Prediction: The long and the short of it.

Antony Millner*[1] and Daniel Heyen[1]

[1]London School of Economics and Political Science

April 30, 2018

Abstract

Commentators often lament forecasters' inability to provide precise predictions of the long-run behaviour of complex economic and physical systems. Yet their concerns often conflate the presence of substantial long-run uncertainty with the need for long-run predictability; short-run predictions can partially substitute for long-run predictions if decision-makers can adjust their activities over time. So what is the relative importance of short- and long-run predictability? We study this question in a model of rational dynamic adjustment to a changing environment. Even if adjustment costs, discount factors, and long-run uncertainty are large, short-run predictability can be much more important than long-run predictability.

Keywords: Value of information, prediction, dynamic adjustment, long-run uncertainty
JEL codes: D80, D83

1 Introduction

Ever since Galileo wrote down his laws of motion in the early seventeenth century, the quantitative sciences have been engaged in the business of prophesy. Scientific ingenuity has rendered a staggering range of phenomena more predictable. Atmospheric scientists forecast the weather, epidemiologists predict the spread of infectious diseases, macroeconomists forecast economic growth, and demographers predict population change. Yet despite many successes, reliable predictions of the long run behaviour of complex social or natural systems often remain elusive (Granger & Jeon, 2007; Palmer & Hagedorn, 2006). Inability to predict the long run is frequently seen as a barrier to effective decision-making, and can be a source of

*Address: London School of Economics and Political Science, Houghton St, London, WC2A 2AE, UK. Tel: +44 207 107 5423. Email: a.millner@lse.ac.uk. We are grateful to audiences at EEA, EAERE, Imperial, Heidelberg, LSE, Montpellier, to numerous colleagues (especially Leo Simon, Larry Karp, and Derek Lemoine) for valuable comments, and to the CCCEP and Grantham Foundation for support.

emotional distress and planning inertia (Grupe & Nitschke, 2013). Concomitantly, improving long-run predictability is often a major goal of the scientific communities that produce forecasts. But just how important is it to be able to predict the distant future? Does substantial long-run uncertainty necessarily imply that accurate long-run predictions would be highly valuable? Or can long-run predictions be effectively substituted by short-run forecasts when decisions can be adjusted dynamically as new information arrives? This paper attempts to shed light on these questions.

It is not uncommon to find the presence of long-run uncertainty conflated with the need for improved long-run predictions.[1] For example, a recent report by The National Academy of Sciences (2016) on planned improvements in long-range weather forecasting suggests that 'Enhancing the capability to forecast environmental conditions outside the well-developed weather timescale – for example, extending predictions out to several weeks and months in advance – could dramatically increase the societal value of environmental predictions, saving lives, protecting property, increasing economic vitality, protecting the environment, and informing policy choices.' Similarly, many commentators have suggested that the lack of reliable projections of the local impacts of climate change, most of which will occur many decades hence, is a significant barrier to effective adaptation planning. Füssel (2007) contends that 'the effectiveness of pro-active adaptation to climate change often depends on the accuracy of [long run] regional climate and impact projections', while scientists at the World Modelling Summit for Climate Prediction in 2008 suggested that 'adaptation strategies require more accurate and reliable predictions of regional weather and climate...' (Dessai et al., 2009). One can find a similar casual identification of the presence of long-run uncertainty with the importance of long-run predictions in economics. Lindh (2011), for example, states that 'Very long-run...forecasts of economic growth are required for many purposes in long-term planning. For example, estimates of the sustainability of pension systems need to be based on forecasts reaching several decades into the future.'

While one-shot decisions with fixed lead times between actions and outcomes (e.g. agricultural planting decisions) doubtless benefit from predictability at decision-relevant time-scales, most long-run decision processes are at least partially flexible, and can thus be adjusted over time. Firms or individuals who anticipate long-run changes in market conditions, regulation, or their physical environments will adjust their actions dynamically as new information becomes available. Similarly, planners concerned with policies that depend on conditions in the distant future (e.g. social security) can alter the level of policy instruments (e.g. payroll taxes) dynamically as the future unfolds. A straightforward identification of the presence

[1] An anecdote related by Kenneth Arrow (1991) about his time as a military weather forecaster during World War Two provides an extreme example: 'Some of my colleagues had the responsibility of preparing long-range weather forecasts, i.e., for the following month. The statisticians among us subjected these forecasts to verification and found they differed in no way from chance. The forecasters themselves were convinced and requested that the forecasts be discontinued. The reply read approximately like this: The Commanding General is well aware that the forecasts are no good. However, he needs them for planning purposes.'

of long-run uncertainty with the importance of long-run predictions neglects this essential fact. Since the long-run today will become the short-run tomorrow, short-run predictions can play an important role in informing decision-making, even when long-run uncertainty is large. Indeed, it is intuitively clear that short-run predictability is a perfect substitute for long-run predictability if adjustment is costless. In general however adjustment is costly, and large abrupt changes in response to short-run warnings are often significantly more costly than managed gradual transitions which may be informed by long-run predictions. This suggests that long run predictions could play an important role in informing anticipatory planning, and avoiding excessive adjustment costs. It is however unclear *a priori* how the importance of predictability at different lead times depends on the magnitude of adjustment costs. We develop a simple analytical model in which this question is answerable.

Our model considers a decision-maker whose period payoffs depend on how well adapted her choices are to the current state of the world. The state of the world is uncertain, and may change over time in a non-stationary manner. The decision-maker may adjust her choices in every period to account for expected changes in her environment, but faces convex adjustment costs. This cost structure makes rapid adjustments in response to short-run warnings more costly than gradual incremental shifts of equal magnitude (which may be informed by long-run predictions).[2] Optimal decisions thus balance the benefits of exploiting current conditions with the need to anticipate future conditions in order to avoid costly rapid adjustments in the future. The decision-maker has access to a prediction system that generates forecasts of all future states in every period. These forecasts have a fixed profile of accuracy as a function of lead time. Thus, if τ_m is a measure of the accuracy of forecasts of lead time m, the decision-maker receives a forecast of accuracy $\tau_1, \tau_2, \ldots$ of states of the world $1, 2, \ldots$ time steps from the present in every period. For example, the decision-maker receives a forecast of accuracy τ_2 about a state two time steps from now in the current period, but knows that in the next period she will receive a new forecast of the same state, this time with accuracy τ_1. She may change her decisions in order to react to new predictions once they become available, but doing so entails a cost. Although the model reduces to a stochastic-dynamic control problem with an infinite number of state variables, we find a closed-form expression for the decision-maker's discounted expected payoffs V as a function of the profile of predictive accuracy that the prediction system exhibits:

$$V = V(\tau_1, \tau_2, \tau_3, \ldots).$$

By exploring the dependence of V on its arguments, and the parameters of the decision problem, we quantify the value of predictability at different lead times. Our central finding is that if we account for sequential forecast updating and agents' ability to adjust their activities

[2]Assuming convex adjustment costs is thus conservative with respect to adjudicating the importance of long-run predictability, as this assumption *favours* long-run predictions. See the text following equation (20) below for further discussion.

over time, short-run predictability is often more important than long-run predictability, even if adjustment costs, discount factors, and long-run uncertainty are large.

Although there is a sizeable literature on the value of information and its role in dynamic decision-making, as far as we know there are few direct antecedents to the questions we seek to address in this paper. The literature on the value of information began with the pathbreaking work of Blackwell (1953) and Marschak & Miyasawa (1968), who defined an incomplete ordering of the 'informativeness' of arbitrary information structures. We share this work's micro-oriented focus on the value of exogenous information sources for individual decision-makers, but also differ from it in important respects. In order to ensure tractability, our model makes strong assumptions about the nature of agents' payoff functions and the signals they receive. The return for this specificity is that we are able to study a much richer set of dynamic decisions than is typically used in this literature. Our focus on the dynamic characteristics of predictions, i.e. their accuracy as a function of lead time, is absent from this literature, and necessitates a pared down approach.

Work on the role of information in optimal dynamic decision-making falls into two categories: two period models that examine the effect of second period learning on optimal first period decisions (e.g. Arrow & Fisher, 1974; Epstein, 1980), or infinite horizon models that involve learning about the realizations of a stochastic state variable (e.g. Merton, 1971), or a parameter of a structural dynamic-stochastic model (e.g. Ljungqvist & Sargent, 2004). Neither of these standard approaches can capture the effects we study here. Two period models cannot capture the repeatedly updated nature of prediction and the dependence of predictability on lead time, both essential features of our model. Finite horizon models also suffer from an inherent bias towards short-run forecasts, as in a model with horizon H there will be H lead time 1 forecasts, but only one lead time H forecast. On the other hand, models based on familiar stochastic processes, or learning about parameters of structural models, do not usually allow the accuracy of predictions at different lead times to be controlled independently, meaning that it is impossible to ask questions about the relative importance of short- and long-run predictability (see the discussion on p. 8 for an elaboration of this point). We thus need a different approach if we are to define a model that is tractable, unbiased, disentangles lead times, and nevertheless retains coarse features of dynamic prediction.

A small applied literature studies the effect of forecasts at different lead times on dynamic decision-making. Costello et al. (2001) study a finite horizon stochastic renewable resource model and show that forecasts of shocks more than one step ahead carry no value for a resource manager. This result follows directly from the fact that their model is linear in the control variable; this removes the interactions between decisions in different periods, rendering long-run forecasts irrelevant. Costello et al. (1998) use numerical methods to study the effect of one and two period ahead forecasts in a calibrated non-linear resource management model, showing that for some parameter values perfect information at these lead times provides substantial value. Our work considerably generalizes these findings. We analyze a non-linear model that exhibits non-trivial interactions between time periods, use an infinite time horizon that removes bias against long-run forecasts, obtain closed-form solutions that enable clean comparative statics without the need for a calibrated numerical model, calculate the contribution of forecasts at all lead times to the overall value of a prediction system, and allow forecasts of arbitrary accuracy.

Finally, a substantial literature delineates the difficulties of long-run forecasting in contexts as diverse as climate science, macroeconomics, demography, epidemiology, and national security (see e.g. Palmer & Hagedorn, 2006; Granger & Jeon, 2007; Lindh, 2011; Lee, 2011; Myers et al., 2000; Yusuf, 2009). A common refrain in much of this work is that accurate long-run forecasting is difficult, but would be of considerable value for decision-makers if achievable. Yet to our knowledge there is no existing analytical framework that provides intuition for if, and when, this is likely to be true. Our work provides a step towards such a framework, illustrating in a simple model how a decision-maker's ability to adapt to changes in her environment dynamically, and the costs she sustains in doing so, co-determine the relative importance of short-run and long-run predictability.

2 The model

The model we develop is a variation on a work-horse model of rational dynamic adjustment to a changing environment that has been deployed in a variety of settings. These include modeling firm behaviour in the face of changing market conditions (e.g. Sargent, 1978; Fischer & Blanchard, 1989), and so-called 'target tracking' in military and engineering applications. The model provides a stylized and analytically tractable representation of a class of decision problems in which a decision-maker's period payoffs depend on an exogenously changing environmental variable, and changes in activities incur adjustment costs. We discuss our model's assumptions and how they differ from existing work below, but first spell out the details.

Consider a decision-maker who faces an uncertain exogenous environment at each time $n \in \mathbb{N}$. The units of time are arbitrary, but should be understood to match the frequency of forecast updates (e.g. days for weather forecasts, quarters for inflation forecasts). We assume that the decision-maker's possible choices can be mapped into the real line, and denote a generic choice by $X \in \mathbb{R}$. We will operate at a high level of abstraction, and thus leave the interpretation of X open. The more literal-minded reader is referred to Appendix A, where we provide a direct interpretation of the decision problem we examine in terms of a competitive firm making production decisions in the face of uncertain future prices. Other interpretations are of course possible, e.g. X could be the level of a tax set by a regulator, or an individual's stock of defensive capital.

The decision-maker may adjust X in each period, at a cost that is convex in the magnitude of the adjustment. Since large abrupt changes in activities are more costly than gradual incremental shifts of equal magnitude, the decision-maker has an incentive to engage in anticipatory planning. The state of the world at time n, denoted by $\tilde{\theta}_n \in \mathbb{R}$, is the loss-minimizing decision in that period. Values of X that are closer to $\tilde{\theta}_n$ are better adapted to conditions at time n, and give rise to higher period payoffs. The decision-maker's choices must achieve a balance between exploiting current conditions (i.e. choosing X close to the current expected value of $\tilde{\theta}$) and preparing for future conditions (i.e. shifting X towards expected future values of $\tilde{\theta}$), thus avoiding excessively large and costly adjustments later on.

For any time n, let $\theta_t = \tilde{\theta}_{n+t}$ for $t \geq 0$, i.e. θ_t is the value of the loss-minimizing decision $\tilde{\theta}$ that will be realized t time steps in the future (see Figure 1 below for an illustration). We denote the agent's beliefs about θ_t at time n by $p_n(\theta_t)$. At $n = 0$ the agents' prior beliefs about the future values θ_t are captured by an infinite sequence of normal distributions with means μ_t^0 and precisions (i.e. inverse variance) λ_t^0, i.e.

$$p_0(\theta_t) \sim \mathcal{N}(\mu_t^0, 1/\lambda_t^0). \tag{1}$$

The values of μ_t^0 and λ_t^0 are unconstrained, allowing us to describe a wide variety of beliefs about the future. In particular, we do not require the agent to believe that the environmental random variable $\tilde{\theta}$ is identically distributed over time.

Let X_n be the value of the decision variable X that the agent inherits *at the beginning of period* n. At the beginning of the period the agent chooses a new value for X, i.e. X_{n+1}. This is the value of X that will affect payoffs in the current period, and be passed forward to the next period. The cost of modifying the decision variable from X_n to X_{n+1} is $\frac{1}{2}\alpha(X_{n+1} - X_n)^2$, where $\alpha \geq 0$ is a parameter that captures the magnitude of the adjustment costs the agent faces. After the choice of X_{n+1} is made, the agent experiences the realization of the current value of $\tilde{\theta}$, i.e. θ_0, and sustains a loss equal to half the squared distance between X_{n+1} and θ_0. Thus, the expected period payoff at the beginning of the current period is given by,

$$W(X_{n+1}, X_n, p_n(\theta_0)) = -\frac{1}{2}\left[\int (X_{n+1} - \theta_0)^2 p_n(\theta_0) d\theta_0 + \alpha(X_{n+1} - X_n)^2\right]. \tag{2}$$

The decision-maker's objective function is the usual discounted sum of expected period payoffs, which will be defined in full below. As advertised, the reader seeking an interpretation of this payoff function in a familiar economic application is referred to Appendix A.

To model the effect of predictions on the agent's beliefs, we assume that at the end of each period n, the agent receives a sequence of forecasts $S^n = (s_t^n)_{t\geq 1}$ of the values of future states θ_t for all $t \geq 1$. We assume that

$$s_t^n = \theta_t + \epsilon_t^n \tag{3}$$

where

$$\epsilon_t^n \sim \mathcal{N}(0, 1/\tau_t) \tag{4}$$

and $\tau_t \geq 0$ is the precision of forecasts of events t time steps ahead. Thus predictions have an exogenous profile of precision as a function of lead time, parameterized by the infinite sequence of parameters

$$\vec{\tau} \equiv (\tau_1, \tau_2, \tau_3, \ldots).$$

The precision sequence $\vec{\tau}$ is time invariant, i.e. the prediction system is assumed to produce forecasts with the same profile of accuracy as a function of lead time at every time n.[3] Consider two prediction systems A and B, with precision sequences $\vec{\tau}^A, \vec{\tau}^B$. If, for a fixed lead time t, $\tau_t^A > \tau_t^B$, then A is more informative (in the sense of Blackwell (1953)) than B about events t time steps in the future. Although in practice we would expect $\vec{\tau}$ to be a decreasing sequence, we place no constraints on its value in what follows.

Notice that the agent receives a new sequence of forecasts S^n at the end of every period n. However, the precision of the information she receives about a particular value $\tilde{\theta}_k$ that lies in her future changes as time progresses and she moves closer to time k. Since the agent's prior beliefs about the future values θ_t in period $n = 0$ are normal, and the conditional distributions of signals s_t given states are normal, her beliefs about the future values of the states will update according to the standard normal-normal Bayesian formulae (see e.g. DeGroot, 1970). In particular, beliefs about future values θ_t will be normally distributed in every period, and are characterized by a mean μ_t^n and precision λ_t^n in period n. Moreover, the agent also knows that the beliefs she currently holds about the future values θ_t for $t \geq 1$, will become her beliefs about θ_{t-1} in the next period. For example, her current beliefs about the next period will become her beliefs about the current period, in the next period. Using these observations, we can write down the state equations that describe how the forecasting system changes the agents' beliefs about the values of the states θ_t from one period to the next:

$$\begin{aligned} \mu_t^{n+1}(s_{t+1}^n) &= \frac{\tau_{t+1}}{\tau_{t+1} + \lambda_{t+1}^n} s_{t+1}^n + \frac{\lambda_{t+1}^n}{\tau_{t+1} + \lambda_{t+1}^n} \mu_{t+1}^n \\ \lambda_t^{n+1} &= \lambda_{t+1}^n + \tau_{t+1}. \end{aligned} \tag{5}$$

As is standard in the normal-normal bayesian updating model, the posterior mean of beliefs about each future value θ_t is a convex combination of the prior mean and the signal realization, with the weight that is placed on the signal increasing in the signal precision. Posterior precisions, however, evolve deterministically. A complete description of the current state of

[3]While we make this assumption for simplicity and clarity, we note that any covariance stationary time series satisfies the time invariance property. We mention this not because our model is covariance stationary (it need not be), but to illustrate that this property is not unusual.

the system at the beginning of period n is thus given by the ordered pair (X_n, Y_n), where

$$Y_n \equiv ((\mu_t^n)_{t\geq 0}, (\lambda_t^n)_{t\geq 0}), \tag{6}$$

collects together the infinitely many 'belief' state variables. The dynamics of Y_n are given by (5), and X_n is a 'decision' state variable whose next value is chosen by the agent in each period. Figure 1 provides a graphical summary of the model setup, and the timing of events.

