Grantham Research Institute on Climate Change and the Environment

Centre for Climate Change Economics and Policy

Valuing predictability

Antony Millner and Daniel Heyen

January 2017

Centre for Climate Change Economics and Policy Working Paper No. 293

Grantham Research Institute on Climate Change and the Environment Working Paper No. 2

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

The Grantham Foundation for the Protection of the Environment

University of Leeds

The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council. Its second phase started in 2013 and there are five integrated research themes:

1. Understanding green growth and climate-compatible development
2. Advancing climate finance and investment
3. Evaluating the performance of climate policies
4. Managing climate risks and uncertainties and strengthening climate services
5. Enabling rapid transitions in mitigation and adaptation

More information about the Centre for Climate Change Economics and Policy can be found at: http://www.cccep.ac.uk.

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and the Global Green Growth Institute. It has nine research programmes:

1. Adaptation and development
2. Carbon trading and finance
3. Ecosystems, resources and the natural environment
4. Energy, technology and trade
5. Future generations and social justice
6. Growth and the economy
7. International environmental negotiations
8. Modelling and decision making
9. Private sector adaptation, risk and insurance

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: http://www.lse.ac.uk/grantham.

This working paper is intended to stimulate discussion within the research community and among users of research, and its content may have been submitted for publication in academic journals. It has been reviewed by at least one internal referee before publication. The views expressed in this paper represent those of the author(s) and do not necessarily represent those of the host institutions or funders.

Valuing Predictability

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

[1]London School of Economics and Political Science

January 19, 2017

Abstract

How important is it to be able to predict the distant future? We study this question in a model of an agent who operates in a non-stationary stochastic environment. Payoffs depend on how well adapted activities are to current conditions, and activities may be adjusted to account for anticipated environmental changes, at a cost. We compute the value of *prediction systems*, which produce forecasts of the future with a given profile of accuracy as a function of lead time in every period. This allows us to quantify the importance of predictive accuracy at each lead time. Even if adjustment costs, discount factors, and long-run uncertainty are large, short-run predictability is often more important than long-run predictability.

'If you have to forecast, forecast often.'
– Edgar R. Fiedler, The Three Rs of Economic Forecasting: Irrational, Irrelevant and Irreverent, 1977.

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, econometricians forecast demand for new public transportation systems, and statisticians predict electoral outcomes. 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 emotional distress and planning inertia. Where Science cannot satisfy our demand for certainty, others may nevertheless oblige us. A motley crew of modern-day

* Millner: a.millner@lse.ac.uk. We gratefully acknowledge funding from the ESRC Centre for Climate Change Economics and Policy and Grantham Foundation for the Protection of the Environment. Daniel Heyen gratefully acknowledges funding from the German Research Foundation, grant HE 7551/1-1.

soothsayers prognosticate on a variety of salient, but difficult to predict, long-run events. Pundits foretell the rise and decline of nations, futurologists prophesy technological revolutions that will transform our grandchildren's lives, and astrologers predict our romantic entanglements decades into the future. But just how important is it to be able to predict the distant future? Does large long-run uncertainty imply that accurate long-run predictions would necessarily be highly valuable? Or can long-run predictions perhaps be substituted by short-run forecasts when decisions can be adjusted dynamically as new information arrives? What might an 'optimal' pattern of predictive accuracy across forecast lead times look like in different decision environments? This paper attempts to shed light on these questions.

We develop a conceptual model that allows us to compute the value of predictability for a rational decision-maker who operates in an uncertain, non-stationary, environment. It is important to distinguish *predictability* from *information* at the outset. To fix ideas consider the following stylized scenario. Anna is the head of a disaster relief organization that operates on the US East coast. On a certain Sunday evening she receives a weather report showing that a hurricane is brewing in the mid-Atlantic. The report details the hurricane's predicted path over the coming week, showing that it is expected to make landfall somewhere on the East coast on Friday. While the position of the hurricane on Monday or Tuesday is well constrained by the forecast, its position at the time of landfall is highly uncertain. Anna nevertheless makes some initial plans to deploy resources based on the current forecast, with the knowledge that she will receive an updated forecast tomorrow, which may cause her to make a costly change of plans. Indeed, Anna knows that a new forecast will be produced at regular intervals, with each successive forecast providing increasingly accurate information about the hurricane's position on Friday.

