{\frac{\omega+\gamma}{2} \frac{e^{2\alpha\Delta} - 1}{\omega\gamma} \left[\sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}} + \left(\frac{\omega+\gamma}{2}\right) \right]} + 1 \end{aligned}$$

This is always true because the LHS is larger than $2e^{2\alpha\Delta}$ and the RHS is smaller than $2e^{2\alpha\Delta}$. \square

Proof of Proposition 2.

$$\begin{aligned} \hat{I}(\Delta|\alpha, \gamma/\omega) &= \lim_{\tau \rightarrow \infty} \frac{I_{\tau\Delta}}{\Delta} \\ &= \frac{1}{2\Delta} \log \left(1 + \gamma \frac{\Phi(\alpha\Delta|\gamma, \omega)(1 - e^{-2\alpha\Delta}) + \omega e^{-2\alpha\Delta}}{\omega\Phi(\alpha\Delta|\gamma, \omega)} \right) \\ &= \frac{1}{2\Delta} \log \left(1 + \frac{\gamma}{\omega} (1 - e^{-2\alpha\Delta}) + \frac{\gamma e^{-2\alpha\Delta}}{\Phi(\alpha\Delta|\gamma, \omega)} \right) \\ &= \frac{1}{2\Delta} \log \left[1 + \frac{\gamma}{\omega} \left(1 - e^{-2\alpha\Delta} \left(1 - \frac{\omega}{\Phi(\alpha\Delta|\gamma, \omega)} \right) \right) \right] \\ &= \frac{1}{2\Delta} \log \left(1 + \frac{\gamma}{\omega} (1 - e^{-2\alpha\Delta}) + \frac{\gamma e^{-2\alpha\Delta}}{\frac{\omega+\gamma}{2} + \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta}-1}}} \right). \end{aligned}$$

We want to show that this rate is decreasing in Δ . Let

$$\begin{aligned} \hat{I}(\Delta|\alpha, \gamma/\omega) &= \frac{1}{2\Delta} \log \left(\frac{\Phi(\alpha\Delta|\gamma, \omega)}{\Phi(\alpha\Delta|\gamma, \omega) - \gamma} \right) \\ &= \frac{1}{2\Delta} Q(\Delta), \end{aligned}$$

where $Q(0) = 0$, $Q(\infty) = \log(1 + \omega/\gamma) > 0$,

$$\begin{aligned} Q'(\Delta) &= \frac{1}{2\Delta} \left[\frac{1}{\Phi(\alpha\Delta|\gamma, \omega)} - \frac{1}{\Phi(\alpha\Delta|\gamma, \omega) - \gamma} \right] \frac{d\Phi(\alpha\Delta|\gamma, \omega)}{d\Delta} \\ &= \frac{\gamma}{2\Delta\Phi(\alpha\Delta|\gamma, \omega)[\Phi(\alpha\Delta|\gamma, \omega) - \gamma]} \frac{\omega\gamma\alpha e^{2\alpha\Delta}}{(e^{2\alpha\Delta} - 1)^2 \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta}-1}}} \\ &= \frac{\alpha\omega\gamma^2 e^{2\alpha\Delta}}{(e^{2\alpha\Delta} - 1)^2 [\Phi(\alpha\Delta|\gamma, \omega) + \frac{\gamma}{e^{2\alpha\Delta}-1}] \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta}-1}}} > 0. \end{aligned}$$

So $\hat{I}(\Delta|\alpha, \gamma/\omega) = Q(\Delta)/\Delta$ is decreasing as claimed in Δ if $Q(\Delta)$ is concave, or $Q''(\Delta) < 0$. This inequality can be verified using the above expression for $Q'(\Delta)$ with lengthy and tedious (thus omitted) algebra.

The information rate $\hat{I}(\Delta|\alpha, \gamma/\omega)$ explodes as $\Delta \rightarrow 0$, because by Hopital

$$\begin{aligned} \lim_{\Delta \rightarrow 0} \hat{I}(\Delta|\alpha, \gamma/\omega) &= \frac{0}{0} \\ &= \lim_{\Delta \rightarrow 0} \hat{I}'(\Delta|\alpha, \gamma/\omega) \end{aligned}$$