Before proceeding to the solution of the model, we now discuss some of its more unusual assumptions, why they are necessary, and how they relate to existing work. The payoff structure in the model, captured by (2), is formally identical to that in previous models of dynamic adjustment we alluded to at the beginning of this section. The novelty in our approach arises from our representation of the decision-maker's dynamic expectations, i.e. the updating process summarized in (5). In existing applications of the dynamic adjustment model the loss-minimizing decisions $\tilde{\theta}$ are always modeled as a stochastic process. Sargent (1978), for example, assumes that $\tilde{\theta}_n$ follows an AR(1) process. However, this approach cannot be used if we are to gain traction on our central question, i.e. understanding the relative importance of short- and long-run predictability. To understand why, suppose that the $\tilde{\theta}_n$ are correlated (as they would be if modeled as a stochastic process), and that a prediction system provides information about some $\tilde{\theta}_k$. Then learning about $\tilde{\theta}_k$ means that we learn something about *all* values of $\tilde{\theta}_n$. It is thus not possible to associate a prediction about an event at a given lead time with a change in uncertainty at only that lead time in a correlated model. Our central question is thus generically unanswerable in correlated models, unless the correlation structure can be fine-tuned to generate arbitrary patterns of predicability as a function of lead time.[4] Hence our assumption that the $\tilde{\theta}_n$ are independent (but not identically distributed).[5] By contrast, the belief updating system described in (5) retains several features of dynamic prediction (i.e. sequentially updated expectations, and predictions whose accuracy depends on lead time), but allows us to control the precision of predictions at each lead time independently. The cost of this approach is that instead of being able to describe dynamic expectations with a single state variable as in standard stochastic process models, we now require an infinite number of state variables, one for each independent belief about each future event. In sum, while the independence assumption may seem unusual to some readers, it is a necessary expediency if we are to generate insights into the relative

[4]This could be possible in models where the state equation depends on many lagged variables (e.g. AR(n) models), however one needs an infinite number of lags to disentangle all lead times, the relationship between parameters of the state equation and predictability at different lead times is highly complex, and such models severely constrain agents' prior beliefs about the future due to their stationarity. In addition, dynamic optimization problems in these high lag environments are often intractable. By contrast our stylized model parsimoniously captures lead time dependent predictive accuracy, maintains tractability, and allows us to describe prior uncertainty in a flexible manner. Costello et al. (1998) use a similar independence assumption.

[5]Of course, if the $\tilde{\theta}_n$ are independent *a priori*, but the signals s_t^n are correlated, this induces correlation between time periods after a single forecast is received. We thus also require the s_t^n to be independent.

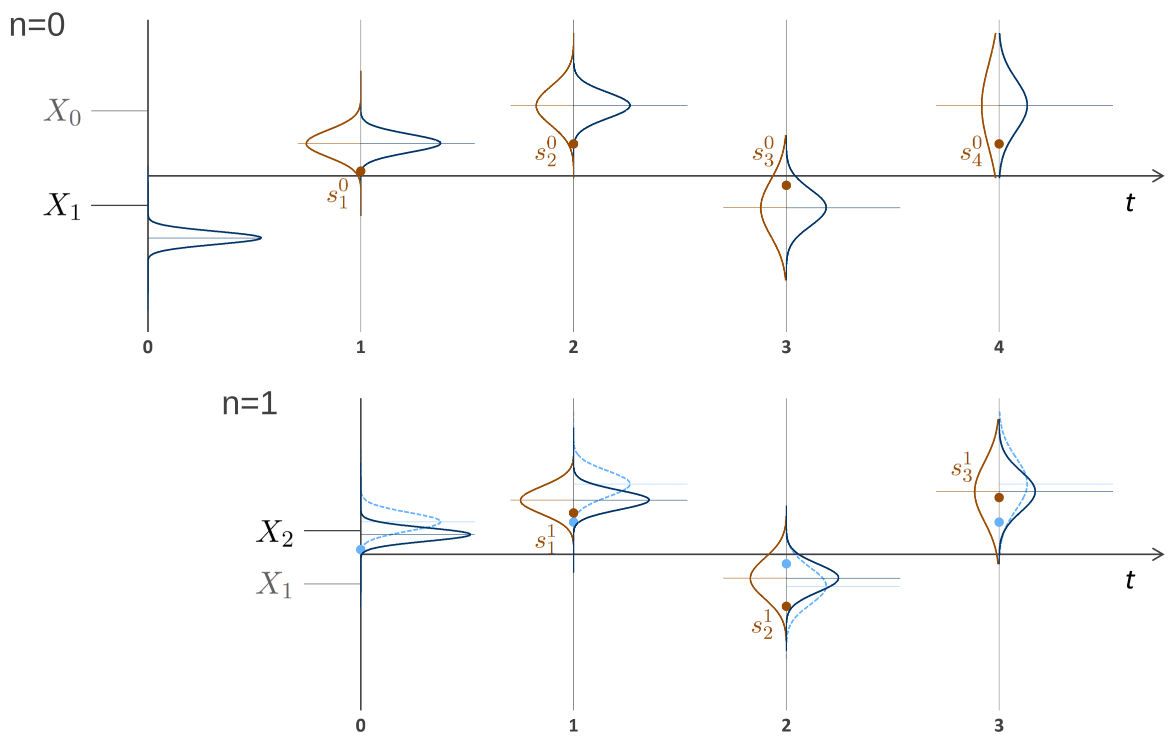


Figure 1: Illustration of the model setup. The figure depicts the agent's beliefs, choices, and the information provided by the prediction system, in the first two periods $n = 0, 1$. At the beginning of period $n = 0$ the agent holds a sequence of prior beliefs about the future values of the state of the world θ_t. Beliefs about each future value of θ_t are normally distributed, indicated by the dark blue distributions at each value of t. The initial value of the decision variable is $X = X_0$, and the agent must choose a new value X_1 at the beginning of the period. At the end of the period the agent receives the infinite sequence of forecasts $S^0 = (s_t^0)_{t \geq 1}$, indicated by the dark red dots, which allow her to update her beliefs about the future values θ_t. The brown distributions at each t capture the agent's initial expectations about the signals s_t^0 she will receive (i.e. $q(s_t^0; Y_0)$ in (7)). Smaller values of τ_t, which are assumed to occur at longer lead times in this example, correspond to wider distributions of expected forecast realizations, and weaker belief updating towards the realized signal. This is demonstrated by the agent's updated beliefs at $n = 1$, where once again beliefs about future values of θ_t at the beginning of the period appear in dark blue, and for comparison the agent's $n = 0$ beliefs and signal realizations are represented by the light blue distributions and dots respectively. Once again, the agent must choose a new value for the decision variable X_2 at the beginning of the period $n = 1$, and will receive a new sequence of forecasts $S^1 = (s_t^1)_{t \geq 1}$ (indicated by dark red dots), with the same profile of precisions at the end of period $n = 1$.

importance of predictability at different lead times.[6] We now turn to the model's solution.

Bellman equation

Let $V(X_n, Y_n)$ be the current value of the infinite dimensional state (X_n, Y_n), where Y_n is defined in (6). The next period value of the state depends on the sequence of signals S^n that the agent will receive at the end of the current period. At the beginning of period n the agent's beliefs about signal s_t^n $(t \geq 1)$ are given by:

$$\begin{aligned} q(s_t^n; Y_n) &= \int p(s_t^n|\theta_t) p_n(\theta_t) d\theta_t \\ &\sim \mathcal{N}(\mu_t^n, 1/\lambda_t^n + 1/\tau_t), \end{aligned} \tag{7}$$

where the last line follows from a simple calculation using (3–4). We denote the agent's beliefs about the probability of receiving a sequence of signals $S^n = (s_t^n)_{t\geq 1}$ by

$$Q(S^n; Y_n) = \prod_{t=1}^{\infty} q(s_t^n; Y_n). \tag{8}$$

We are now ready to state the Bellman equation for the value function $V(X_n, Y_n)$. Denote the next period value of the belief states Y_{n+1} as a function of the previous value Y_n and the realized signal sequence S^n as

$$Y_{n+1} = F(Y_n, S^n). \tag{9}$$

where $F(Y_n, S^n)$ is given by (5). Then,

$$V(X_n, Y_n) = \max_{X_{n+1}} W(X_{n+1}, X_n, Y_n) + \beta \int V(X_{n+1}, F(Y_n, S^n)) Q(S^n; Y_n) dS^n, \tag{10}$$

where $dS^n = \prod_{t=1}^{\infty} ds_t^n$, $\beta \in (0,1)$ is the agent's discount factor, and we have changed notation slightly to emphasize the dependence of the period payoff function (2) on the belief state variables Y_n. Note that the dependence of the value function on the profile of forecast precisions $\vec{\tau}$ comes both through the updating rule $F(Y_n, S^n)$ (see eq. (5)), and through the agent's expectations about the values of future forecast realizations (see eq. (7)). Thus, increases in predictability affect both the quality of future decisions (by reducing the variance of outcomes), and the agent's expectations about the information that will be available in the future.

Optimal policy

The model is a stochastic dynamic control problem with an infinite number of state variables, since the agent holds an independent belief about each future value θ_t. Despite the infinite

[6] Sequential event forecasts of the kind we consider have been studied in e.g. Clements (1997); Selten (1998).

dimensionality of the state space in our model,[7] standard methods based on the Benveniste-Scheinkman condition (Benveniste & Scheinkman, 1979) yield simple closed form solutions for the optimal control rule. We state this rule in some detail, as it will help us to interpret the main results below. All proofs can be found in the appendices.

Proposition 1. *The optimal policy* $X_{n+1} = \pi(X_n, Y_n)$ *is given by*

$$\pi(X, Y) = aX + \sum_{t=0}^{\infty} b_t \mu_t \tag{11}$$

where

$$\begin{aligned} a &= \frac{1 + \alpha(1+\beta) - \sqrt{(1+\alpha(1+\beta))^2 - 4\alpha^2\beta}}{2\alpha\beta}, \\ b_t &= \frac{a}{\alpha} (a\beta)^t . \end{aligned}$$

It is straightforward to demonstrate the following properties of the coefficients a, b_t:

$$\begin{aligned} & a + \sum_{t=0}^{\infty} b_t = 1 \\ & \lim_{\alpha \to 0} a = 0, \ \lim_{\alpha \to \infty} a = 1, \ \frac{\partial a}{\partial \alpha} > 0 \\ & \lim_{\alpha \to 0} b_t = \begin{cases} 1 & t = 0 \\ 0 & t > 0 \end{cases}, \ \lim_{\alpha \to \infty} b_t = 0 \\ & \frac{\partial}{\partial \alpha} \left(\frac{b_{t+1}}{b_t} \right) > 0, \ \frac{\partial b_0}{\partial \alpha} < 0. \end{aligned}$$

Proposition 1 thus shows that the optimal policy function $\pi(X, Y)$ chooses the next value of X to be a convex combination of the current value of X and the expected values of θ_t. The policy rule exhibits the certainty equivalence property, i.e. it is independent of the agent's uncertainty about future events. This is a well-known consequence of the quadratic payoff function in our model, which makes the model tractable (e.g. Ljungqvist & Sargent, 2004). Although the policy rule does not depend on uncertainty, the value function certainly will, and it is it's dependence on the precision profile $\vec{\tau}$ that we are ultimately interested in.

The coefficients of the policy rule have an intuitive dependence on the adjustment cost

[7]The curse of dimensionality normally renders stochastic-dynamic control problems in state spaces of even moderate dimension (e.g. more than 10) intractable. Our model is, as far as we know, the only example of a non-trivial stochastic-dynamic model on an infinite dimensional state space that admits an exact closed-form solution for the value function.

parameter α. Consider the extreme cases $\alpha \to 0$, and $\alpha \to \infty$. The proposition shows that

$$\lim_{\alpha \to 0} \pi(X, Y) = \mu_0$$
$$\lim_{\alpha \to \infty} \pi(X, Y) = X.$$

When adjustment costs tend to zero, the policy rule does not depend on either X or μ_t for $t \geq 1$. This occurs since with costless adjustment the decision problem separates into a sequence of static optimization problems, and the payoff maximizing choice in each of these problems is simply to choose X equal to the expected value of the current value of $\tilde{\theta}$, i.e. μ_0. When $\alpha \to \infty$, any change in the value of X is very costly, so the optimal action is to leave X where it is. In between these extremes the policy rule depends on expectations about all future values of θ_t. As α increases from zero the decision maker's choice depends more on both the inherited value of X, and her expectations about the future. This occurs since the convexity of adjustment costs penalizes large adjustments later on. Current choices thus account for both the benefits of adjusting to current conditions and the need to anticipate future conditions. The larger is α, the more important it is to anticipate future conditions, and this is reflected in the fact that coefficients b_t decrease at a slower rate as α increases. At the same time, larger α makes adjustments more costly, leading the policy rule to place greater weight on the inherited value of X. Finally, to understand the finding that $a + \sum_{t=0}^{\infty} b_t = 1$, consider the case in which $\mu_t = X$ for all t. In this case the agent believes that her choice is perfectly adapted to conditions now and in the future, and she should thus not want to change X. This occurs if $aX + \sum_{t=0}^{\infty} b_t X = X$.

It will be helpful in what follows to have some quantitative understanding of which values of α are 'large' and 'small' in some absolute sense. To benchmark how α affects optimal policies, consider a deterministic version of the model in which the states $\tilde{\theta}_n$ are chosen to be a fixed sequence of draws from an arbitrary univariate random variable with finite variance. When $\alpha = 0$, optimal decisions coincide with the current value of $\tilde{\theta}_n$, i.e. $X_n = \mu_0 = \tilde{\theta}_n$ for all n. As α increases, adjustment becomes more costly, and the values of X_n fluctuate less than $\tilde{\theta}_n$ itself. Appendix C derives an expression for the asymptotic variance of the policy choices X_n as a function of α and β. For a wide range of β, $\alpha > 1.5$ implies that the decision maker adjusts to less than 20% of the variability in $\tilde{\theta}$, and $\alpha > 3$ implies adjustment to less than 10% of the variability. Thus, $\alpha = 3$ is already a fairly large value of the adjustment cost parameter. In addition, the appendix demonstrates that changes in α have a greater effect on behaviour when α is small (e.g. $\alpha < 1$) than when it is large.

Value function

In order to understand the effect of the precision sequence $\vec{\tau}$ on the agent's expected payoffs, we need to compute the value function. This would seem to be difficult, as the model's state space is infinite dimensional, the period payoff depends non-quadratically[8] on the precision state variables λ_t, and we need to take expectations of the value function over an infinite sequence of signals, the distribution of which depends on all the belief state variables. Despite these apparent obstacles it is possible to obtain a closed form solution for the value function, which enables the remainder of our analysis.

Begin by defining a shift operator Δ that acts on infinite vectors $\vec{Z} = (z_t)_{t\geq 0}$ as follows:

$$\Delta(\vec{Z}) = \Delta((z_0, z_1, z_2, \ldots)) \equiv (z_1, z_2, z_3, \ldots).$$

Thus Δ simply deletes the first element of $\vec{Z}$ and shifts all the other elements forward one position. The belief updating rule (5) for the vector of prior precisions $\vec{\lambda} = (\lambda_t)_{t\geq 0}$ can thus be written as:

$$F(\vec{\lambda}) = \Delta(\vec{\lambda}) + \vec{\tau}, \tag{12}$$

where $F(\vec{\lambda})$ denotes the $\vec{\lambda}$ components of the updating rule in (5). Define $F_k^t(\vec{\lambda})$ to be the $(k+1)$-th element of $F^t(\vec{\lambda})$.[9] In addition, recall that b_k is the coefficient of μ_k in the optimal control rule (11). Then,

Proposition 2. *The value function $V(X, Y)$ is given by*

$$V(X, Y) = T(\vec{\tau}) + \text{terms independent of } \vec{\tau},$$

where

$$\begin{aligned} T(\vec{\tau}) &\equiv -\frac{1}{2} b_0 \left[\sum_{t=1}^{\infty} \beta^t \sum_{k=0}^{\infty} \left(\frac{b_k}{b_0} \right)^2 \frac{1}{F_k^t(\vec{\lambda})} \right] \qquad (13) \\ &\propto - \Bigg[\left(\frac{1}{\lambda_1 + \tau_1} + (a^2\beta^2) \frac{1}{\lambda_2 + \tau_2} + (a^2\beta^2)^2 \frac{1}{\lambda_3 + \tau_3} + \ldots \right) \\ &+ \beta \left(\frac{1}{(\lambda_2 + \tau_2) + \tau_1} + (a^2\beta^2) \frac{1}{(\lambda_3 + \tau_3) + \tau_2} + (a^2\beta^2)^2 \frac{1}{(\lambda_4 + \tau_4) + \tau_3} + \ldots \right) \\ &+ \beta^2 \left(\frac{1}{(\lambda_3 + \tau_3 + \tau_2) + \tau_1} + (a^2\beta^2) \frac{1}{(\lambda_4 + \tau_4 + \tau_3) + \tau_2} + (a^2\beta^2)^2 \frac{1}{(\lambda_5 + \tau_5 + \tau_4) + \tau_3} + \ldots \right) + \mathcal{O}(\beta^3) \Bigg]. \end{aligned}$$

[8]From (2) we have

$$W(X_{n+1}, X_n, Y_n) = -\frac{1}{2} \left[(1+\alpha) X_{n+1}^2 + \alpha X_n^2 - 2X_{n+1}(\mu_0 + \alpha X_n) + \frac{1}{\lambda_0^n} + (\mu_0^n)^2 \right].$$

One could write this payoff in terms of the variance of the agent's beliefs about θ_t, making it linear in variances, but then the state equations for the evolution of variances would be non-linear (see (5)). Standard methods from linear quadratic control are not applicable.