Clearly the forecast Anna receives on Sunday contains information about where her resources should be deployed on Friday. However, in order to assign a value to our ability to predict hurricane paths we must account for the continually updated nature of forecasts, their accuracy as a function of lead time, and the costs Anna's organization sustains when adjusting its decisions dynamically in response to new information. To formalize the distinction between predictability and information we define a *prediction system* to be an information structure (Blackwell, 1953) that travels through time with a decision-maker, producing forecasts of all future events with a given profile of accuracy as a function of lead time in every period. Not only do prediction systems provide decision-makers with information about future events in the current period, they also determine their expectations about the information that will be available when they take future decisions.

One may think of a prediction system as an abstract representation of the dynamic models economists and natural scientists use for forecasting. These models usually posit a law of motion that translates current values of state variables into predictions of the future values of states. They produce an updated view of all future periods in every period, based on prevailing conditions. Forecast errors from such models are usually highly dependent on lead time. Since model errors compound over time, long-run forecasts are almost always less reliable than short run forecasts.[1] The concept of a prediction system captures these aspects of dynamic models in a tractable and flexible manner, without the need to specify any specific model based on laws of motion. This approach allows us to flexibly manipulate a prediction system's ability to predict events at different temporal distances, making it possible to study a decision-maker's appetite for predictability as a function of lead time in a manner that would be impossible with a traditional dynamic model.

More specifically, our model considers a decision-maker whose period payoffs depend on how well adapted her choices are to the current, uncertain, state of the world. The decision-maker may adjust her choices in every period to account for changes in her environment, but faces convex adjustment costs, which penalize large adjustments more heavily than small ones. 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, captured by the magnitude of the adjustment costs she faces.

Although the resulting model corresponds to a non-trivial stochastic dynamic programming problem with learning about an infinite number of state variables (corresponding to the decision-maker's beliefs about each future period), we are able to find an analytic expression for the decision-maker's value function V for arbitrary 'normal' prediction systems.[2] That is, we compute the decision-maker's discounted expected payoffs as a function of the profile of predictive accuracy that the prediction system exhibits:

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

[1] Even if a model is structurally flawless, small uncertainties in the measurement of initial conditions can grow exponentially with time. This is a defining feature of chaotic dynamics, and can occur in the simplest of systems. See Taleb (2008, p. 177) for compelling examples.

[2] The reader may wonder why we don't simply study a finite horizon model. The reason is simple. Consider a two period model, in which the decision-maker receives one forecast at the beginning of period 1, and another at the beginning of period 2. The period 1 forecast will contain both short- and long-run predictions, but the period 2 forecast will contain only a short-run prediction. There is thus a substantial bias towards short-run predictions in such a model. An infinite horizon model is necessary in order to place all forecast lead times on an equal footing.

This allows us to quantify the essential tradeoffs between predictability at different lead times. These tradeoffs are concisely captured by Myers et al. (2000) in their work on forecasting the emergence of infectious diseases:

'At the heart of early warning is a basic trade-off between the specificity of predictions…and the lead times which those predictions can provide. In general, long-range forecasts give the least specific warnings, but have the advantage of providing planners with relatively long lead times. At the other extreme, systems based on early detection of cases provide highly specific information on the timing and location of outbreaks, but allow little time for implementing remedial measures.'

Although phrased in terms of the time available to react to a warning, translated into economic terms this quote highlights the fact that warnings that come late are more costly to react to than warnings that come early (i.e. adjustment costs are convex). But exactly how do the accuracy of predictions at different lead times and the costs of reacting to them jointly determine the value of a given prediction system? It is intuitively clear that current planning decisions need not account much for the future if adjustment costs are small, since the decision-maker may cheaply react to new conditions as they arise. This suggests that when adjustment is cheap, short-run predictability will be more important than long-run predictability. What is less clear intuitively is how demand for predictability as a function of lead time changes as the costs of adjustment increase. Our solution for the decision-maker's value function V provides a general quantitative answer to this question within the context of our model. Since V is a non-separable function of the τ_m, the interactions between predictive accuracy at different lead times are potentially important determinants of the overall value of a prediction system. These interactions complicate the problem of determining the contribution of each lead time to the value function. We use two different techniques to extract the information the value function contains about the importance of predictive accuracy as a function of lead time.

We first perform a Taylor expansion of the value function, assuming that the precision of predictions is small. This results in a linear approximation that is separable in the τ_m, allowing the contribution of each lead time to overall value to be computed independently. In this approximation we show that the value function exhibits one of two qualitative patterns of dependence on predictability as a function of lead time. When agents discount the future heavily, or have low adjustment costs, the value of a marginal unit of predictive accuracy about events m time steps in the future is a declining function of m. If however decision-makers are sufficiently patient, and adjustment costs sufficiently large, the value of a marginal unit of predictability about events at lead time m is a unimodal function of m, with a global maximum at some $m^* > 1$. The most valuable forecast lead time m^* is frequently small, even when the decision maker faces large adjustment costs and has a low discount rate.