$$\begin{aligned}
 &= \lim_{\Delta \rightarrow 0} \frac{\frac{\gamma}{\left(\frac{\omega+\gamma}{2} + \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2x\Delta}-1}}\right)^2} - \frac{2\alpha e^{2x\Delta}\omega\gamma}{(e^{2x\Delta}-1)^2}}{1 + \frac{\gamma}{\frac{\omega+\gamma}{2} + \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2x\Delta}-1}}} - \frac{2\alpha e^{2x\Delta}\omega\gamma}{(e^{2x\Delta}-1)^2}} \\
 &= \lim_{\Delta \rightarrow 0} \frac{\gamma}{1 + \frac{\gamma}{\frac{\omega+\gamma}{2} + \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2x\Delta}-1}}}} \\
 &\quad \times \frac{\frac{\alpha e^{2x\Delta}\omega\gamma}{(e^{2x\Delta}-1)^2}}{\left(\frac{\omega+\gamma}{2} + \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2x\Delta}-1}}\right)^2 \sqrt{\left(\frac{\omega+\gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2x\Delta}-1}}}
 \end{aligned}$$

ignoring all terms other than the exploding $\omega\gamma/(e^{2x\Delta} - 1)^2$

$$= \lim_{\Delta \rightarrow 0} 1 \frac{\frac{\alpha\omega\gamma}{(e^{2x\Delta}-1)^2}}{\left(\sqrt{\frac{\omega\gamma}{e^{2x\Delta}-1}}\right)^2 \sqrt{\frac{\omega\gamma}{e^{2x\Delta}-1}}} = \lim_{\Delta \rightarrow 0} \frac{\alpha}{\sqrt{\omega\gamma(e^{2x\Delta} - 1)}} = \infty.$$

Although continuous sampling leads to an infinite steady-state precision $\Phi(0+) = \infty$, which makes the informational content of each observation negligible, the information rate per unit time explodes as sampling becomes continuous. In this sense, continuous-time sampling and filtering is not a physically feasible operation.

At the other extreme:

$$\lim_{\Delta \rightarrow \infty} \hat{I}(\Delta|\alpha, \gamma/\omega) = \lim_{\Delta \rightarrow \infty} \frac{1}{2\Delta} \log\left(\frac{\omega + \gamma}{\omega}\right) = 0.$$

Next,

$$\begin{aligned}
 &\frac{\partial \hat{I}(\Delta|\alpha, \gamma/\omega)}{\partial \omega} \\
 &= \frac{-\frac{\gamma}{\omega^2} \left(1 - e^{-2x\Delta} \left(1 - \frac{\omega}{\Phi(\alpha\Delta|\gamma, \omega)}\right)\right) + \frac{\gamma}{\omega} (1 - e^{-2x\Delta}) \left(-\frac{1}{\Phi(\alpha\Delta|\gamma, \omega)} + \frac{\omega}{\Phi^2(\alpha\Delta|\gamma, \omega)} \frac{\partial \Phi(\alpha\Delta|\gamma, \omega)}{\partial \omega}\right)}{2\Delta \left[1 + \frac{\gamma}{\omega} \left(1 - e^{-2x\Delta} \left(1 - \frac{\omega}{\Phi(\alpha\Delta|\gamma, \omega)}\right)\right)\right]} \\
 &\propto -\frac{\gamma}{\omega^2} \left(1 - e^{-2x\Delta} \left(1 - \frac{\omega}{\Phi(\alpha\Delta|\gamma, \omega)}\right)\right) \\
 &\quad + \frac{\gamma}{\omega\Phi(\alpha\Delta|\gamma, \omega)} e^{-2x\Delta} \left(\frac{\partial \Phi(\alpha\Delta|\gamma, \omega)}{\partial \omega} \frac{\omega}{\Phi(\alpha\Delta|\gamma, \omega)} - 1\right) \\
 &< -\frac{\gamma}{\omega^2} e^{-2x\Delta} \frac{\omega}{\Phi(\alpha\Delta|\gamma, \omega)} + \frac{\gamma e^{-2x\Delta}}{\omega\Phi(\alpha\Delta|\gamma, \omega)} \left(\frac{\partial \Phi(\alpha\Delta|\gamma, \omega)}{\partial \omega} \frac{\omega}{\Phi(\alpha\Delta|\gamma, \omega)} - 1\right) \\
 &= \frac{\gamma e^{-2x\Delta}}{\omega\Phi(\alpha\Delta|\gamma, \omega)} \left[-1 + \frac{\partial \Phi(\alpha\Delta|\gamma, \omega)}{\partial \omega} \frac{\omega}{\Phi(\alpha\Delta|\gamma, \omega)} - 1\right].
 \end{aligned}$$