[9]For example,

$$F^2(\vec{\lambda}) = F(F(\vec{\lambda})) = F(\Delta(\vec{\lambda}) + \tau) = \Delta(\Delta(\vec{\lambda}) + \vec{\tau}) + \vec{\tau} = \Delta^2(\vec{\lambda}) + \Delta(\vec{\tau}) + \vec{\tau}.$$

Thus say $F_0^2(\vec{\lambda}) = \lambda_2 + \tau_2 + \tau_1$, where it is important to recall that $\vec{\tau} = (\tau_1, \tau_2, \ldots)$.

To interpret this result notice that the term $\sum_{k=0}^{\infty}\left(\frac{b_k}{b_0}\right)^2 \frac{1}{F_k^t(\vec{\lambda})}$ in (13) represents the contribution to the value function from the uncertainty the agent faces when she takes a decision t time steps in the future. $1/F_k^t(\vec{\lambda})$ is the agent's uncertainty about events that are k time steps in the future, in period t. The exponentially declining factor $(b_k/b_0)^2 = (a^2\beta^2)^k$ captures the importance of uncertainty about events at temporal distance k for decision-making, as can be seen from the optimal policy rule (11). $T(\vec{\tau})$ is thus the discounted sum of the cost of uncertainty for each future decision. The forecasting system reduces this uncertainty cost by providing information about all future periods, in every period. The agent's uncertainty about events that are k time steps in the future in a period t time steps from now is reduced by forecasts of precision $\tau_{t+k}, \tau_{t+k-1}, \ldots, \tau_k$.

3 The relative value of short- and long-run predictability

The previous section derived an expression for the decision-maker's value function for an arbitrary prediction system that obeys (5). In this section we unpack this result in order to study the relative importance of short- and long-run predictability. Given prior uncertainty $\vec{\lambda} = (\lambda_t)_{t\geq 1}$, the function $T(\vec{\tau})$ in Proposition 2 depends on the sequence of forecast precisions $\vec{\tau}$. Our goal now is to understand the dependence of $T(\vec{\tau})$ on its arguments. In general this is a complex task, as $T(\vec{\tau})$ is a non-separable function of the τ_m. The following subsections consider different methods for extracting the information $T(\vec{\tau})$ contains about the relative importance of predictability at different lead times.

3.1 Marginal predictability

To make initial progress we begin by finding a linear approximation to $T(\vec{\tau})$ at $\vec{\tau} = \vec{0}$. This approximation will only be accurate when forecast precisions are *marginal*. Studying a linearized version of $T(\vec{\tau})$ has two purposes. First, the linear approximation to $T(\vec{\tau})$ is a separable function of $\vec{\tau}$, allowing the contribution of each τ_m to the value function to be computed easily in this case. This allows us to form clear intuition for the effects that determine the relative importance of different lead times, *to first order*. Second, and more importantly, in this approximation all interactions between forecast lead times are neglected. Since we expect the interesting effects in the model to be a consequence of the sequential updating of forecasts, which allow short-run predictability to partially substitute for long-run predictability, it is useful to first examine a baseline case in which those substitution effects (i.e. interactions between lead times) are effectively switched off. This will allow us to demonstrate later on how accounting for interactions between lead times alters the relative importance of short- and long-run predictability.

Begin by defining the function

$$g(m) \equiv \sum_{k=0}^{\infty} \frac{\beta^k}{\lambda_{m+k}^2}, \tag{14}$$

and assuming that

$$\lim_{t\to\infty} \frac{\lambda_{t+1}^2}{\lambda_t^2} > \beta, \tag{15}$$

implying that $g(m)$ is finite for all m (by the ratio test). Then,

Proposition 3. *If the interactions between forecast lead times are neglected, the increase in the value function due to the prediction system (relative to an uninformative baseline) is*

$$dV = T(d\vec{\tau}) - T(\vec{0}) \approx \frac{a}{\alpha(1-a^2\beta)} \sum_{m=1}^{\infty} r_m d\tau_m, \tag{16}$$

where

$$r_m \equiv g(m)\beta^m \left(1 - (a^2\beta)^m\right). \tag{17}$$

To understand the intuition behind this result we now derive it heuristically (the appendix contains a formal proof). Recall that the agent receives a forecast of lead time m in every period. The effect of the forecast the agent receives in the current period is to reduce uncertainty about events at temporal distance m. But, in doing so, this forecast gives rise to a cascade of uncertainty reductions at shorter lead times in future periods. This occurs since a reduction in uncertainty about lead time m events in the current period is equivalent to a reduction in uncertainty about events at lead time $m-1$ in the next period, and lead time $m-2$ in the period after that, etc. As (13) makes clear, the value of a reduction in uncertainty about events k time steps in the future is proportional to $(b_k/b_0)^2 = (a^2\beta^2)^k$. Since uncertainty reductions in future periods are discounted, a marginal unit of precision in the first forecast of lead time m that the agent receives increases payoffs by an amount proportional to

$$\sum_{k=0}^{m-1} \beta^{m-k}(a^2\beta^2)^k = \beta^m \sum_{k=0}^{m-1} (a^2\beta)^k.$$

Because a marginal increase in the precision of forecasts of lead time m increases payoffs in proportion to $\frac{d}{d\lambda_m}(-1/\lambda_m) = 1/\lambda_m^2$, the total effect of the first forecast of lead time m is to increase payoffs by an amount proportional to

$$\frac{1}{\lambda_m^2}\beta^m \sum_{k=0}^{m-1} (a^2\beta)^k.$$

This quantity accounts for the uncertainty reduction effect of the first forecast of lead time m, which the agent receives at the end of the current period. At the end of the next period, the agent receives another forecast of lead time m. This forecast gives rise to the same cascade of uncertainty reductions, and has the same value as the initial forecast, up to a normalization. The normalization is simply the discounted value of the change in lead time m uncertainty that the agent faces in the next period, i.e. $\beta\frac{1}{\lambda_{m+1}^2}$. This occurs in all future periods. Thus, the total value of a marginal unit of precision in forecasts of lead time m is proportional to:

$$\begin{aligned}
&\frac{1}{\lambda_m^2}\left(\beta^m\sum_{k=0}^{m-1}(a^2\beta)^k\right)+\frac{\beta}{\lambda_{m+1}^2}\left(\beta^m\sum_{k=0}^{m-1}(a^2\beta)^k\right)+\frac{\beta^2}{\lambda_{m+2}^2}\left(\beta^m\sum_{k=0}^{m-1}(a^2\beta)^k\right)+\dots\\
&\propto\left(\sum_{t=0}^{\infty}\frac{\beta^t}{\lambda_{m+t}^2}\right)\beta^m\left(1-(a^2\beta)^m\right)
\end{aligned}$$

This is exactly the expression we obtained in (17). Notice how the derivation of this expression makes it clear that sequential updating of forecasts is not a major determinant of the value of a *marginal* unit of predictability. The fact that forecasts are updated sequentially gives rise to the factor $\left(\sum_{t=0}^{\infty}\frac{\beta^t}{\lambda_{m+t}^2}\right)=g(m)$ in (17), but if only a *single* marginal forecast of lead time m were received in the first period, the expression in (17) would look very similar, with this factor simply replaced by $\frac{1}{\lambda_m^2}$. Thus, neglecting the interactions between lead times is qualitatively similar to neglecting sequential forecast updating itself (we make this analogy exact in a special case below).

Equation (17) makes it clear that the dependence of prior uncertainty on lead time can have an important influence on the value of marginal predictability at different lead times through the function $g(m)$. To understand these effects in a parsimonious way we will focus on a simple parametric model of prior beliefs. We suppose that the precisions of prior beliefs about the locations of θ_t are given by,

$$\lambda_t^2=\phi^t\lambda_0^2+(1-\phi^t)\lambda_\infty^2, \tag{18}$$

where $\phi\in(0,1]$, and $0<\lambda_\infty^2<\lambda_0^2$. In this model the squared precision of prior beliefs about events decays geometrically from λ_0^2 for the current period to λ_∞^2 for events in the infinite future. It is straightforward to verify that (15) is always satisfied in this case as long as $\lambda_\infty^2>0$. Moreover, notice that if beliefs about the infinitely distant future are arbitrarily uncertain, i.e. $\lambda_\infty^2\to 0$, we have

$$\lim_{\lambda_\infty^2\to 0^+}\frac{g(m)}{g(1)}=\lim_{\lambda_\infty^2\to 0^+}\frac{\sum_{k=0}^{\infty}\frac{\beta^k}{\lambda_{m+k}^2}}{\sum_{k=0}^{\infty}\frac{\beta^k}{\lambda_{1+k}^2}}=\lim_{\lambda_\infty^2\to 0^+}\frac{\sum_{k=0}^{\infty}\frac{\beta^k}{\phi^{m+k}\lambda_0^2+(1-\phi^{m+k})\lambda_\infty^2}}{\sum_{k=0}^{\infty}\frac{\beta^k}{\phi^{1+k}\lambda_0^2+(1-\phi^{1+k})\lambda_\infty^2}}=\left(\frac{1}{\phi}\right)^{m-1}. \tag{19}$$

Thus in this (not implausible) limit the ratio $g(m)/g(1)$ takes an especially simple form. The limiting ratio in (19) is well defined for all $\phi\in(0,1]$, even though $g(m)$ itself diverges if $\lambda_\infty^2=0$ and $\phi<\beta$.

In the limit as $\lambda_\infty^2 \to 0$, we can thus define a simple measure of the value of a unit of predictability about events at distance m, relative to the value of a unit of predictability about events at distance 1:

$$R_m \equiv \frac{r_m}{r_1} = \underbrace{\beta^{m-1}}_{\text{Discounting}} \underbrace{\left(\frac{1}{\phi}\right)^{m-1}}_{\text{Uncertainty}} \underbrace{\left[\frac{1-(a^2\beta)^m}{1-a^2\beta}\right]}_{\text{Early warning}}. \tag{20}$$

Using this expression, the relative value of the predictability of events at different lead times may be computed as a function of the three parameters α, β and ϕ. These parameters characterize the decision-maker's flexibility, impatience, and prior uncertainty about the future respectively. Aside from being simple to analyze, the choice of priors in (18) makes the formulas for r_m/r_1 for updating from a single forecast vs. updating from sequential forecasts coincide exactly in the limit as $\lambda_\infty^2 \to 0$, since $\frac{1/\lambda_m^2}{1/\lambda_1^2} = g(m)/g(1)$ in this case. To a first approximation there is thus no difference between once-off and sequential forecasting in this model of priors. This is thus an especially good baseline from which to assess how accounting for the interactions between lead times alters the balance between short- and long-run predictability.

The formula (20) shows that there are three effects that determine the relative value of marginal predictability at different lead times. First, since forecasts at larger lead times relate to more distant payoffs, they are more heavily discounted. This gives rise to the first term in (20), which is *decreasing* in m. Second, since the prior precision of beliefs is smaller for larger lead times (i.e. uncertainty increases with the time horizon), and payoffs are concave in precisions, the effect of a marginal increase in the precision of beliefs is increasing in lead times. This leads to the second term in (20), which is *increasing* in m, reflecting the fact that the long-run is more uncertain than the short-run. Finally, the third term captures the cumulative effect of an early warning about events at lead time m on all subsequent decisions that are made until that event is realized. Since warnings of lead time m give rise to improved decision-making for $m-1$ subsequent adjustment decisions, longer lead times are associated with greater cost savings. Thus the third term in (20) is *increasing* in m, reflecting the fact that earlier warnings give rise to cheaper adjustments (since adjustment costs are convex). It is moreover straightforward to verify that

$$\frac{\partial^2}{\partial\alpha\partial m}\left[\frac{1-(a^2\beta)^m}{1-a^2\beta}\right] > 0$$

when $m \geq 1$. Thus, the larger are adjustment costs α, the faster the third term in (20) increases with m. This is intuitive, since the more costly adjustments are, the more important it is to get early warning of the need for them (again due to convexity). This term thus demonstrates how our assumption of convex adjustment costs *favours* long-run predictions.[10]

The overall dependence of R_m on m depends on the relative rates of increase and decrease of the three terms in (20). Some simple analysis (see Appendix F) shows that R_m can exhibit only three kinds of qualitative behaviour. First, if $\phi \leq \beta$, R_m is an increasing function of m. In this case, the benefits of reducing large long-run uncertainties outweigh the effects of discounting, making long-run predictability more important than short-run predictability (when interactions between lead times are neglected). If $\phi > \beta$, R_m either decreases monotonically with m or is a unimodal function with a global maximum at some $m \geq 2$. Appendix F characterizes the regions of parameter space where these two qualitative behaviours occur. In general, when β is sufficiently small, R_m will be declining in m for all values of α. However, when β exceeds some critical value $\hat{\beta}$, there exists an $\hat{\alpha} > 0$ such that for all $\alpha > \hat{\alpha}$ R_m is unimodal. Analytic expressions for $\hat{\beta}$ and $\hat{\alpha}$ show that the faster prior uncertainty increases with lead time (i.e. the lower is ϕ), the lower are $\hat{\beta}$ and $\hat{\alpha}$. Figure 2 plots R_m for several values of the parameters.

Taken at face value, the first order analysis in this sub-section would seem to suggest that long-run predictability is often significantly more important than short-run predictability. When $\phi < \beta$ in (20), i.e. when the long-run is significantly more uncertain than the short run, the analysis of R_m suggests that long-run predictability has a greater effect on discounted expected payoffs than short-run predictability, regardless of the adjustment cost parameter α. Moreover, even when $\phi > \beta$ it is possible to find values of the parameters for which R_m increases for a long time, before declining.[11] However, as we have emphasized, the first-order analysis largely neglects the dynamic nature of decision-making; it does not account for the interactions between lead-times, and thus cannot reflect the substitution possibilities that are a consequence of sequential forecast updating. This analysis thus defines a naive baseline that is conceptually similar to a conflation of the presence of long-run uncertainty with the need for long-run predictions.

[10]By contrast, linear adjustment costs give rise to no incentive to anticipate future changes in the environment (since rapid and gradual adjustments of equal magnitude are equally costly in this case) – in this case short-run predictions can substitute perfectly for long-run predictions (see Costello et al., 2001). Concave costs (including fixed costs) would give rise to lumpy optimal adjustments in which activities are only adjusted infrequently when the marginal benefit of adjustment is believed to exceed its marginal cost (which is high for small adjustments). In this case the agent obtains no cost savings from gradualism, and thus has less opportunity to exploit early warnings. Since adjustment occurs only infrequently in this case, intuition suggests that discounting will be the dominant determinant of the relative value of predictability at different lead times, as indeed it is in once-off adjustment decisions. Tractability issues prevent us from handling concave costs formally in this model, but this may be an interesting avenue for future research.

[11]If R_m is unimodal, its maximum occurs at one of the two integers closest to

$$m^* = \frac{\ln\left(\frac{\ln(\beta/\phi)}{\ln(a^2\beta^2/\phi)}\right)}{\ln(a^2\beta)}.$$

It is straightforward to show that $\frac{\partial m^*}{\partial \alpha} > 0$, $\frac{\partial m^*}{\partial \beta} > 0$ and $\frac{\partial m^*}{\partial \phi} < 0$. Moreover m^* diverges as $\phi \to \beta^+$, and may be very large when ϕ is close to β.