We then extend our analysis to non-marginal prediction systems, focusing on characterizing an 'optimal distribution' of predictive accuracy across forecast lead times. The weight on lead time m in an optimal distribution is defined as the share of a hypothetical total predictability budget that the agent would allocate to lead time m. Unlike the marginal analysis, this approach captures the interactions between forecast lead times. We show that flexible agents should optimally like predictability to be strongly concentrated on short lead times, while inflexible agents should prefer a more diffuse distribution of predictability across all lead times, which nevertheless places more weight on shorter lead-times. Optimal predictability distributions either assign declining weight to larger lead times, or are unimodal functions. For a wide range of parameters the most important forecast lead time is small (i.e. one or two time steps ahead).

Finally, we discuss applications of the insights our model delivers to applied prediction problems in the natural and social sciences. Our leading empirical example considers changes in the predictability of Atlantic hurricane paths over the past 45 years. Scientific advances over this period have improved hurricane predictability at all lead times, with the greatest reduction in forecast errors coming for long-run forecasts. Nevertheless, we show that even though gains in the accuracy of short and medium-run forecasts have been more modest, they account for the dominant share of the increase in value for many decision-makers. We suggest that these findings could have consequences for how existing proposals to fund research aimed at improving predictability should be directed if they are to maximize value for decision-makers. We discuss applications in other areas, suggesting that accounting for the dynamic nature of decision-making and prediction can substantially alter intuitive conceptions of the relative importance of short- and long-run predictability, even when long-run uncertainty is large.

The paper is structured as follows. We discuss related literature next, before presenting the model and our results in Section 2. Section 3 discusses the findings, their applications, and limitations, and concludes.

Related literature

As far as we know the questions we seek to address in this paper are novel in the literature. Nevertheless, our contribution bears a family resemblance to two strands of work. The first deals with general rankings of information structures in static environments. The second is concerned with the value of forecasts in applied settings.

The literature on the value of information begins with the pathbreaking work of Blackwell (1953) and Marschak & Miyasawa (1968), which sought to define a notion of 'more information' that is independent of the details of agents' decision problems and preferences. While extremely general, the ordering of information structures that Blackwell obtained is partial, and fails to rank even very simple informational environments (Lehmann, 1988). Subsequent authors have obtained more decisive rankings of information structures by placing constraints on the nature of decision-maker's payoff functions (Lehmann, 1988; Athey & Levin, 1998; Persico, 2000; Cabrales et al., 2013). Our work is notionally related to this literature, in that we are also concerned with the value of informational systems, but differs 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 generated by information structures. We thus depart from this literature's emphasis on obtaining general informational orderings that rely on few assumptions about agents' decision-problems. The return for this loss of generality is that we are able to study a much richer set of dynamic decisions and information production systems than are typically used in this literature.[3] This allows us to capture essential features of real-world forecasting products, and enables our analysis of the relative importance of short and long-run predictability.

A related applied literature uses stylized models to quantify the economic value of forecasting products for rational decision-makers. Much of the early work used simple static binary decision problems to estimate the value of weather forecasts (Angström, 1919; Nelson & Winter, 1964). These models have become staples of forecast evaluation (see e.g. Katz & Murphy, 1997), but because they are static they cannot capture the dynamic characteristics of forecasting products.[4] Two leading handbooks on economic forecasting (Clements & Hendry, 2011; Elliott et al., 2006; Elliot & Timmermann, 2013) devote little space to the issue of valuing forecasting products. The former has a single chapter on forecast value that covers much the same ground as Katz & Murphy (1997). The latter has no explicit discussion of forecast value, but does discuss the relationship between forecast evaluation and decision theory (Granger & Machina, 2006), focussing on the specification of loss functions. A substantial statistical literature delineates the practical difficulties of long-run forecasting in a variety of contexts (see e.g. Granger & Jeon (2007); Lindh (2011)). Implicit in this work is an assumption that accurate long-run forecasting is difficult, but would be of considerable value if achievable. Our work illustrates how a decision-maker's ability to adapt to new in-

[3]There is of course a substantial literature that examines the effect of uncertainty and learning on optimal dynamic decision-making. Most of this work falls into two categories: two period models that examine the effect of second period learning on optimal first period decisions (Arrow & Fisher, 1974; Epstein, 1980), or infinite horizon models that involve learning about the realizations of a stochastic state variable (e.g. Dixit & Pindyck, 1994), or a parameter of a structural dynamic-stochastic model (e.g. Demers, 1991). Neither of these approaches can capture the effects we study here. Two period models cannot capture the repeatedly updated nature of prediction and the dependence of forecast accuracy on lead time, both essential features of our model (see also footnote 2 on the need for an infinite horizon model). Models based on stochastic processes or learning about parameters of state equations do not allow a prediction system's accuracy 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.