Thus this partial derivative is negative as claimed provided that

$$\omega \frac{\partial \Phi(\alpha \Delta | \gamma, \omega)}{\partial \omega} < 2\Phi(\alpha \Delta | \gamma, \omega).$$

Using the expression for $\Phi(\alpha \Delta | \gamma, \omega)$ from Lemma 1

$$\begin{aligned} \omega \left(\frac{1}{2} + \frac{\omega + \gamma + \frac{\gamma}{e^{2\alpha\Delta} - 1}}{2\sqrt{\left(\frac{\omega + \gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}}} \right) &< \omega + \gamma + 2\sqrt{\left(\frac{\omega + \gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}} \\ \omega \frac{\omega + \gamma + \frac{\gamma}{e^{2\alpha\Delta} - 1}}{2\sqrt{\left(\frac{\omega + \gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}}} &< \gamma + 2\sqrt{\left(\frac{\omega + \gamma}{2}\right)^2 + \frac{\omega\gamma}{e^{2\alpha\Delta} - 1}} \\ \omega \left(\omega + \gamma + \frac{\gamma}{e^{2\alpha\Delta} - 1} \right) &< 2\gamma\sqrt{\omega + \gamma + \frac{4\omega\gamma}{e^{2\alpha\Delta} - 1}} + (\omega + \gamma)^2 + \frac{4\omega\gamma}{e^{2\alpha\Delta} - 1} \\ -2\omega\gamma - \frac{3\omega\gamma}{e^{2\alpha\Delta} - 1} &< \frac{\gamma}{2}\sqrt{\omega + \gamma + \frac{4\omega\gamma}{e^{2\alpha\Delta} - 1}} \end{aligned}$$

which is always true.

Finally, the negative dependence of the information rate $\hat{I}(\Delta | \alpha, \gamma / \omega)$ on α given ω , and on α and ω given $\sigma^2 = 2\alpha/\omega$ follows directly from simple differentiation and from the previous result. \square

References

- Ash, R.B., 1965. *Information Theory*. Constable and Co., New York.
- Caballero, R., 1989. Time dependent rules, aggregate stickiness and information externalities. Columbia University Discussion Paper 428.
- Calvo, G., 1983. Staggered wages in a utility maximization framework. *Journal of Monetary Economics* 12 (3), 383–393.
- Clarida, Galy, J., Gertler, M., 1999. The science of monetary policy: a new Keynesian perspective. *Journal of Economic Literature* 37 (4), 1661–1707.
- Elliott, R.J., Aggoun, L., Moore, J.B., 1995. *Hidden Markov models; estimation and control*. Springer Verlag Applications of Mathematics, Vol. 29, Springer, Berlin.
- Gabaix, X., Laibson, D., 2001. The 6D bias and the equity premium puzzle. In: Bernanke, B., Rogoff, K. (Eds.), 2001 NBER Macroeconomics Annual. MIT Press, Cambridge, MA.
- Karlin, S., Taylor, H.M., 1981. *A Second Course in Stochastic Processes*. Academic Press, New York.
- Lucas, R.E., 1972. Expectations and the neutrality of money. *Journal of Economic Theory* 4, 103–124.
- Lucas, R.E., 1976. Econometric policy evaluation: a critique. In: Brunner, K., Meltzer, A. (Eds.), *Stabilization of the Domestic and International Economy*. North-Holland, Amsterdam.
- Mankiw, G., Reis, R., 2002. Sticky information versus sticky prices: a proposal to replace the new Keynesian Phillips curve. *Quarterly Journal of Economics* 117 (4), 1295–1328.
- Moscarini, G., Smith, L., 2001. The optimal level of experimentation. *Econometrica* 69 (6), 1629–1644.
- Moscarini, G., Smith, L., 2002. The law of large demand for information. *Econometrica* 70 (6), 2351–2366.

- Shannon, C.E., 1948. A mathematical theory of communication. *Bell System Technology Journal* 27, 379–423, 623–656.
- Sheshinski, E., Weiss, Y. (Eds.), 1993. *Optimal Pricing, Inflation, and the Cost of Price Adjustment*. MIT Press, Cambridge, MA.
- Sims, C., 1998. Stickiness. *Carnegie–Rochester Conference Series on Public Policy* (49)1, 317–356.
- Sims, C., 2003. Implications of rational inattention. *Journal of Monetary Economics* 50 (3), 665–690.
- van Zandt, E., 1999. Real-time decentralized information processing as a model of organizations with boundedly rational agents. *Review of Economic Studies* 66 (3), 633–658.