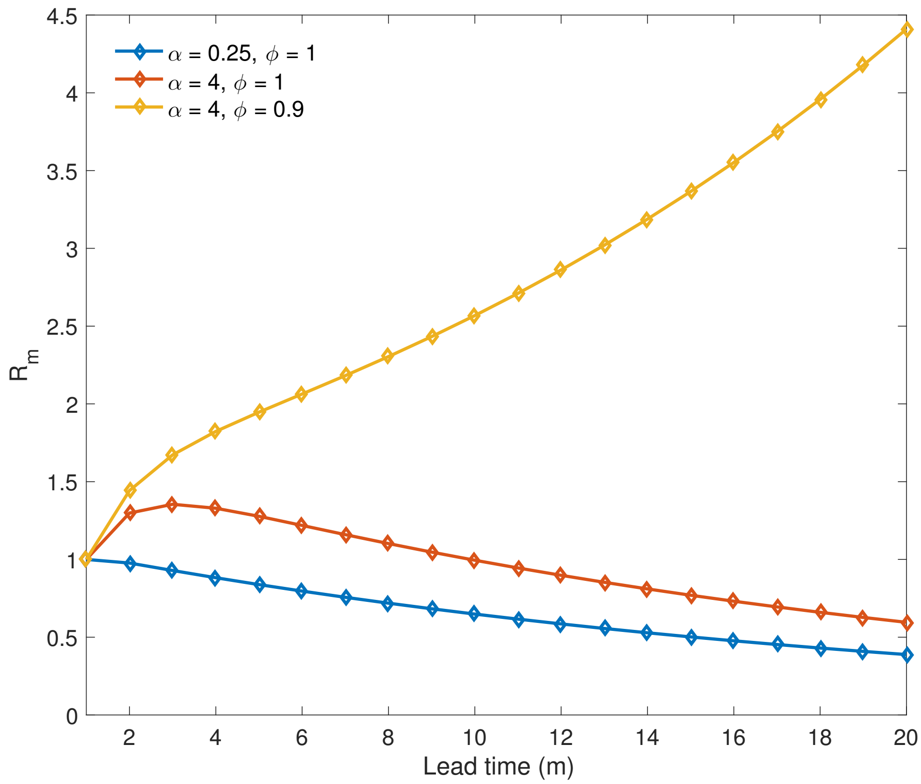


Figure 2: Typical dependence of R_m on adjustment costs (α) and prior uncertainty (ϕ). $\beta = 0.95$ in these examples.

3.2 Accounting for interactions

In this section we move beyond first-order results, aiming to summarize the dependence of $T(\vec{\tau})$ on $\vec{\tau}$ in a manner that accounts for the interactions between lead times. It is intuitively clear that these interactions are important determinants of the overall value of a prediction system. If we are able to predict events at lead time m very accurately, the value of an improvement in the predictability of events at lead time $m+1$ must surely be quite low. Indeed, inspection of the expression for $T(\vec{\tau})$ in (13) shows that for any positive integers m, k,

$$\frac{\partial T}{\partial \tau_m} > 0, \quad \frac{\partial^2 T}{\partial \tau_m \partial \tau_k} < 0.$$

Thus, $T(\vec{\tau})$ is a concave function of $\vec{\tau}$; predictabilities at different lead times are substitutes. Since $T(\vec{\tau})$ is a non-separable function of the infinite vector of parameters $\vec{\tau} = (\tau_m)_{m \geq 1}$, there is no unique way of computing the contribution of each individual τ_m to the value function. We will exploit the fact that $T(\vec{\tau})$ is concave in $\vec{\tau}$ to present a measure of the importance of different lead times that we find especially intuitive.

In order to summarize the dependence of $T(\vec{\tau})$ on its arguments we imagine that the decision-maker has a hypothetical total predictability budget $B = \sum_{m=1}^{\infty} \tau_m$, and study how she would like this budget to be allocated between lead times. Since $T(\vec{\tau})$ is concave this allocation problem has a unique solution, and the budget share allocated to each lead time captures its importance in a manner that accounts for interactions. We emphasize that the predictability budget B is a purely hypothetical construct; it does *not* represent the costs of increasing predictability, which are very likely to vary by lead time. Rather, B is merely a mathematical device that allows us to summarize the relative importance of predictability at different lead times in the value function. As in the rest of the value of information literature, our focus throughout the paper is on the benefit side of predicability.

Formally, we are interested in computing the following quantity:

$$\vec{\sigma} \equiv \frac{1}{B}\left(\text{argmax}_{\vec{\tau}}\, T(\vec{\tau}) \text{ s.t. } \sum_{m=1}^{\infty} \tau_m = B\right). \tag{21}$$

The m-th component of $\vec{\sigma}$, denoted σ_m, is the share of the total predictability budget B that the agent would like to allocate to lead time m. By definition, $\sigma_m \in [0, 1]$ for all m, and $\sum_{m=1}^{\infty} \sigma_m = 1$. Although it is not possible to solve for $\vec{\sigma}$ analytically, it is straightforward to find an arbitrarily good approximation to the solution using standard numerical constrained optimization routines.

In general $\vec{\sigma}$ depends on the vector of prior precisions $\vec{\lambda}$. To maintain consistency with the marginal analysis of the previous section, we assume that $\lambda_t = \phi^{t/2}\lambda_0$, corresponding to the $\lambda_\infty^2 \to 0$ limit of (18). The relative importance of prior beliefs and predictions in determining the expectations that enter the value function is captured by the ratio λ_0/B. To see this, notice from (13) that the optimization problem in (21) is equivalent to finding a sequence of values $\vec{\sigma}$ that maximizes

$$-\frac{1}{B}\Bigg[\left(\frac{1}{\frac{\lambda_0}{B}\phi^{1/2}+\sigma_1} + (a^2\beta^2)\frac{1}{\frac{\lambda_0}{B}\phi^{2/2}+\sigma_2} + (a^2\beta^2)^2\frac{1}{\frac{\lambda_0}{B}\phi^{3/2}+\sigma_3} + \dots\right)$$
$$+\beta\left(\frac{1}{(\frac{\lambda_0}{B}\phi^{2/2}+\sigma_2)+\sigma_1} + (a^2\beta^2)\frac{1}{(\frac{\lambda_0}{B}\phi^{3/2}+\sigma_3)+\sigma_2} + (a^2\beta^2)^2\frac{1}{(\frac{\lambda_0}{B}\phi^{4/2}+\sigma_4)+\sigma_3} + \dots\right) + \mathcal{O}(\beta^2)\Bigg],$$

subject to $\sum_m \sigma_m = 1$. When $\lambda_0/B \ll 1$, predictions are highly non-marginal relative to the prior, and dominate the value function. Interactions between forecast lead times will be most important in this case. By contrast, when $\lambda_0/B \gg 1$ predictions are marginal relative to the prior, and interactions between lead times are of second order importance. So what is a reasonable value of λ_0/B? It is clear that if we want to study substitution between lead times we must not choose λ_0/B to be too large. On the other hand, if λ_0/B is very small the prior plays no role in the analysis. This case is of interest (see below), but we would also like to investigate the role of the prior, so we cannot only choose small values for λ_0/B. In practice, we would expect forecast errors to be roughly comparable to prior uncertainty. Indeed, for many phenomena forecasts themselves are responsible for forming our priors. Our current uncertainty about climate change, for example, is intimately related to the uncertainty in scientific projections of climate change. We will thus initially work with a conservative representative value of $\lambda_0/B = 1$. Note that this implies that the sum of forecast precisions over *all* lead times is comparable in size to the precision of our beliefs about the *current* period. Almost all σ_m are thus small relative to λ_0/B. Indeed, the importance of the prior is likely overestimated in this parameterization. Appendix G contains a sensitivity analysis and detailed discussion of the cases where $\lambda_0/B \gg 1$ and $\lambda_0/B \ll 1$. Figure 3 demonstrates the typical dependence of $\vec{\sigma}$ on adjustment costs α and prior uncertainty ϕ when $\lambda_0/B = 1$.

Several features of Figure 3 deserve highlighting. First, the budget allocations in this figure tell a very different story from the marginal analysis in Figure 2. Even when $\phi < \beta$ (i.e. the top panel in Fig. 3) so that the long run is significantly more uncertain than the short run, the decision-maker would like to allocate most of her predictability budget to short lead times, i.e. 1-4 time steps ahead. By contrast, the analysis in Fig. 2 showed that the value of a marginal unit of predictability is *increasing* in lead time when $\phi < \beta$; at face value this would seem to suggest that the decision-maker should allocate her entire budget to long-run prediction. The results in Fig. 3 are very different, because $\vec{\sigma}$ accounts for the substitution possibilities between lead times that arise from sequential forecast updating. These substitution effects favour the short-run because accurate short-run forecasts can compensate for any errors in long-run forecasts once the long-run events in question come nearer to the present, in addition to providing information about current near-term conditions. If adjustment is highly costly (i.e. α is large) it is clearly more costly to react to short-run warnings, and accurate early warnings become more important. This is reflected in the fact that $\vec{\sigma}$ places larger weight on long lead times as α increases. However, perhaps surprisingly, even for very large values of α substitution effects cause the agent to allocate very small budget shares to lead times larger than 10 times steps. The presence of predictability at these shorter lead times renders long-run predictability essentially irrelevant. Fig. 3 also shows that prior uncertainty does not have nearly as large an effect on how the decision-maker would like to allocate her predictability budget as the marginal analysis suggested it might. The two panels of Fig. 3, which correspond to values of ϕ below and above β respectively, differ in their details, with lower values of ϕ (i.e. greater long-run uncertainty) giving rise to greater weight on longer lead times. But in both cases very little weight is given to long lead times. In contrast, the marginal analysis suggested that these cases should give rise to very different behaviour, with long-run forecasts always being more valuable than short-run forecasts when $\phi < \beta$. This indicates that substitution effects dominate the role of the prior in determining the budget

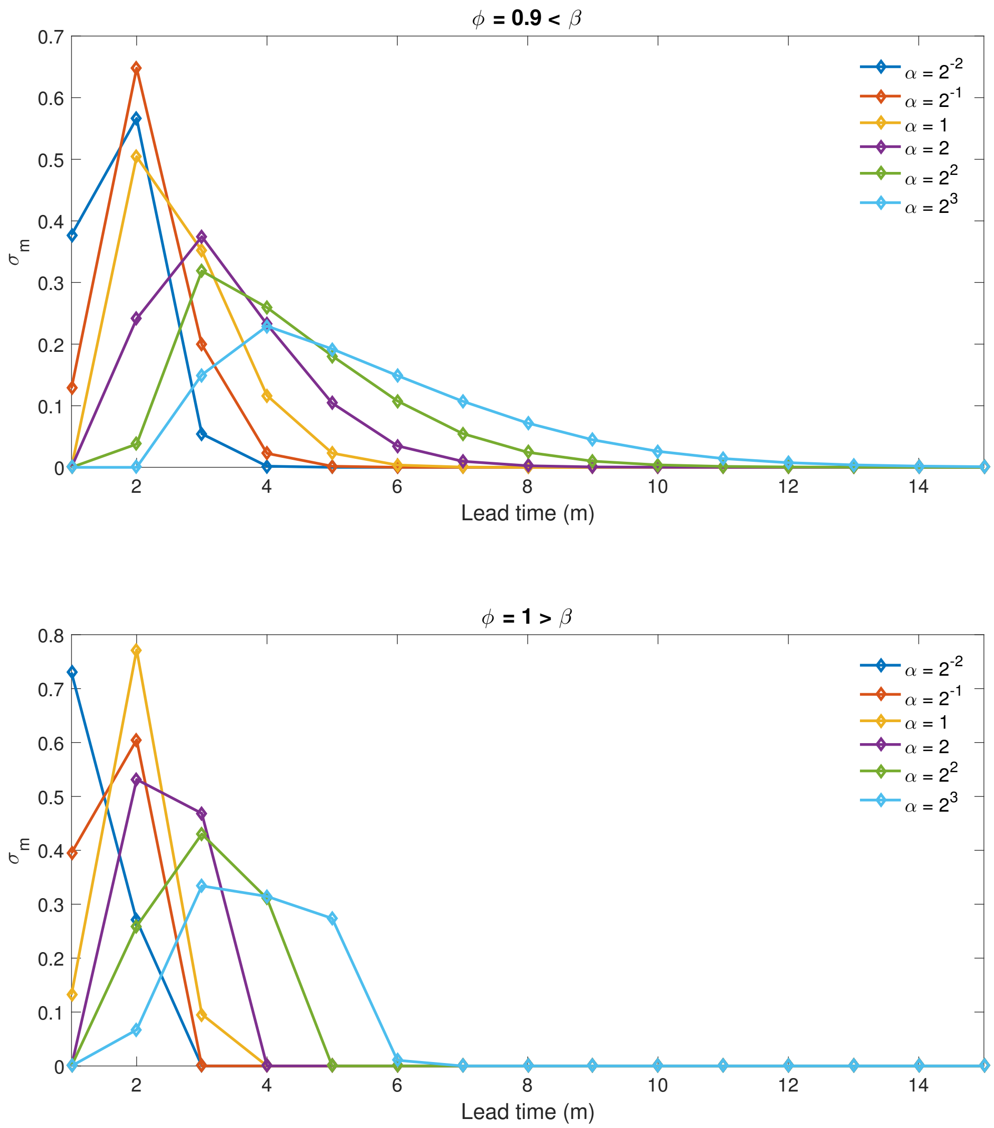


Figure 3: Budget share σ_m allocated to lead time m in the optimization problem in (21). $\beta = 0.95$, $\frac{\lambda_0}{B} = 1$.

allocation, despite the precision of the prior being comparable to or greater than that of predictions.

While Fig. 3 pertains to the baseline case $\lambda_0/B = 1$, it is also interesting to examine $\vec{\sigma}$ when $\lambda_0/B \to 0$. In this limit the agent's beliefs about all future periods are *entirely* determined by the prediction system, the prior plays no role. In this case $\vec{\sigma}$ captures the

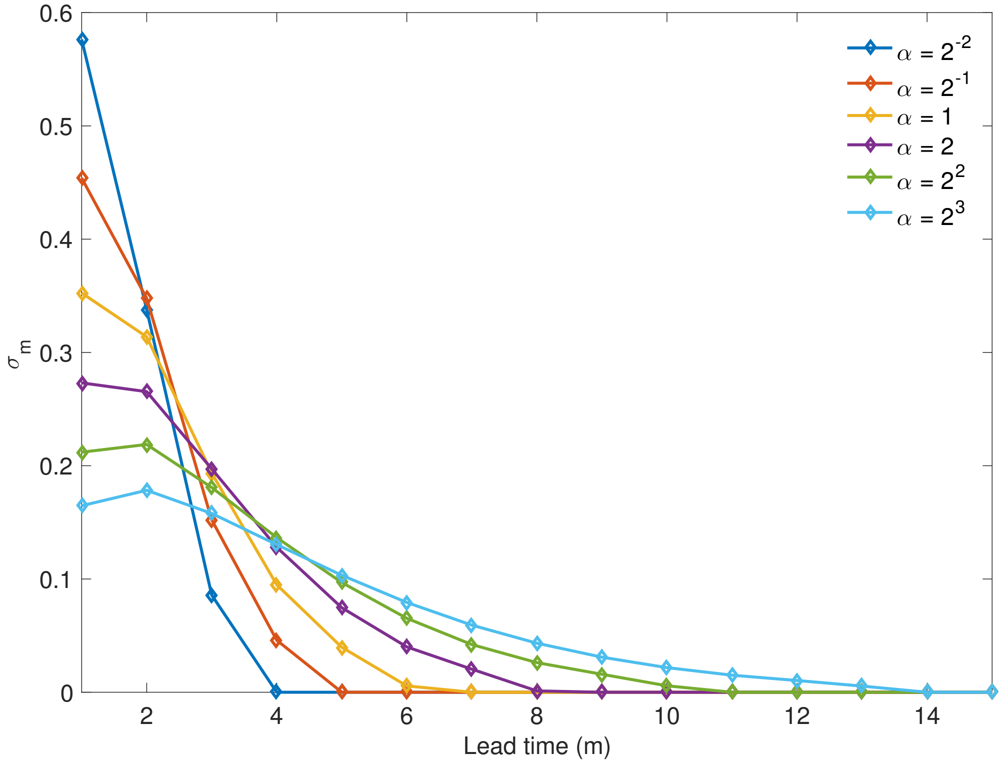


Figure 4: Budget share σ_m allocated to lead time m in the optimization problem in (21). $\beta = 0.95$, $\frac{\lambda_0}{B} \to 0$. This figure illustrates the 'pure' effect of substitution between lead times, when priors play no role in the analysis.

'pure' effects of substitution between lead times, unadulterated by the prior (which favours long lead times). Results for this case are depicted in Fig. 4. This figure clearly demonstrates how substitution between lead times leads the decision-maker to place most of his budget on short run predictability, even when adjustment costs are very large.

4 Conclusions

We have developed a simple analytical model that allows us to compute decision-makers' induced preferences over prediction systems with different profiles of accuracy as a function of lead time. Valuing prediction systems correctly requires an explicitly dynamic model that accounts for the fact that forecasts of events at different temporal distance have different accuracies, and that agents may adapt their decisions to new information as forecasts are updated over time. The essential novel feature of our model is that it disentangles the predictability of events at different temporal distances, allowing us to compute the contribution of predictive accuracy at each lead time to the overall value of a forecasting product in a simple and tractable manner. This enables a study of the relative importance of short- and long-run predictability that is, we believe, novel in the literature.

Our results point to potentially important lessons for decision-makers, and for efforts to improve the social value of forecasts. As observed in the introduction, it is not uncommon to find the presence of long-run uncertainty conflated with a need for long-run predictability in policy circles. In general however this is a logical fallacy, as it neglects decision-makers' abilities to adjust activities over time in response to updated forecasts. Our analysis suggests that if adjustments in response to sequential forecast updates are accounted for, short-run predictability is often more valuable than long-run predictability, even if adjustment costs and long-run uncertainty are large. It is perhaps surprising just how effectively short-run predictability can substitute for long-run predictability in our model, as the convexity of adjustment costs would seem to imply that accurate long-run forecasts would give rise to significant cost savings when adjustment costs are large. The fact that this result would be difficult to guess *a priori* (at least for us) points to the necessity of modeling approaches that aim to disentangle the contribution of predictions at each lead time to the overall value of forecasts. Such models could also provide forecast producers with valuable information about where they should focus their efforts at forecast improvement. While improvements in long-run predictions often require new scientific approaches that reduce model misspecification errors, short-run predictions can often be substantially improved by simply reducing measurement errors in initial conditions (i.e. increasing the quality of observations). Our results suggest that the latter activity may carry significant value for decision-makers concerned with adapting to long-run changes, even though such improvements will yield little new information about long-run conditions.