[4]Chapter 6 of Katz & Murphy (1997) discusses some dynamic decision models, but the forecasts they study only provide information about events one time step ahead. These models are thus of little use for assessing the relative value of short- and long-run predictions.

formation dynamically affects the relative importance of long- and short-run predictability. This demonstrates that lack of long-run predictability need not have a substantial impact on payoffs, even if long-run uncertainty is large.

2 The model

We consider a decision-maker who faces an uncertain exogenous state of the world at each time $n \in \mathbb{N}$. The units of time are arbitrary, but should be understood to match the frequency of forecast updates in applications (e.g. days for weather forecasts, quarters for inflation forecasts). We assume that the decision-maker's possible activities can be mapped into the real line, and denote a generic choice by $X \in \mathbb{R}$. The decision-maker may adjust her choice X in each period, at a cost that is convex in the magnitude of the adjustment. 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. Since adjustment costs are convex, the decision-maker's choices must achieve a balance between exploiting current conditions and preparing for future conditions, thus avoiding excessively large and costly adjustments later on.

We assume that the loss-minimizing decisions $\tilde{\theta}_n$ are independent, but not identically distributed.[5] For any n, let $\theta_t = \tilde{\theta}_{n+t}$ for $t \geq 0$, i.e. θ_t is the value of $\tilde{\theta}$ that will be realized t time steps in the future. 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 a sequence of normal distributions with means μ_t^0 and precisions (i.e. inverse variance) λ_t^0, i.e.

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

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 $\frac{1}{2}\alpha(X_{n+1} - X_n)^2$. 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.[6]

[5] The independence assumption is essential if we are to cleanly define a notion of predictability as a function of lead time. Suppose that the $\tilde{\theta}_n$ are correlated, and 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 $\tilde{\theta}_n$. Since any signal from a prediction system about a future event affects information about all future events in a correlated model, no signal can be uniquely associated with predictability at a specific lead time. See Section 3 for further discussion.

[6] The use of quadratic loss functions has been criticized for placing excessively strong restrictions on agent's preferences in some decision contexts (Granger & Machina, 2006; Elliott & Timmermann, 2008). While we acknowledge that this is a potential limitation of our analysis, the fact that our model reduces to a stochastic dynamic control problem with an infinite number of state variables means that progress is entirely dependent on being about to solve for the agent's value function in closed form. This is only possible when the loss

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.

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 the prediction system has an exogenous profile of precision as a function of lead time, parameterized by the infinite sequence

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

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. Consider two forecast 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 system A is more informative (in the sense of Blackwell) than system 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 infinite 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

function is quadratic and location independent. See Section 3 for further discussion.

next:

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

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

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

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.

Bellman equation

Using our new notation, we can write the current expected period payoff as a function of the state variables (X_n, Y_n) and the decision variable X_{n+1} as:

$$W(X_{n+1}, X_n, Y_n) = -\frac{1}{2}\left[\int (X_{n+1}-\theta_0)^2 p_n(\theta_0; \mu_0^n, \lambda_0^n) d\theta_0 + \alpha(X_{n+1}-X_n)^2\right]$$
$$= -\frac{1}{2}\left[(1+\alpha)X_{n+1}^2 + \alpha X_n^2 - 2X_{n+1}(\mu_0 + \alpha X_n) + \frac{1}{\lambda_0^n} + (\mu_0^n)^2\right]. \tag{6}$$

Let $V(X_n, Y_n)$ be the current value of the infinite dimensional state (X_n, Y_n). 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:

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

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

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

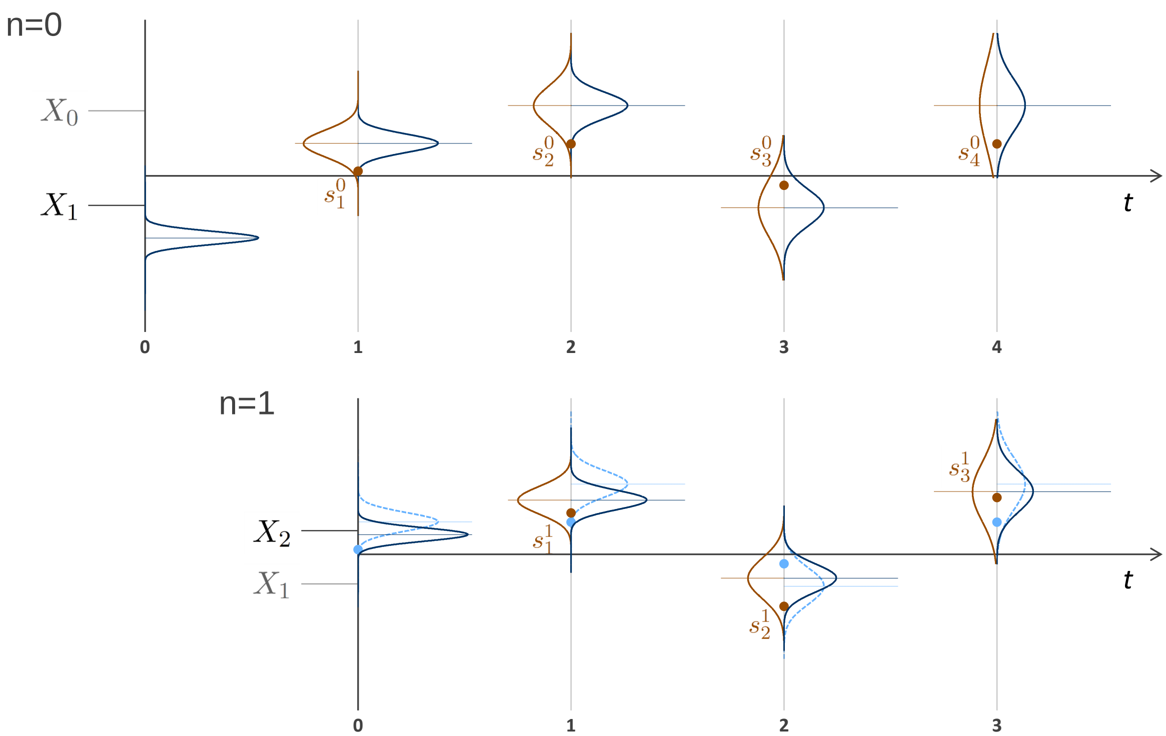


Figure 1: Illustration of the model setup. The figure depicts the agent's beliefs, choices, and the information provided by the prediction system, in the first two periods $n = 0, 1$. At the beginning of period $n = 0$ the agent holds a sequence of prior beliefs about the future values of the state of 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 $n = 1$.

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 $dS^n = \prod_{t=1}^{\infty} ds_t^n$, and $\beta \in (0, 1)$ is the agent's discount factor. 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 agents' expectations about the values of future signal 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 dimensionality of the state space in our model, standard methods based on the Benveniste-Scheinkman condition (Benveniste & Scheinkman, 1979) yield simple closed form solutions for the optimal control rule.

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

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

and the coefficients a, b_t satisfy:

$$\lim_{\alpha\to 0} a = 0,\ \lim_{\alpha\to\infty} a = 1,\ \frac{\partial a}{\partial \alpha} > 0 \tag{12}$$

$$a + \sum_{t=0}^{\infty} b_t = 1 \tag{13}$$

$$\lim_{\alpha\to 0} b_t = \begin{cases} 1 & t = 0 \\ 0 & t > 0 \end{cases},\ \lim_{\alpha\to\infty} b_t = 0 \tag{14}$$

$$\frac{\partial}{\partial \alpha}\left(\frac{b_{t+1}}{b_t}\right) > 0,\ \frac{\partial b_0}{\partial \alpha} < 0. \tag{15}$$

This proposition 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. We emphasize however that although the policy rule does not depend on uncertainty, the value function certainly will. Since our goal is to study the value of predictability, it is the dependence of the value function 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 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 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 $a + \sum_{t=0}^{\infty} b_t = 1$,