Although we believe that our model provides important conceptual insights into the determinants of rational demand for predictability at different lead times, it is clearly limited in some respects. The modeling exercise is made possible by judicious assumptions which render an otherwise impossibly complex infinite dimensional stochastic control problem solvable in closed form. We highlight two of these assumptions here.

First, the model relies on a location-independent quadratic loss function. It is clear that if some states of the world are intrinsically more valuable than others, information about these states will be of greater importance. Since our model assumes a payoff function that penalizes actions purely according to their distance from a state-dependent optimal choice, the costs of a maladapted choice do not depend on the state of the world. It is therefore best to think of our results as defining a symmetric baseline case in which the ability of the decision-maker to adapt to her environment is not state-contingent. We believe that this captures the essence of the problems we are interested in, but extensions to asymmetric loss functions would naturally be of interest, although we expect them to face analytical difficulties.

Second, as in the rest of the value of information literature, our model focuses on a decision-maker who faces an *exogenously* changing environment. Thus, its conceptual lessons apply to e.g. individuals and firms, but less to large entities whose actions may strongly affect the uncertainties in their operating environments. For example, we feel that the model is a fair abstract representation of the problem of adapting to climate change at the local level, but *not* of mitigating climate change at the global level. In the latter case actions the world takes to reduce greenhouse gas emissions clearly affect uncertainties, whereas in the former any small country or firm may reasonably take changes in the climate as exogenous to its own activities.

References

K. J. Arrow (1991). '"I Know a Hawk From a Handsaw"'. In M. Szenberg (ed.), *Eminent Economists: Their Life Philosophies*. Cambridge University Press, Cambridge England ; New York.

K. J. Arrow & A. C. Fisher (1974). 'Environmental Preservation, Uncertainty, and Irreversibility'. *The Quarterly Journal of Economics* **88**(2):312–319.

L. M. Benveniste & J. A. Scheinkman (1979). 'On the Differentiability of the Value Function in Dynamic Models of Economics'. *Econometrica* **47**(3):727–732.

D. Blackwell (1953). 'Equivalent Comparisons of Experiments'. *The Annals of Mathematical Statistics* **24**(2):265–272.

M. P. Clements (1997). 'Evaluating the Rationality of Fixed-event Forecasts'. *Journal of Forecasting* **16**(4):225–239.

C. Costello, et al. (1998). 'The Value of El Niño Forecasts in the Management of Salmon: A Stochastic Dynamic Assessment'. *American Journal of Agricultural Economics* **80**(4):765–777.

C. Costello, et al. (2001). 'Renewable resource management with environmental prediction'. *Canadian Journal of Economics* **34**(1):196–211.

M. DeGroot (1970). *Optimal Statistical Decisions*. McGraw-Hille, New York, WCL edition edn.

S. Dessai, et al. (2009). 'Climate prediction: A limit to adaptation?'. In W.N. Adger, I. Lorenzoni, & K. O'Brien (eds.), *Adapting to Climate Change: Thresholds, Values, Governance*, pp. 64–78. Cambridge University Press.

L. G. Epstein (1980). 'Decision Making and the Temporal Resolution of Uncertainty'. *International Economic Review* **21**(2):269–283.

S. Fischer & O. Blanchard (1989). *Lectures on Macroeconomics*. MIT Press.

H.-M. Füssel (2007). 'Vulnerability: A generally applicable conceptual framework for climate change research'. *Global Environmental Change* **17**(2):155–167.

C. W. J. Granger & Y. Jeon (2007). 'Long-term forecasting and evaluation'. *International Journal of Forecasting* **23**(4):539–551.

D. W. Grupe & J. B. Nitschke (2013). 'Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective'. *Nature Reviews Neuroscience* **14**(7):488–501.

R. Lee (2011). 'The outlook for Population Growth'. *Science* **333**:569–573.

T. Lindh (2011). 'Long-horizon growth forecasting and demography'. In M. P. Clements & D. F. Hendry (eds.), *Oxford handbook of economic forecasting*. Oxford University Press.

L. Ljungqvist & T. J. Sargent (2004). *Recursive Macroeconomic Theory*. MIT Press, Cambridge, Mass, 2nd edn.

J. Marschak & K. Miyasawa (1968). 'Economic Comparability of Information Systems'. *International Economic Review* **9**(2):137–174.

R. C. Merton (1971). 'Optimum consumption and portfolio rules in a continuous-time model'. *Journal of Economic Theory* **3**(4):373–413.

M. F. Myers, et al. (2000). 'Forecasting disease risk for increased epidemic preparedness in public health'. *Advances in Parasitology* **47**:309–330.

National Academy of Sciences (2016). *Next genertaion earth system prediction: Strategies for subseasonal to seasonal foreacasts*. doi:https://doi.org/10.17226/21873. The National Academies Press, Washington D.C.

T. Palmer & R. Hagedorn (eds.) (2006). *Predictability of Weather and Climate*. Cambridge University Press.

T. J. Sargent (1978). 'Estimation of Dynamic Labor Demand Schedules under Rational Expectations'. *Journal of Political Economy* **86**(6):1009–1044.

R. Selten (1998). 'Axiomatic Characterization of the Quadratic Scoring Rule'. *Experimental Economics* **1**(1):43–61.

M. Yusuf (2009). 'Prediction Proliferation: The History of the Future of Nuclear Weapons'. Tech. rep., Brookings Institution Policy Paper no. 11.

Appendix

A A microeconomic interpretation of the model

Here we provide an interpretation of our model in a familiar microeconomic setting. Consider a competitive firm that produces a quantity q_t of product at time t, and faces linear marginal costs of production $C'(q) = c_0 + c_1 q$. The firm faces uncertain prices $(p_t)_{t\geq 0}$ in the future, and quadratic adjustment costs $k(q_t - q_{t-1})^2$. These costs are a reduced form representation of costs sustained due to rejigging operations at the intensive and extensive margins. Small changes in production are handled at the intensive margin, and are thus cheap – current employees work longer/shorter hours, installed capital is used more/less intensively, and orders from existing suppliers are tweaked to meet small fluctuations in demand. However, large changes in production require extensive margin changes – large numbers of new employees must be hired/fired, new machinery must be bought or rented, and new suppliers found and terms negotiated. The form of our adjustment costs amounts to assuming that extensive margin adjustments are cheaper if done in a planned sequence of steps, rather than in an abrupt transition. There are several reasons why this might be. With early warning the firm could find creative ways of adapting its existing resources to new market conditions. Early warning may also place the firm in a stronger negotiating position with respect to employment and supply contracts (because of the lack of urgency), and could reduce the opportunity costs associated with under/over capacity. Quadratic adjustment costs are an analytically convenient reduced form way of representing these inertial forces on adjustment.

Since the firm takes prices p_t as given, its instantaneous profit function can be written as:

$$\begin{aligned}\Pi_t &= p_t q_t - (c_0 q_t + \frac{1}{2}c_1 q_t^2) - k(q_t - q_{t-1})^2 \\ &= -\frac{c_1}{2}(q_t - \theta_t)^2 - k(q_t - q_{t-1})^2 + M_t\end{aligned}$$

where $\theta_t = (p_t - c_0)/c_1$, and M_t is a decision irrelevant constant, which may be neglected when computing the value of information about the sequence of values $(\theta_t)_{t\geq 0}$. Thus the firm's profit function is of the form (2), up to an irrelevant factor of c_1.

B Proof of Proposition 1

We use the Bellman equation (10) to solve for the optimal policy function $X_{n+1} = \pi(X_n, Y_n)$. When referring to functions and operations on functions, we will adopt a notation in which primed variables denote next period quantities, and unprimed variables denote current period quantities, i.e. $W = W(X', X, Y)$ and $V = V(X, Y)$. So $\frac{\partial W}{\partial X'}$, for example, refers to the function whose value is the partial derivative of W with respect to its first argument, i.e. the *next period* value of X. When we evaluate functions and their derivatives at specific times, we will still use e.g. X_n, X_{n+1} to denote function arguments. Thus $\frac{\partial W}{\partial X'}(X_{n+1}, X_n, Y_n)$ is the partial derivative of W with respect to its first argument, evaluated at (X_{n+1}, X_n, Y_n). With this notation, the first order condition for X_{n+1} is

$$\frac{\partial W}{\partial X'}(\pi(X_n, Y_n), X_n, Y_n) + \beta \int \frac{\partial V}{\partial X}(\pi(X_n, Y_n), F(Y_n, S))Q(S; Y_n)dS = 0. \tag{22}$$

By the envelope theorem,

$$\frac{\partial V}{\partial X}(X_n, Y_n) = \frac{\partial W}{\partial X}(\pi(X_n, Y_n), X_n, Y_n) \tag{23}$$

From (2), and (23) evaluated at time $n + 1$, we have

$$\begin{aligned}
\frac{\partial W}{\partial X'}(\pi(X_n, Y_n), X_n, Y_n) &= \mu_0 + \alpha X_n - (1 + \alpha)\pi(X_n, Y_n) \\
\frac{\partial V}{\partial X}(\pi(X_n, Y_n), F(Y_n, S)) &= \frac{\partial W}{\partial X}(\pi(\pi(X_n, Y_n), F(Y_n, S)), \pi(X_n, Y_n), F(Y_n, S)) \\
&= \alpha(X' - X)\big|_{X'=\pi(\pi(X_n,Y_n),F(Y_n,S)),X=\pi(X_n,Y_n)} \\
&= \alpha(\pi(\pi(X_n, Y_n), F(Y_n, S)) - \pi(X_n, Y_n)).
\end{aligned}$$

Substituting into (22), we find that the policy rule must satisfy

$$\mu_0 + \alpha X_n - (1 + \alpha)\pi(X_n, Y_n) + \beta \int [\alpha(\pi(\pi(X_n, Y_n), F(Y_n, S)) - \pi(X_n, Y_n))]\, Q(S; Y_n)dS = 0. \tag{24}$$

We solve this equation by the 'guess and verify' method. The certainty equivalence property of the quadratic control problem suggests that we should look for a control rule of the form

$$\pi(X, Y) = aX + \sum_{t=0}^{\infty} b_t \mu_t$$

where the coefficients $(a, (b_t)_{t\geq 0})$ are to be determined. Plugging this guess into (24), and now suppressing the index n, we find:

$$\begin{aligned}
&[\mu_0 + \alpha X - (1 + \alpha)(aX + \sum_{t=0}^{\infty} b_t\mu_t)]+ \\
&\beta\alpha \left[\int \left(a(aX + \sum_{t=0}^{\infty} b_t\mu_t) + \sum_{t=0}^{\infty} b_t\mu_t'(s_{t+1}) - (aX + \sum_{t=0}^{\infty} b_t\mu_t)\right) Q(S, Y)dS\right] = 0
\end{aligned}$$

where $\mu_t'(s_{t+1})$ is the next period value of μ_t conditional on receiving a signal s_{t+1}, given by (5). Since $E_{s_t}\mu_t'(s_{t+1}) = \mu_{t+1}$, we can simplify this to:

$$\begin{aligned}
&\mu_0 + \alpha X - (1+\alpha)(aX + \sum_{t=0}^{\infty} b_t\mu_t) + \\
&\beta[\alpha a^2 X + a\alpha \sum_{t=0}^{\infty} b_t\mu_t + \alpha \sum_{t=0}^{\infty} b_t\mu_{t+1} - a\alpha X - \alpha \sum_t b_t\mu_t] = 0.
\end{aligned}$$

Since this equation must hold for all values of X, μ_t, we must equate the coefficients of each state variable to zero. The equation for the coefficient of X is:

$$\alpha\beta a^2 - (1 + \alpha(1+\beta))a + \alpha = 0 \tag{25}$$

$$\Rightarrow a = \frac{1 + \alpha(1+\beta) \pm \sqrt{(1+\alpha(1+\beta))^2 - 4\alpha^2\beta}}{2\alpha\beta} \tag{26}$$

To pick the correct root, note that if $\alpha \to 0$, the policy rule should reduce to

$$\pi(X, Y) = \mu_0.$$

This follows since when adjustment is costless, the optimal policy simply maximizes period payoffs. For the positive root we have

$$\lim_{\alpha\to 0} a(\alpha) \to \infty,$$

thus giving incorrect behaviour. By contrast, we show below that the correct behaviour is obtained if we select the negative root. Thus we conclude that

$$a = a(\alpha, \beta) = \frac{1 + \alpha(1+\beta) - \sqrt{(1+\alpha(1+\beta))^2 - 4\alpha^2\beta}}{2\alpha\beta} \tag{27}$$

The equation for b_0 is:

$$1 - (1+\alpha)b_0 + a\beta\alpha b_0 - \alpha\beta b_0 = 0$$

$$\Rightarrow b_0 = \frac{1}{1 + \alpha + \alpha\beta(1-a)}. \tag{28}$$

For $t \geq 1$, the equation for b_t is:

$$\begin{aligned}
&-(1+\alpha)b_t + a\beta\alpha b_t + \alpha\beta b_{t-1} - \alpha\beta b_t = 0 \\
\Rightarrow &b_t = \frac{\alpha\beta}{1 + \alpha + \alpha\beta(1-a)} b_{t-1}.
\end{aligned}$$

Thus for all $t \geq 0$,

$$b_t = \frac{1}{1+\alpha+\alpha\beta(1-a)} \left[\frac{\alpha\beta}{1+\alpha+\alpha\beta(1-a)}\right]^t \tag{29}$$

We can simplify this further by using the equation for a in (25). Define

$$\Lambda \equiv 1+\alpha+\alpha\beta(1-a) \tag{30}$$

From (25) we have

$$(\alpha\beta)a^2 - (1+\alpha(1+\beta))a + \alpha = 0$$

Now

$$\begin{aligned} &1+\alpha(1+\beta) = \Lambda + \alpha\beta a \\ &\Rightarrow (\alpha\beta)a^2 - (\Lambda+\alpha\beta a)a + \alpha = 0 \\ &\Rightarrow \Lambda = \frac{\alpha}{a}. \end{aligned}$$

Thus

$$b_t = \frac{a}{\alpha}\left(a\beta\right)^t. \tag{31}$$

We now prove the properties of the coefficients a, b_t, stated below the proposition:

1. $\lim_{\alpha\to 0} a(\alpha,\beta) = 0$

   Use l'Hopital's rule: differentiate the numerator and denominator of a with respect to α, and evaluate the limit of each as $\alpha \to 0$:

$$\begin{aligned} \lim_{\alpha\to 0} a(\alpha,\beta) &= \frac{1+\beta-\frac{1}{2\times 1}(2\times 1\times(1+\beta)-0)}{2\beta} \\ &= 0. \end{aligned}$$

2. $\lim_{\alpha\to\infty} a(\alpha,\beta) = 1$:

$$\begin{aligned} \lim_{\alpha\to\infty} a(\alpha,\beta) &= \frac{1+\beta}{2\beta} - \frac{1-\beta}{2\beta} \\ &= 1. \end{aligned}$$

3. $\frac{\partial a}{\partial \alpha} > 0$.

   From (27) we have

$$\frac{\partial a}{\partial \alpha} = -\frac{1}{2}\frac{-\alpha\beta+\sqrt{\alpha^2(1-\beta)^2+2\alpha(1+\beta)+1}-\alpha-1}{\alpha^2\beta\sqrt{\alpha^2(1-\beta)^2+2\alpha(1+\beta)+1}}. \tag{32}$$

Hence, $\frac{\partial a}{\partial \alpha} a > 0$ iff

$$\begin{aligned} & -\alpha\beta + \sqrt{\alpha^2(1-\beta)^2 + 2\alpha(1+\beta)+1} - \alpha - 1 < 0 \\ \iff & \sqrt{\alpha^2(1-\beta)^2 + 2\alpha(1+\beta)+1} < 1 + \alpha + \alpha\beta \\ \iff & \alpha^2(1-\beta)^2 + 2\alpha(1+\beta) + 1 < \alpha^2(1+\beta)^2 + 2\alpha(1+\beta) + 1 \end{aligned}$$

which is obviously satisfied for all $\alpha > 0, \beta \in (0,1)$.

4. $a + \sum_{t=0}^{\infty} b_t = 1$.

   From the previous calculations we know that $a \in [0,1] \Rightarrow a\beta \in [0,1]$. It follows from (31) that

$$\begin{aligned} a + \sum_{t=0}^{\infty} b_t - 1 &= a + \frac{a}{\alpha}\frac{1}{1-a\beta} - 1 \\ &= \frac{-\alpha\beta a^2 + a(1+\alpha(1+\beta)) - \alpha}{\alpha(1-a\beta)} \\ &= 0 \end{aligned}$$

   where the last equality follows from the defining equation for a in (25).

5. $\frac{\partial}{\partial \alpha}(b_{t+1}/b_t) > 0$, $\frac{\partial b_0}{\partial \alpha} < 0$.

   Since $a + \sum_{t=0}^{\infty} b_t = 1$, and a is increasing in α, we know that $\sum_{t=0}^{\infty} b_t$ must be decreasing in α. From (31) we see that

$$\frac{b_{t+1}}{b_t} = a\beta$$

   and hence this ratio is increasing in α. Since b_t declines more slowly as α increases, it must be the case that $\frac{\partial b_0}{\partial \alpha} < 0$ in order to ensure that $\sum_{t=0}^{\infty} b_t$ is decreasing in α.

C Illustration of the dependence of optimal policies on adjustment costs α.

To illustrate how the adjustment cost parameter α affects decisions quantitatively, consider a deterministic version of the model in which the values $\tilde{\theta}_n$ are chosen be a fixed sequence of draws from an arbitrary univariate random variable with finite variance σ^2. When $\alpha = 0$, optimal decisions coincide with the current value of $\tilde{\theta}_n$, i.e. $X_n = \mu_0 = \tilde{\theta}_n$ for all n. As α increases, adjustment becomes more costly, and the values of X_n fluctuate less than $\tilde{\theta}_n$ itself.

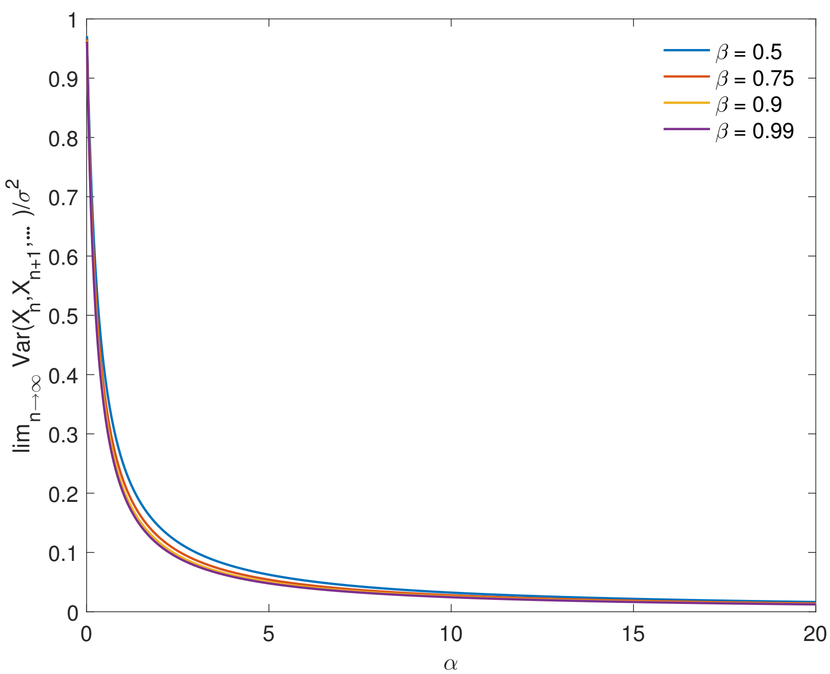


Figure 5: Asymptotic variability of the decision variable X relative to the variability of the loss-minimizing decisions $\tilde{\theta}$, assuming that the values of $\tilde{\theta}_n$ are deterministic and given by a fixed sequence of draws from a random variable with variance σ^2.

Using the formula (11) and some simple ergodic arguments one can show that

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Var}(X_n, X_{n+1}, \ldots) &= \frac{\sum_{t=0}^{\infty} b_t^2}{1-a^2}\sigma^2 \\ &= \left[\left(\frac{a}{\alpha}\right)^2 \frac{1}{(1-a^2)(1-a^2\beta^2)}\right]\sigma^2\end{aligned}$$

for arbitrary initial condition X_0. Figure 5 plots the asymptotic variance of the sequence of decisions as a function of α for several β. The figure illustrates how α controls the magnitude of the adjustments the decision-maker makes to adapt to fluctuations in a stationary environment. For a wide range of β, $\alpha > 1.5$ implies that the decision maker adjusts to less than 20% of the variability in $\tilde{\theta}$, and $\alpha > 3$ implies adjustment to less than 10% of the variability. In addition, changes in α have a greater effect on behaviour when α is small (e.g. $\alpha < 1$) than when it is large.

D Proof of Proposition 2

As in the derivation of the optimal policy function, we use the 'guess and verify' method. Begin by guessing that the value function has the form

$$V(X,Y) = kX^2 + \sum_{t=0}^{\infty} c_t \mu_t X + \sum_{t=0}^{\infty} \sum_{p=t+1}^{\infty} D_{t,p} \mu_t \mu_p + \sum_{t=0}^{\infty} d_t \mu_t^2 + \sum_{t=0}^{\infty} \sum_{i=0}^{\infty} \frac{f_{i,t}}{\lambda_t + h_{i,t}}. \tag{33}$$

All except the last term of this expression are straightforward to guess simply by inspection of the formula for the period payoff in (2). The last term will however be the most important, as it will turn out that this is the only term that depends on the precision sequence $\vec{\tau} = (\tau_t)_{t \geq 1}$.

Consider the quadratic terms in this guess of the form $\mu_t \mu_p$. We are going to need to know how these will transform under the updating rule (5) and after the expectation over signal realizations has been applied. Letting a prime denote the next period value of a variable, we are interested in computing expectations of the form

$$\mathbf{E}_S \mu_t' \mu_p' = \mathbf{E}_{s_{t+1}, s_{p+1}} \mu_t'(s_{t+1}) \mu_p'(s_{p+1})$$

where signals are distributed according to the agents' current posterior predictive distribution, given by (7). Recall that

$$\mu_t'(s_{t+1}) = \frac{\tau_{t+1}}{\tau_{t+1} + \lambda_{t+1}} s_{t+1} + \frac{\lambda_{t+1}}{\tau_{t+1} + \lambda_{t+1}} \mu_{t+1}$$

When $t \neq p$, we can immediately write down the answer, as means are martingales, and signals are independent:

$$\mathbf{E}_{s_{t+1}, s_{p+1}} \mu_t'(s_{t+1}) \mu_p'(s_{p+1}) = \mu_{t+1} \mu_{p+1}$$

For $t = p$ however, things are different:

$$\mathbf{E}_{s_{t+1}} \mu_t'(s_{t+1}) \mu_t'(s_{t+1}) = \mathbf{E}_{s_{t+1}} \left[\frac{\tau_{t+1}}{\tau_{t+1} + \lambda_{t+1}} s_{t+1} + \frac{\lambda_{t+1}}{\tau_{t+1} + \lambda_{t+1}} \mu_{t+1} \right]^2$$

Consider the quadratic term in s_{t+1} in this expression:

$$\begin{aligned}
\mathbf{E}_{s_{t+1}} \left(\frac{\tau_{t+1}}{\tau_{t+1} + \lambda_{t+1}} \right)^2 s_{t+1}^2 &= \left(\frac{\tau_{t+1}}{\tau_{t+1} + \lambda_{t+1}} \right)^2 [\mathrm{Var}(s_{t+1}) + \mu_{t+1}^2] \\
&= \left(\frac{\tau_{t+1}}{\tau_{t+1} + \lambda_{t+1}} \right)^2 [\frac{\lambda_{t+1} + \tau_{t+1}}{\lambda_{t+1} \tau_{t+1}} + \mu_{t+1}^2] \\
&= \frac{\tau_{t+1}}{\lambda_{t+1}(\lambda_{t+1} + \tau_{t+1})} + \left(\frac{\tau_{t+1}}{\tau_{t+1} + \lambda_{t+1}} \right)^2 \mu_{t+1}^2
\end{aligned}$$

When we combine this expression with the other terms in the expression for $\mathbf{E}_{s_{t+1}}\mu'_t(s_{t+1})\mu'_t(s_{t+1})$, the factor in front of μ^2_{t+1} in the second term will cancel to 1 (as occurs in the case $t \neq p$), and we are left with

$$\mathbf{E}_{s_{t+1}}\mu'_t(s_{t+1})\mu'_t(s_{t+1}) = \frac{\tau_{t+1}}{\lambda_{t+1}(\lambda_{t+1}+\tau_{t+1})} + \mu^2_{t+1}. \tag{34}$$

Hence, in summary:

$$\mathbf{E}_{s_{t+1},s_{p+1}}\mu'_t(s_{t+1})\mu'_p(s_{p+1}) = \begin{cases} \mu_{t+1}\mu_{p+1} & t \neq p \\ \frac{\tau_{t+1}}{\lambda_{t+1}(\lambda_{t+1}+\tau_{t+1})} + \mu^2_{t+1} & t = p. \end{cases} \tag{35}$$

It will be more convenient in what follows to write the terms that depend on λ_{t+1} is this expression as

$$\frac{\tau_{t+1}}{\lambda_{t+1}(\lambda_{t+1}+\tau_{t+1})} = \frac{1}{\lambda_{t+1}} - \frac{1}{\lambda_{t+1}+\tau_{t+1}}. \tag{36}$$

We now want to write down the Bellman equation for our assumed functional form for the value function. The first step is to compute the period payoff:

$$\begin{aligned}
W(\pi(X,Y),X,Y) &= -\frac{1}{2}\left[(1+\alpha)[\pi(X,Y)]^2 + \alpha X^2 - 2\pi(X,Y)(\mu_0+\alpha X) + \frac{1}{\lambda_0} + (\mu_0)^2\right] \\
&= -\frac{1}{2}\left[(1+\alpha)(aX + \sum_{t=0}^{\infty} b_t\mu_t)^2 + \alpha X^2 - 2(aX + \sum_{t=0}^{\infty} b_t\mu_t)(\mu_0+\alpha X) + \frac{1}{\lambda_0} + (\mu_0)^2\right] \\
&= -\frac{1}{2}\Bigg[(1+\alpha)(a^2X^2 + 2aX\sum_{t=0}^{\infty} b_t\mu_t + \sum_{t=0}^{\infty}\sum_{p=t+1}^{\infty} 2b_tb_p\mu_t\mu_p + \sum_{t=0}^{\infty} b_t^2\mu_t^2) + \alpha X^2 \\
&\quad -2aX\mu_0 - 2a\alpha X^2 - 2\mu_0\sum_{t=0}^{\infty} b_t\mu_t - 2\alpha X\sum_{t=0}^{\infty} b_t\mu_t + (\mu_0)^2 + \frac{1}{\lambda_0}\Bigg]
\end{aligned}$$

We also have

$$\begin{aligned}
\mathbf{E}_S V(\pi(X,Y),F(Y,S)) &= \mathbf{E}_S\Bigg[k(\pi(X,Y))^2 + \sum_{t=0}^{\infty} c_t\mu'_t(s_{t+1})\pi(X,Y) + \sum_{t=0}^{\infty}\sum_{p=t+1}^{\infty} D_{t,p}\mu'_t(s_{t+1})\mu'_p(s_{p+1}) \\
&\quad + \sum_{t=0}^{\infty} d_t(\mu'_t(s_{t+1}))^2 + \sum_{i=0}^{\infty}\sum_{t=0}^{\infty}\frac{f_{i,t}}{\lambda'_t + h_{i,t}}\Bigg] \\
&= k[aX + \sum_t b_t\mu_t]^2 + \sum_{t=0}^{\infty} c_t\mu_{t+1}[aX + \sum_{p=0}^{\infty} b_p\mu_p] + \sum_{t=0}^{\infty}\sum_{p=t+1}^{\infty} D_{t,p}\mu_{t+1}\mu_{p+1} \\
&\quad + \sum_{t=0}^{\infty} d_t(\mu_{t+1})^2 + \sum_{t=0}^{\infty} d_t\left[\frac{1}{\lambda_{t+1}} - \frac{1}{\lambda_{t+1}+\tau_{t+1}}\right] + \sum_{i=0}^{\infty}\sum_{t=0}^{\infty}\frac{f_{i,t}}{\lambda_{t+1}+\tau_{t+1}+h_{i,t}}
\end{aligned}$$

We now have expressions for each of the three terms $V(X,Y)$, $W(\pi(X,Y),X,Y)$, $\mathbf{E}_S V(\pi(X,Y),F(Y,S))$, and must choose the free coefficients of the value function so that

$$V(X,Y) = W(\pi(X,Y),X,Y) + \beta \mathbf{E}_S V(\pi(X,Y),F(Y,S))$$

holds as an identity. We begin by focussing on the terms that depend on λ_t. If we focus just on these terms, the Bellman equation reads

$$\sum_{i=0}^{\infty}\sum_{t=0}^{\infty}\frac{f_{i,t}}{\lambda_t + h_{i,t}} = -\frac{1}{2}\frac{1}{\lambda_0} + \beta\left(\sum_{t=0}^{\infty} d_t\left[\frac{1}{\lambda_{t+1}} - \frac{1}{\lambda_{t+1}+\tau_{t+1}}\right] + \sum_{i=0}^{\infty}\sum_{t=0}^{\infty}\frac{f_{i,t}}{\lambda_{t+1}+\tau_{t+1}+h_{i,t}}\right) \tag{37}$$

We must determine values for the sequences $f_{i,t}, h_{i,t}$ such that this equation holds as an identity. Since the right hand side of this equation contains terms of the form $1/\lambda_t$ for all t, we must have terms of this form on the left hand side as well. We thus begin by choosing

$$h_{0,t} = 0$$

for all $t \geq 0$. Then if (37) is to hold as an identity for all λ_t, τ_t we require

$$f_{0,0} = -\frac{1}{2} \tag{38}$$

$$f_{0,t} = \beta d_{t-1} \text{ for } t \geq 1. \tag{39}$$

Notice that setting $h_{0,t} = 0$ creates an imbalance of terms of the form

$$\sum_{t=0}^{\infty}\frac{f_{0,t}}{\lambda_{t+1}+\tau_{t+1}}$$

on the right hand side of the Bellman equation through the last term in (37). To correct this imbalance through terms on the left hand side, we must choose

$$h_{1,t} = \tau_t$$

implying in turn that we must choose

$$f_{1,0} = 0$$

$$f_{1,t} = \beta[-d_{t-1} + f_{0,t-1}] \text{ for } t \geq 1.$$

Again we create an imbalance of terms on the right hand side, which we correct by picking

$$h_{2,t} = \tau_t + h_{1,t-1} = \tau_t + \tau_{t-1}$$

and we find that

$$f_{2,0} = 0$$
$$f_{2,t} = \beta f_{1,t-1}.$$

We can complete this imbalance/rebalance procedure indefinitely to solve for all the coefficients $f_{i,t}, h_{i,t}$. We find:

$$h_{0,t} = 0; \quad h_{i,t} = \tau_t + h_{i-1,t-1} \qquad i \geq 1 \tag{40}$$
$$f_{0,0} = -\frac{1}{2}; \quad f_{0,t} = \beta d_{t-1} \qquad t \geq 1 \tag{41}$$
$$f_{1,0} = 0; \quad f_{1,t} = \beta[-d_{t-1} + f_{0,t-1}] \qquad t \geq 1 \tag{42}$$
$$f_{i,0} = 0; \quad f_{i,t} = \beta f_{i-1,t-1}. \qquad n \geq 2, t \geq 1. \tag{43}$$

It is straightforward to solve the set of recurrence relations for $f_{i,t}$. It is convenient to write the solution as an infinite dimensional matrix:

$$\mathbf{f} = \begin{pmatrix} -\frac{1}{2} & \beta d_0 & \beta d_1 & \beta d_2 & \beta d_3 & \ldots \\ 0 & -\beta(d_0 + \frac{1}{2}) & \beta(\beta d_0 - d_1) & \beta(\beta d_1 - d_2) & \beta(\beta d_2 - d_3) & \ldots \\ 0 & 0 & -\beta^2(d_0 + \frac{1}{2}) & \beta^2(\beta d_0 - d_1) & \beta^2(\beta d_1 - d_2) & \ldots \\ 0 & 0 & 0 & -\beta^3(d_0 + \frac{1}{2}) & \beta^3(\beta d_0 - d_1) & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ldots \end{pmatrix} \tag{44}$$

The i, t entry of this matrix corresponds to $f_{i-1,t-1}$, i.e. the rows correspond to fixed values of i, and the columns to fixed values of t, both starting at zero.[12]

Clearly $f_{i,t} = 0$ for any $i > t$. Thus the only parameters $h_{i,t}$ that are relevant have

[12] Notice that $\sum_{i=0}^{\infty} f_{i,t} = -\frac{1}{2}\beta^t$. To understand this suppose that $\tau_t = 0$ for all t, i.e. the agent receives no forecasts. Then her beliefs will not change over time, and the variance of her beliefs about $\tilde{\theta}_{n+t}$ will be the same once time $n + t$ rolls around as they are in the current period n. The contribution of the variance terms to the value function in this case is thus straightforward to compute, since variance terms only enter the period payoff through the term $-\frac{1}{2}\lambda_0$. Thus, when $\tau_t = 0$, we should expect the following term in the value function: $-\frac{1}{2}\sum_{t=0}^{\infty}\beta^t \frac{1}{\lambda_t}$. Now when $\tau_t = 0$ for all t, we have

$$\sum_{i=0}^{\infty}\sum_{t=0}^{\infty} \frac{f_{i,t}}{\lambda_t + h_{i,t}} = \sum_{t=0}^{\infty} \frac{\sum_{i=0}^{\infty} f_{i,t}}{\lambda_t}$$
$$= -\frac{1}{2}\sum_{t=0}^{\infty} \beta^t \frac{1}{\lambda_t}$$

as expected.