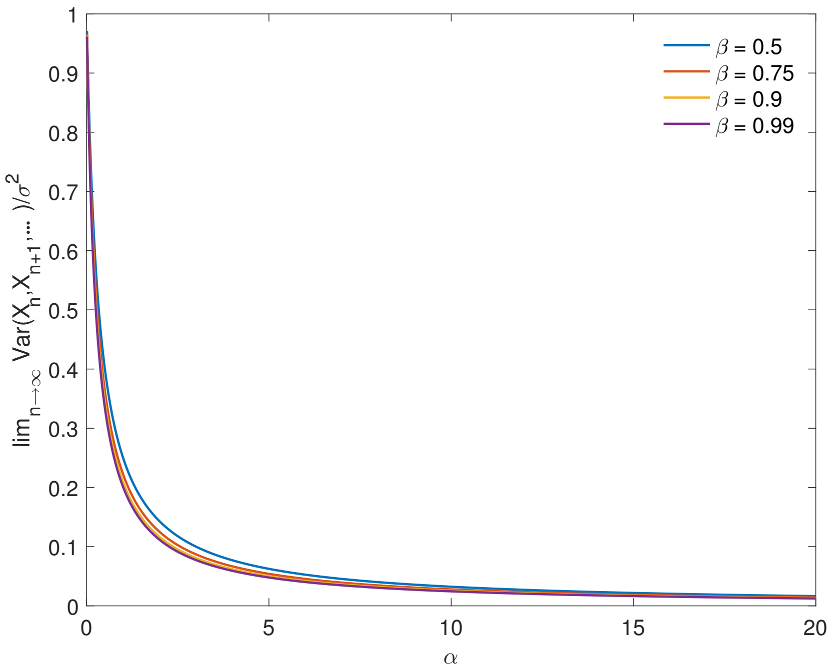


Figure 2: Asymptotic variability of the decision variable X relative to the variability of the loss-minimizing decisions $\tilde{\theta}$, assuming that the values of $\tilde{\theta}_n$ are deterministic and given by a fixed sequence of draws from a random variable with variance σ^2.

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 $aX + \sum_{t=0}^{\infty} b_t X = X$.

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

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

for arbitrary initial condition X_0. Figure 2 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. Thus, $\alpha = 3$ is already a fairly large value of the adjustment cost parameter. In addition, 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 welfare, 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 on the precision state variables λ_t (see eq. 6)[7], 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:

Proposition 2. *The value function $V(X,Y)$ has the form*

$$V(X,Y) = kX^2 + \sum_{t=0}^{\infty} c_t \mu_t X + \sum_{t=0}^{\infty} \sum_{p=t+1}^{\infty} D_{t,p} \mu_t \mu_p + \sum_{t=0}^{\infty} d_t \mu_t^2 + \sum_{t=0}^{\infty} \sum_{i=0}^{t} \frac{f_{i,t}}{\lambda_t + h_{i,t}}. \tag{16}$$

Closed form expressions for the parameters $k, c_t, D_{t,p}, d_t, f_{i,t}, h_{i,t}$ may be obtained. The only terms of the value function that depend on the forecast precision profile $\vec{\tau}$ are

$$T = T(\vec{\tau}) \equiv \sum_{t=1}^{\infty} \sum_{i=1}^{t} \frac{f_{i,t}}{\lambda_t + h_{i,t}} \tag{17}$$

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

$$h_{i,t} = \sum_{k=t+1-i}^{t} \tau_k, \tag{18}$$

$$f_{i,t} = -\frac{1}{2} \frac{a}{\alpha} \beta^t (a^2 \beta)^{t-i}. \tag{19}$$

To interpret T, begin by letting $F(\lambda_k)$ be the λ_k component of the belief updating rule $F(Y,S)$ in (5), where we suppress the dependence on signals S, as precisions update deterministically. In addition, let F^t be the t-th iterate of F. Then by reordering terms we see

[7] One could of course write the model in terms of the variance of the agent's beliefs about θ_t, in which case payoffs would be linear in variances, but then the state equations for the variances would be non-linear (see (5)), rendering standard methods from linear quadratic control inapplicable.

that

$$
\begin{aligned}
T = &-\frac{1}{2}\frac{a\beta}{\alpha}\Bigg[\left(\frac{1}{\lambda_1+\tau_1}+(a^2\beta^2)\frac{1}{\lambda_2+\tau_2}+(a^2\beta^2)^2\frac{1}{\lambda_3+\tau_3}+\dots\right)\\
&+\beta\left(\frac{1}{(\lambda_2+\tau_2)+\tau_1}+(a^2\beta^2)\frac{1}{(\lambda_3+\tau_3)+\tau_2}+(a^2\beta^2)^2\frac{1}{(\lambda_4+\tau_4)+\tau_3}+\dots\right)\\
&+\beta^2\left(\frac{1}{(\lambda_3+\tau_3+\tau_2)+\tau_1}+(a^2\beta^2)\frac{1}{(\lambda_4+\tau_4+\tau_3)+\tau_2}+(a^2\beta^2)^2\frac{1}{(\lambda_5+\tau_5+\tau_4)+\tau_3}+\dots\right)+\dots\Bigg]\\
=&-\frac{1}{2}\frac{a\beta}{\alpha}\Bigg[\left(\frac{1}{F(\lambda_0)}+(a^2\beta^2)\frac{1}{F(\lambda_1)}+(a^2\beta^2)^2\frac{1}{F(\lambda_2)}+\dots\right)\\
&+\beta\left(\frac{1}{F^2(\lambda_0)}+(a^2\beta^2)\frac{1}{F^2(\lambda_1)}+(a^2\beta^2)^2\frac{1}{F^2(\lambda_2)}+\dots\right)\\
&+\beta^2\left(\frac{1}{F^3(\lambda_0)}+(a^2\beta^2)\frac{1}{F^3(\lambda_1)}+(a^2\beta^2)^2\frac{1}{F^3(\lambda_2)}+\dots\right)+\dots\Bigg]\\
=&-\frac{1}{2}b_0\left[\sum_{t=1}^{\infty}\beta^t\sum_{k=0}^{\infty}\left(\frac{b_k}{b_0}\right)^2\frac{1}{F^t(\lambda_k)}\right] \qquad (20)
\end{aligned}
$$