$0 \le i \le t$. It is straightforward to solve the recurrence relation (40) to find

$$h_{0,t} = 0$$
$$h_{i,t} = \sum_{k=t+1-i}^{t} \tau_k,\ 1 \le i \le t$$

The matrix $\mathbf{f}$ makes it clear that we will need to understand the parameters d_t if we are to solve for $f_{i,t}$. We can find these parameters by solving the μ_t^2 terms of the Bellman equation. Define

$$\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \ne j \end{cases} \tag{45}$$

Then the Bellman equation for the μ_t^2 terms yields

$$\begin{aligned} d_t &= -\frac{1}{2}[(\alpha+1)(b_t)^2 + (1-2b_0)\delta_{t,0}] + \beta(k(b_t)^2 + c_{t-1}b_t(1-\delta_{t,0}) + d_{t-1}(1-\delta_{t,0}) \\ &= \left(k\beta - \frac{1}{2}(\alpha+1)\right)(b_t)^2 - \frac{1}{2}(1-2b_0)\delta_{t,0} + \beta c_{t-1}b_t + \beta d_{t-1} \end{aligned} \tag{46}$$

where $d_{-1} \equiv 0 \equiv c_{-1}$. This equation in turn depends on the coefficients of X^2 and $\mu_t X$, i.e. k and c_t. The X^2 terms of the Bellman equation give:

$$\begin{aligned} k &= -\frac{1}{2}((1+\alpha)a^2 + \alpha - 2a\alpha) + \beta(ka^2) \\ \Rightarrow k &= -\frac{1}{2}\left(\frac{(1+\alpha)a^2 + \alpha - 2a\alpha}{1-\beta a^2}\right), \end{aligned} \tag{47}$$

which is a known quantity. As a check, another way to compute k is to use the envelope theorem result:

$$\begin{aligned} \frac{\partial V}{\partial X} &= \alpha(\pi(X,Y) - X) \\ &= \alpha((a-1)X + \sum_t b_t\mu_t) \end{aligned}$$

Integrating this, we should find that

$$k = \alpha\frac{a-1}{2}.$$

Using (25) it can be shown that these two formulae for k agree, and we thus use the second, simpler, expression.

Equating coefficients of the $\mu_t X$ terms in the Bellman equation gives:

$$\begin{aligned}c_t &= -\frac{1}{2}((1+\alpha)2ab_t - 2a\delta_{t,0} - 2\alpha b_t) + \beta(2kab_t + ac_{t-1}(1-\delta_{t,0}))\\ &= (\alpha - a(1+\alpha) + 2\beta ka)b_t + a\delta_{t,0} + a\beta c_{t-1}\end{aligned}$$

Consider the factor in front of b_t in this expression. Substituting $k = \frac{\alpha}{2}(a-1)$ into this factor we see that it is equal to

$$\alpha\beta a^2 - a(1+\alpha(1+\beta)) + \alpha$$

But from the definition of a in (25) this expression is identically zero. Thus c_t satisfies

$$c_t = a\delta_{t,0} + a\beta c_{t-1}$$

where $c_{-1} = 0$. Thus, we conclude that

$$c_t = a(a\beta)^t \tag{48}$$

for all $t \geq 0$.

Equation (46) thus becomes:

$$\begin{aligned}d_t &= \left(\alpha\beta\frac{a-1}{2} - \frac{1}{2}(\alpha+1)\right)b_t^2 + \left(b_0 - \frac{1}{2}\right)\delta_{t,0} + (a\beta)^t b_t + \beta d_{t-1}\\ &= -\frac{1}{2}(1+\alpha+\alpha\beta(1-a))b_t^2 + (a\beta)^t b_t + (b_0 - \frac{1}{2})\delta_{t,0} + \beta d_{t-1}\end{aligned}$$

From (29) and the definition of Λ in (30) we have

$$b_t = \frac{1}{\Lambda}\left(\frac{\alpha\beta}{\Lambda}\right)^t.$$

Thus

$$\begin{aligned}d_0 &= -\frac{\Lambda}{2}b_0^2 + b_0 - \frac{1}{2}\\ &= -\frac{\Lambda}{2}(\frac{1}{\Lambda})^2 + \frac{1}{\Lambda} - \frac{1}{2}\\ &= \frac{1}{2}(\frac{1}{\Lambda} - 1)\end{aligned}$$

Also for $t \geq 1$:

$$\begin{aligned}d_t &= -\frac{1}{2\Lambda}\left(\frac{\alpha\beta}{\Lambda}\right)^{2t} + (a\beta)^t\frac{1}{\Lambda}\left(\frac{\alpha\beta}{\Lambda}\right)^t + \beta d_{t-1}\\ &= -\frac{1}{2\Lambda}\left(\frac{\alpha\beta}{\Lambda}\right)^2\left(\frac{\alpha\beta}{\Lambda}\right)^{2(t-1)} + \frac{1}{\Lambda}\frac{a\alpha\beta^2}{\Lambda}\left(\frac{a\alpha\beta^2}{\Lambda}\right)^{t-1} + \beta d_{t-1}\end{aligned}$$

This is a non-homogeneous first order difference equation. The solution for $t \geq 1$ is

$$\begin{aligned}
d_t &= \beta^t d_0 - \frac{1}{2\Lambda}\left(\frac{\alpha\beta}{\Lambda}\right)^2 \sum_{k=0}^{t-1} \beta^{t-k-1}\left(\frac{\alpha\beta}{\Lambda}\right)^{2k} + \frac{1}{\Lambda}\frac{a\alpha\beta^2}{\Lambda}\sum_{k=0}^{t-1}\beta^{t-k-1}\left(\frac{a\alpha\beta^2}{\Lambda}\right)^k \\
&= \beta^t d_0 - \frac{1}{2\Lambda}\left(\frac{\alpha\beta}{\Lambda}\right)^2 \beta^{t-1}\sum_{k=0}^{t-1}\left(\frac{\alpha^2\beta}{\Lambda^2}\right)^k + \frac{1}{\Lambda}\frac{a\alpha\beta^2}{\Lambda}\beta^{t-1}\sum_{k=0}^{t-1}\left(\frac{a\alpha\beta}{\Lambda}\right)^k \\
&= \beta^t d_0 - \frac{1}{2\Lambda}\left(\frac{\alpha\beta}{\Lambda}\right)^2 \beta^{t-1}\left(\frac{1-(\alpha^2\beta/\Lambda^2)^t}{1-\alpha^2\beta/\Lambda^2}\right) + \frac{1}{\Lambda}\frac{a\alpha\beta}{\Lambda}\beta^t\frac{1-\left(\frac{a\alpha\beta}{\Lambda}\right)^t}{1-\frac{a\alpha\beta}{\Lambda}} \\
&= \beta^t\left(\frac{1}{2\Lambda} - \frac{1}{2} - \frac{1}{2\Lambda}\frac{\alpha^2\beta}{\Lambda^2}\left(\frac{1-(\alpha^2\beta/\Lambda^2)^t}{1-\alpha^2\beta/\Lambda^2}\right) + \frac{1}{\Lambda}\frac{a\alpha\beta}{\Lambda}\frac{1-\left(\frac{a\alpha\beta}{\Lambda}\right)^t}{1-\frac{a\alpha\beta}{\Lambda}}\right)
\end{aligned}$$

Since $\Lambda = \frac{\alpha}{a}$ we see that

$$\frac{\alpha^2\beta}{\Lambda^2} = \frac{a\alpha\beta}{\Lambda} = a^2\beta.$$

Thus the solution for d_t simplifies to

$$d_t = \frac{1}{2}\beta^t\left[\frac{a}{\alpha} - 1 + \frac{a}{\alpha}\frac{a^2\beta}{1-a^2\beta}(1-(a^2\beta)^t)\right]. \tag{49}$$

We can use this solution for d_t, and the matrix $\mathbf{f}$ to find an explicit solution for the coefficients $f_{i,t}$. Let $i = t-n$ where $1 \leq n < t$. Then the matrix $\mathbf{f}$ shows that

$$\begin{aligned}
f_{t-n,t} &= \beta^{t-n}(\beta d_{n-1} - d_n) \\
&= \beta^{t-n}\left[\beta\left(\frac{1}{2}\beta^{n-1}\left[\frac{a}{\alpha} - 1 + \frac{a}{\alpha}\frac{a^2\beta}{1-a^2\beta}(1-(a^2\beta)^{n-1})\right]\right) - \frac{1}{2}\beta^n\left[\frac{a}{\alpha} - 1 + \frac{a}{\alpha}\frac{a^2\beta}{1-a^2\beta}(1-(a^2\beta)^n)\right]\right] \\
&= -\frac{1}{2}\beta^{t-n}\frac{a}{\alpha}(a\beta)^{2n}
\end{aligned}$$

For $t = i \geq 1$, $\mathbf{f}$ gives

$$\begin{aligned}
f_{t,t} &= -\beta^t(d_0 + \frac{1}{2}) \\
&= -\beta^t\left[\frac{1}{2}(\frac{a}{\alpha} - 1) + \frac{1}{2}\right] \\
&= -\frac{1}{2}\beta^t\frac{a}{\alpha}.
\end{aligned}$$

Thus we conclude that for any $t \geq 1$, $1 \leq i \leq t$,

$$f_{i,t} = -\frac{1}{2}\frac{a}{\alpha}\beta^i(a\beta)^{2(t-i)}.$$

The values of $f_{0,t}$ and $f_{i,0}$ can be read directly off the matrix $\mathbf{f}$. Summarizing these results, the terms of interest to us are given by:

$$T(\vec{\tau}) = \sum_{t=1}^{\infty}\sum_{i=1}^{t} \frac{f_{i,t}}{\lambda_t + h_{i,t}} \tag{50}$$

where for $t \geq 1, 1 \leq i \leq t$,

$$h_{i,t} = \sum_{k=t+1-i}^{t} \tau_k, \tag{51}$$

$$f_{i,t} = -\frac{1}{2}\frac{a}{\alpha}\beta^t (a^2\beta)^{t-i}. \tag{52}$$

Using the definition of $F_k^t(\vec{\lambda})$ as the $(k+1)$-th element of $F^t(\vec{\lambda})$, where $F(\vec{\lambda})$ is given by (12), we can reorder the terms of the sum in our expression for $T(\vec{\tau})$ to see that

$$\begin{aligned}
T(\vec{\tau}) = &-\frac{1}{2}\frac{a\beta}{\alpha}\Bigg[\left(\frac{1}{\lambda_1+\tau_1} + (a^2\beta^2)\frac{1}{\lambda_2+\tau_2} + (a^2\beta^2)^2\frac{1}{\lambda_3+\tau_3} + \dots\right) \\
&+ \beta\left(\frac{1}{(\lambda_2+\tau_2)+\tau_1} + (a^2\beta^2)\frac{1}{(\lambda_3+\tau_3)+\tau_2} + (a^2\beta^2)^2\frac{1}{(\lambda_4+\tau_4)+\tau_3} + \dots\right) \\
&+ \beta^2\left(\frac{1}{(\lambda_3+\tau_3+\tau_2)+\tau_1} + (a^2\beta^2)\frac{1}{(\lambda_4+\tau_4+\tau_3)+\tau_2} + (a^2\beta^2)^2\frac{1}{(\lambda_5+\tau_5+\tau_4)+\tau_3} + \dots\right) + \dots\Bigg] \\
= &-\frac{1}{2}\frac{a\beta}{\alpha}\Bigg[\left(\frac{1}{F_0(\vec{\lambda})} + (a^2\beta^2)\frac{1}{F_1(\vec{\lambda})} + (a^2\beta^2)^2\frac{1}{F_2(\vec{\lambda})} + \dots\right) \\
&+ \beta\left(\frac{1}{F_0^2(\vec{\lambda})} + (a^2\beta^2)\frac{1}{F_1^2(\vec{\lambda})} + (a^2\beta^2)^2\frac{1}{F_2^2(\vec{\lambda})} + \dots\right) \\
&+ \beta^2\left(\frac{1}{F_0^3(\vec{\lambda})} + (a^2\beta^2)\frac{1}{F_1^3(\vec{\lambda})} + (a^2\beta^2)^2\frac{1}{F_2^3(\vec{\lambda})} + \dots\right) + \dots\Bigg] \\
= &-\frac{1}{2}b_0\left[\sum_{t=1}^{\infty}\beta^t\sum_{k=0}^{\infty}\left(\frac{b_k}{b_0}\right)^2\frac{1}{F_k^t(\vec{\lambda})}\right].
\end{aligned} \tag{53}$$

where in the last line we've used the solution for b_t in (29). This is the expression stated in the proposition.

In the process of solving for the parameters that enter the term T we solved for k, c_t, d_t, $f_{i,t}$ and $h_{i,t}$. To show that our guess for the value function does indeed yield the solution, we now derive expressions for the final outstanding coefficients of the value function, $D_{t,p}$. From the Bellman equation we see that

$$D_{t,p} = -\frac{1}{2}\left[(1+\alpha)2b_tb_p - \delta_{t,0}2b_p\right] + \beta\Big[2kb_tb_p + (1-\delta_{t,0})c_{t-1}b_p + b_tc_{p-1} + (1-\delta_{t,0})D_{t-1,p-1}\Big].$$

For $t = 0$, we find

$$D_{0,p} = A(a\beta)^p \quad , \quad A = \frac{a}{\alpha} \ .$$

For $t \geq 1$,

$$D_{t,p} = A(a\beta)^{t+p} + \beta D_{t-1,p-1} \quad , \quad A = \frac{a}{\alpha} \ .$$

The recursive equation

$$y(m,n) = A\xi^{m+n} + By(m-1,n-1) \quad , \ m < n$$

has the solution

$$y(m,n) = A\xi^{m+n}\frac{1-\left(\frac{B}{\xi^2}\right)^m}{1-\frac{B}{\xi^2}} + B^m y(0,n-m) \ .$$

Applying this general formula with $\xi = a\beta$ leads to

$$D_{t,p} = \frac{a}{\alpha}(a\beta)^{t+p}\frac{1-(a^2\beta)^{-t}}{1-(a^2\beta)^{-1}} + \frac{a}{\alpha}\beta^t(a\beta)^{p-t} \ .$$

Thus we have found unique solutions for all the free coefficients of our guess for the value function, confirming that the initial guess does indeed yield the solution.

E Proof of Proposition 3

From the proof of Proposition 2 we have

$$\frac{dV}{d\tau_m} = \frac{dT}{d\tau_m} = -\sum_{i=1}^{\infty}\sum_{t=1}^{\infty}\frac{f_{i,t}}{(\lambda_t + h_{i,t})^2}\frac{dh_{i,t}}{d\tau_m}.$$

From (40),

$$\frac{dh_{i,t}}{d\tau_m} = \begin{cases} 1 & t \geq m \text{ and } t \geq i \geq t+1-m \\ 0 & \text{otherwise} \end{cases}$$

Hence,

$$\frac{dV}{d\tau_m} = -\sum_{t=m}^{\infty}\sum_{i=t+1-m}^{t}\frac{f_{i,t}}{(\lambda_t + \sum_{k=t+1-i}^{t}\tau_k)^2}$$

Evaluate this quantity at $\tau_t = 0$ for all t:

$$\left.\frac{dV}{d\tau_m}\right|_0 = -\sum_{t=m}^{\infty}\frac{1}{\lambda_t^2}\sum_{i=t+1-m}^{t}f_{i,t}$$

Let $t = m+k$ where $k \geq 0$, and consider the sum $\sum_{i=t+1-m}^{t} f_{i,t} = \sum_{i=k+1}^{m+k} f_{i,m+k}$. This sum is equivalent to starting at diagonal element $m+k+1, m+k+1$ of the matrix $\mathbf{f}$ in (44), and summing the m terms above this diagonal element (including the diagonal). Reading off the matrix, we see that this sum simplifies to:

$$\sum_{i=k+1}^{m+k} f_{i,m+k} = -\beta^{k+1} d_{m-1} - \frac{1}{2}\beta^{m+k}.$$

and hence

$$\begin{aligned}\left.\frac{dV}{d\tau_m}\right|_0 &= -\sum_{t=m}^{\infty} \frac{1}{\lambda_t^2} \sum_{i=t+1-m}^{t} f_{i,t} \\ &= \left(d_{m-1}\sum_{k=0}^{\infty}\frac{\beta^{k+1}}{\lambda_{m+k}^2} + \frac{1}{2}\beta^m \sum_{k=0}^{\infty}\frac{\beta^k}{\lambda_{m+k}^2}\right).\end{aligned}$$

Using the definition of $g(m)$ this expression becomes

$$\left.\frac{dV}{d\tau_m}\right|_0 = \beta g(m)\left(d_{m-1} + \frac{1}{2}\beta^{m-1}\right).$$

From (49) we have

$$\begin{aligned} d_{m-1} + \frac{1}{2}\beta^{m-1} &= \beta^{m-1}\left(\frac{a}{\alpha} + \frac{a}{\alpha}\frac{a^2\beta}{1-a^2\beta}(1-(a^2\beta)^{m-1})\right) \\ &= \frac{a}{\alpha}\beta^{m-1}\left(1 + \frac{a^2\beta}{1-a^2\beta}(1-(a^2\beta)^{m-1})\right) \\ &= \frac{a}{\alpha}\beta^{m-1}\left[\frac{1-(a^2\beta)^m}{1-a^2\beta}\right].\end{aligned}$$

The result follows.