where in the last line we've used the solution for b_t in (35). The term $\sum_{k=0}^{\infty}\left(\frac{b_k}{b_0}\right)^2\frac{1}{F^t(\lambda_k)}$ in (20) represents the contribution to the value function from the uncertainty the agent faces when she takes a decision t time steps in the future. $1/F^t(\lambda_k)$ is the agent's uncertainty about events that are $k+1$ 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 fact that b_k is the coefficient of μ_k in the optimal policy rule (11). T 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 $\tau_{t+k}, \tau_{t+k-1}, \dots, \tau_k$.

Given prior uncertainty $\vec{\lambda} = (\lambda_t)_{t\geq 1}$, the term $T(\vec{\tau})$ is a function of 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. The following two subsections consider two methods for extracting the information $T(\vec{\tau})$ contains about the relative importance of predictability at different lead times.

2.1 Valuing marginal predictability

To make initial progress we begin by finding a linear approximation to $T(\vec{\tau})$ at $\vec{\tau} = \vec{0}$. The linearized version of T has the advantage of being a separable function of τ_t, allowing the contribution of each τ_t to the value function to be computed independently. The results in this section hold only when the precision of predictions is small, or equivalently, when the interactions between forecast lead times in the value function can be neglected.

Proposition 3. *Define*

$$g(m) \equiv \sum_{k=0}^{\infty} \frac{\beta^k}{\lambda_{m+k}^2}, \tag{21}$$

and assume that

$$\lim_{t\to\infty} \frac{\lambda_{t+1}^2}{\lambda_t^2} > \beta, \tag{22}$$

implying that $g(m)$ is finite for all m. Then for small forecast precisions (i.e. $\tau_t = d\tau_t \ll 1$), the increase in the value function due to the prediction system (relative to an uninformative baseline) is approximately

$$dV = T(d\vec{\tau}) - T(\vec{0}) \approx \frac{a}{\alpha(1-a^2\beta)} \sum_{m=1}^{\infty} r_m d\tau_m, \tag{23}$$

where

$$r_m \equiv g(m)\beta^m \left(1 - (a^2\beta)^m\right). \tag{24}$$

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 (20) makes clear, the importance of a reduction in uncertainty of events k time steps in the future is proportional to $(a^2\beta^2)^k$. 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

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

The change in payoffs due to the uncertainty reduction from this forecast must also be proportional to $\frac{d}{d\lambda}\left(-1/\lambda_m\right) = 1/\lambda_m^2$. Thus, 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 (24).

The dependence of prior uncertainty on lead time can have an important influence on the value of predictions of 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 precision of prior beliefs about the location of θ_t is given by,

$$\lambda_t^2=\lambda_0^2\phi^t. \tag{25}$$

where $\phi\in[0,1]$ is a parameter, and λ_0 is the precision of beliefs about the current loss-minimizing decision θ_0. This implies that,

$$
\begin{aligned}
g(m)&=\sum_{k=0}^{\infty}\frac{\beta^k}{\lambda_{m+k}^2}\\
&=\frac{1}{\lambda_0^2\phi^m}\sum_{k=0}^{\infty}\left(\frac{\beta}{\phi}\right)^k.
\end{aligned}
$$

In order for $g(m)$ to be defined for all m (see (22)), we must restrict attention to $\phi>\beta$. Under this constraint, the ratio

$$\frac{g(m)}{g(1)}=\left(\frac{1}{\phi}\right)^{m-1}.$$

is uniquely defined. In this parametric model, we can thus define a 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:

$$R_m\equiv\frac{r_m}{r_1}=\left(\frac{\beta}{\phi}\right)^{m-1}\left[\frac{1-(a^2\beta)^m}{1-a^2\beta}\right] \tag{26}$$