F Behaviour of R_m

When $\phi < \beta$, it is obvious from (20) that R_m is increasing in m. We thus focus on the case $\phi > \beta$. From the formula (20), and the requirement $\phi > \beta$, it is clear that $\lim_{m\to\infty} R_m = 0$. Here we show that R_m is either monotonically decreasing in m, or has a unique global maximum for some $m \geq 2$, and characterize the parameter ranges where these two behaviours occur.

The fact that R_m has at most one maximum at $m \geq 2$ can be shown by treating m as a continuous variable. Then R_m has a stationary point iff $\frac{d}{dm}R_m = 0$, which a little algebra shows occurs if

$$(a^2\beta)^m \ln\left(\frac{a^2\beta^2}{\phi}\right) = \ln\left(\frac{\beta}{\phi}\right). \tag{54}$$

This condition has at most one solution for $m \geq 1$. Since $R_m > 0$ for all m, $R_1 = 1$,

$\lim_{m\to\infty} R_m = 0$, and dR_m/dm changes sign at most once, R_m cannot have a local minimum. Thus R_m must be either monotonically declining, or be unimodal with a global maximum at some $m \geq 2$.

It is simple to determine conditions under which these different qualitative behaviours occur. Since if R_m is not monotonically declining it must be unimodal, the condition $R_2 > R_1 = 1$ is both necessary and sufficient for R_m to be unimodal. A little algebra shows that $R_2 > 1 \iff \Gamma \equiv a^2\beta^2 + \beta - \phi > 0$. Since $a = 0$ at $\alpha = 0$, we know $\Gamma = \beta - \phi < 0$ when $\alpha = 0$. Also, since a is increasing in α, so is Γ. Combining these facts we see that the set of parameters values for which $\Gamma > 0$ must either be empty, or of the form $\alpha > \hat{\alpha}(\beta, \phi)$, where $\hat{\alpha}(\beta, \phi)$ is some critical value of α at which $\Gamma = 0$. Solving the condition $\Gamma = 0$ for α, we find two solutions:

$$\alpha_1 = \frac{(\phi-\beta)(1+\beta) + \phi\sqrt{\phi-\beta}}{\beta^2 + (\phi-\beta)^2 - (\phi-\beta)(1+\beta^2)} \quad , \quad \alpha_2 = \frac{(\phi-\beta)(1+\beta) - \phi\sqrt{\phi-\beta}}{\beta^2 + (\phi-\beta)^2 - (\phi-\beta)(1+\beta^2)} \ .$$

α_2 is negative for all β and $\phi \in [\beta, 1]$ so we conclude that

$$\hat{\alpha}(\beta, \phi) = \frac{(\phi-\beta)(1+\beta) + \phi\sqrt{\phi-\beta}}{\beta^2 + (\phi-\beta)^2 - (\phi-\beta)(1+\beta^2)} \ . \tag{55}$$

Observe that $\hat{\alpha}(0, \phi) = \frac{\phi(1+\sqrt{\phi})}{\phi(\phi-1)} < 0$ so Γ is negative at $\beta = 0$ irrespective of α. To find the conditions on β under which $\hat{\alpha}(\beta, \phi) \geq 0$ we solve $\hat{\alpha}(\beta, \phi) = 0$ for β, finding the following three roots:

$$\beta_1 = \phi \ , \ \beta_2 = -\frac{1+\sqrt{1+4\phi}}{2} \ , \ \beta_3 = \frac{\sqrt{1+4\phi}-1}{2} \ .$$

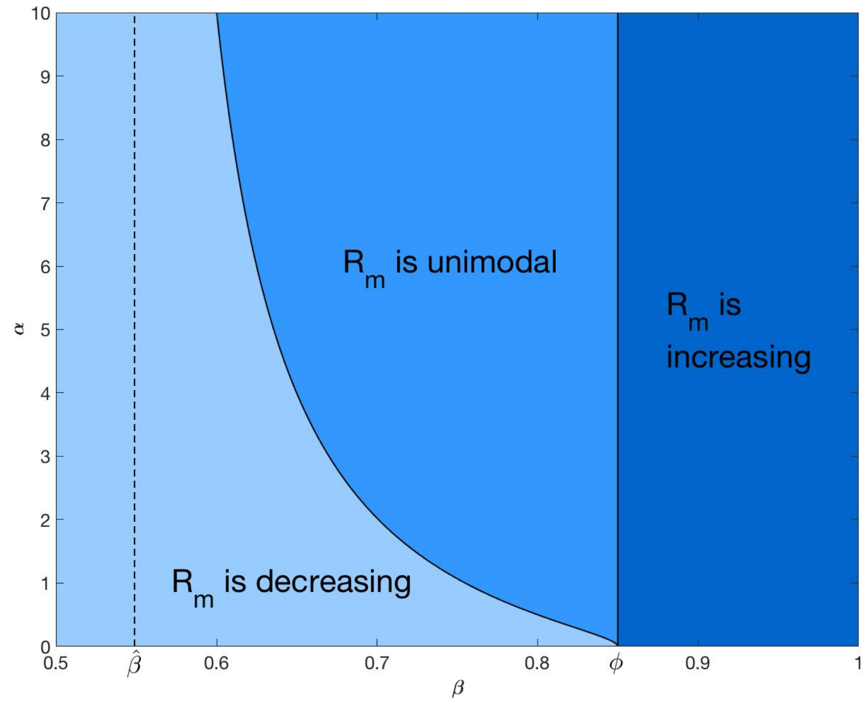
β_1 violates the condition $\beta < \phi$, β_2 is always negative, but $\beta_3 < \phi$ which makes the latter the relevant critical level of β at which $\hat{\alpha}(\beta, \phi) \geq 0$. Thus we define the critical value of β as

$$\hat{\beta}(\phi) = \frac{\sqrt{1+4\phi}-1}{2} \ . \tag{56}$$

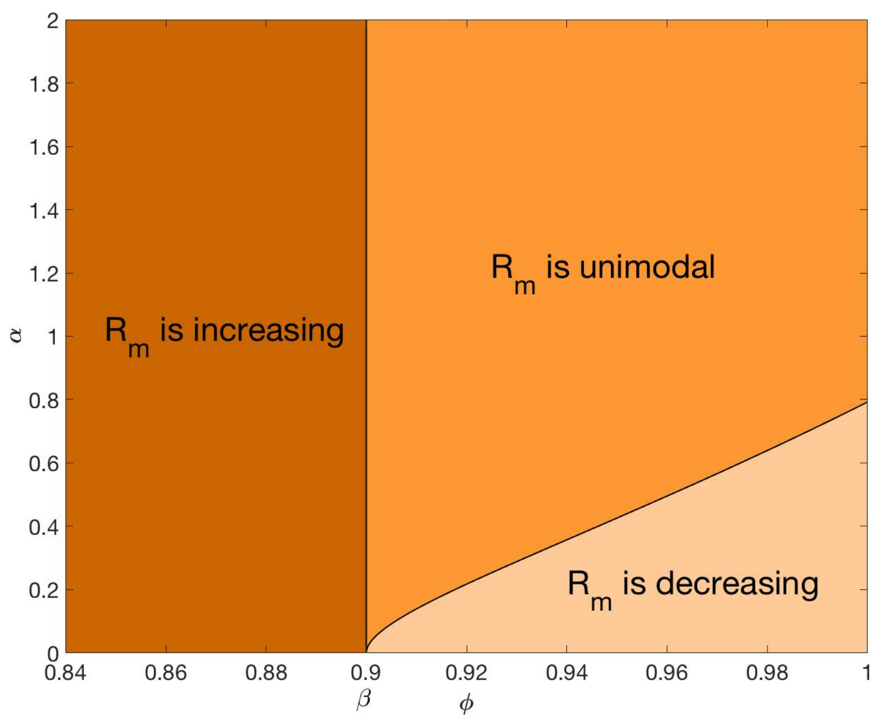
Thus, when $\beta \in [\hat{\beta}, \phi)$, R_m has a maximum at some $m > 1$ if $\alpha > \hat{\alpha}$, otherwise R_m is decreasing. Figure 6 below demonstrates these results graphically.

G Sensitivity analysis for Figure 3

Figures 7 and 8 below represents the outcome of a calculation identical to that in Fig. 3, but for $\lambda_0/B = 1/50$ and $\lambda_0/B = 50$ respectively. Fig. 7 simply demonstrates that the rate of decline of the prior with the time horizon has no effect on budget allocations when λ_0/B is small. In Fig. 8 predictions are marginal relative to the prior, making interactions between lead times unimportant. To a good approximation then, the value function is linear in forecast precisions in this case, as discussed in Proposition 3. Thus, when $\phi > \beta$, we expect the entire



(a) $\phi = 1$



(b) $\beta = 0.9$

Figure 6: Qualitative behaviour of R_m in different regions of parameter space.

budget to be allocated to the most valuable forecast lead time (i.e. the value of m for which R_m in (20) is maximized). The bottom panel of Fig. 8 confirms this expectation. When $\phi < \beta$ however, the marginal analysis in Proposition 3 shows that the value of a marginal unit of predictability is increasing in lead time m – there is no 'most valuable' lead time. Since the agent cannot allocate her entire budget to infinite lead times, and λ_0/B is large, but not infinite in Fig. 8 (so forecasts are only approximately marginal), interaction effects are still at work in this case, and lead to the spread out peaks in the top panel of Fig. 8. Notice however that these peaks place more weight on the long run than the analysis for $\lambda_0/B = 1$ in Fig. 3, indicating that first order effects are more important in this case than in Fig. 3, as we would expect when choosing a very large value of λ_0/B. We emphasize however that $\lambda_0/B = 50$ is an unrealistically large value. As discussed in the text, priors and forecasts usually have roughly the same precisions in practice, so the results in Fig. 8 grossly underestimate the importance of the interactions between lead times.

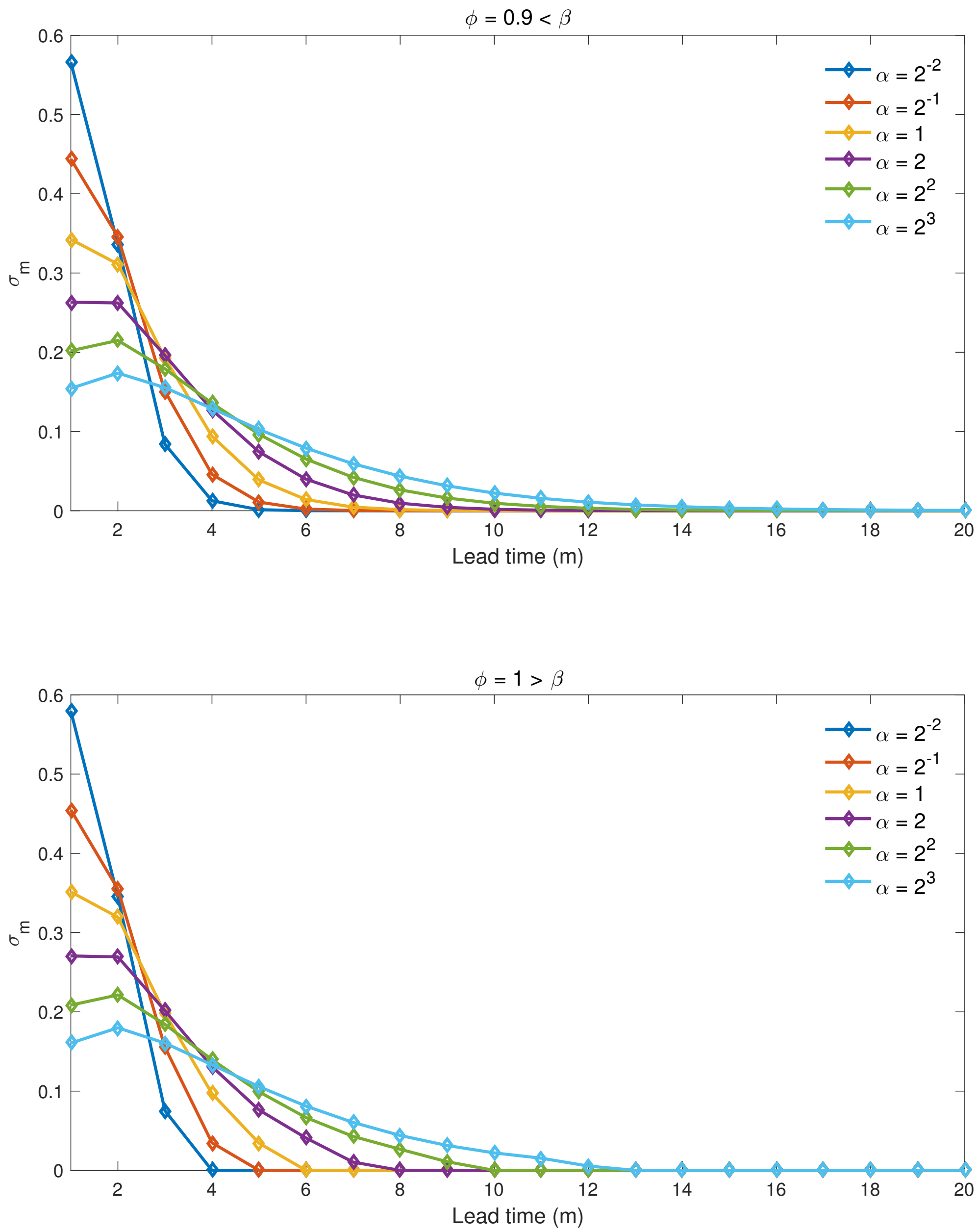


Figure 7: Budget share σ_m allocated to lead time m in the optimization problem in (21). $\beta = 0.95$, $\frac{\lambda_0}{B} = 1/50$.

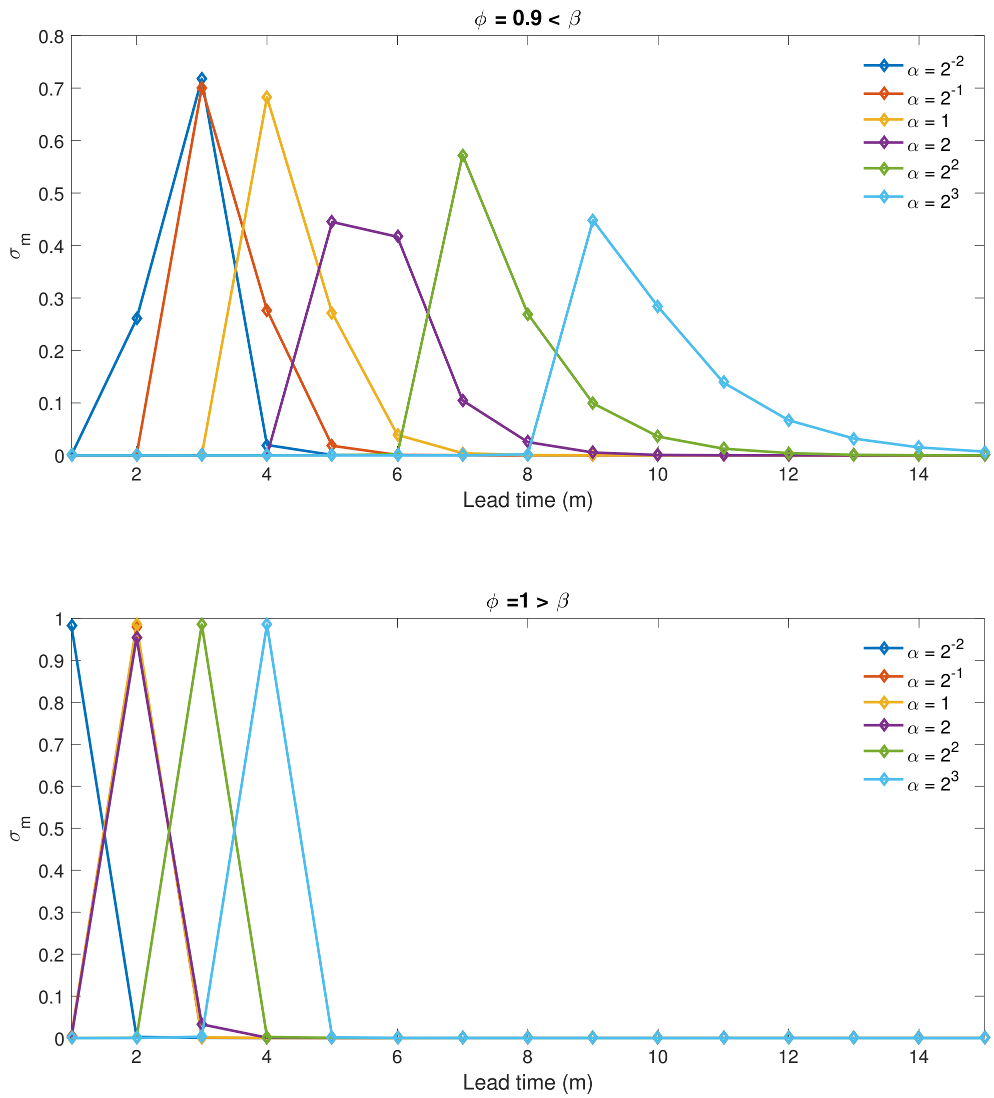


Figure 8: Budget share σ_m allocated to lead time m in the optimization problem in (21). $\beta = 0.95, \frac{\lambda_0}{B} = 50$.