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

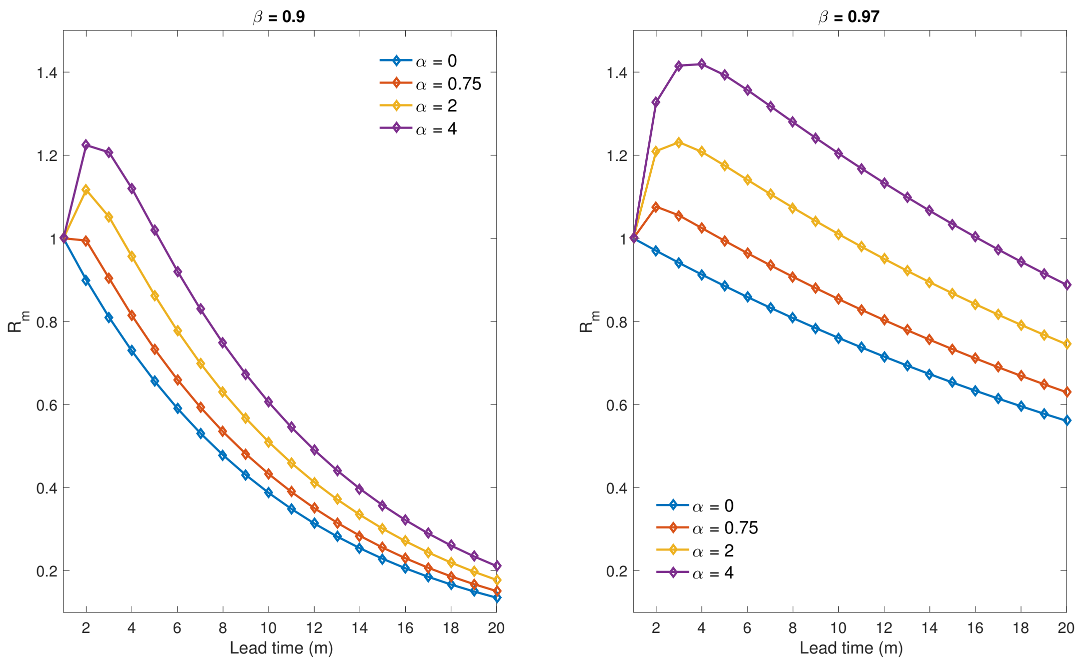


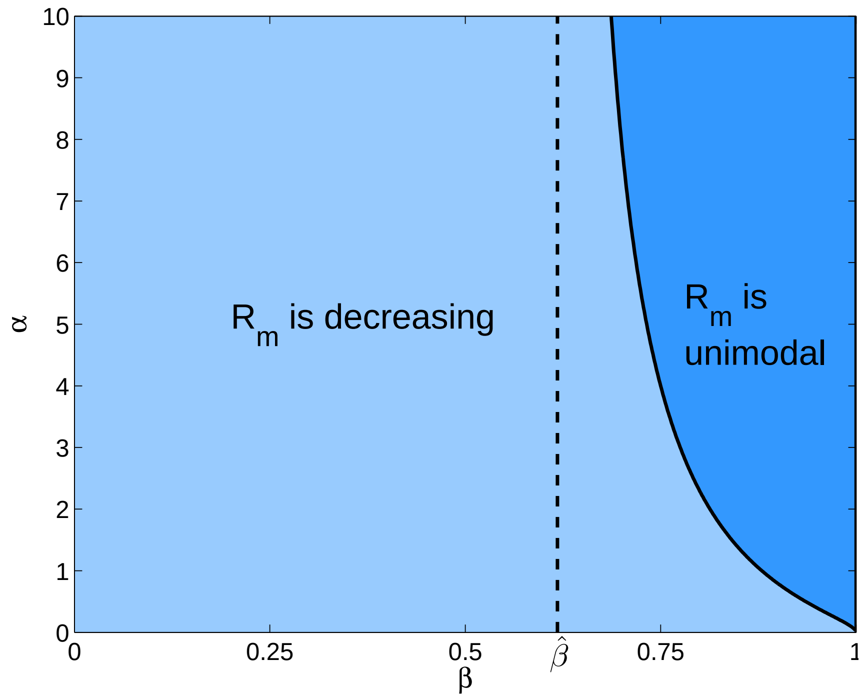
Figure 3: Dependence of R_m on adjustment costs (α) and discount factor (β). Prior uncertainty is assumed to be a constant function of time in this example (i.e. $\phi = 1$).

respectively.

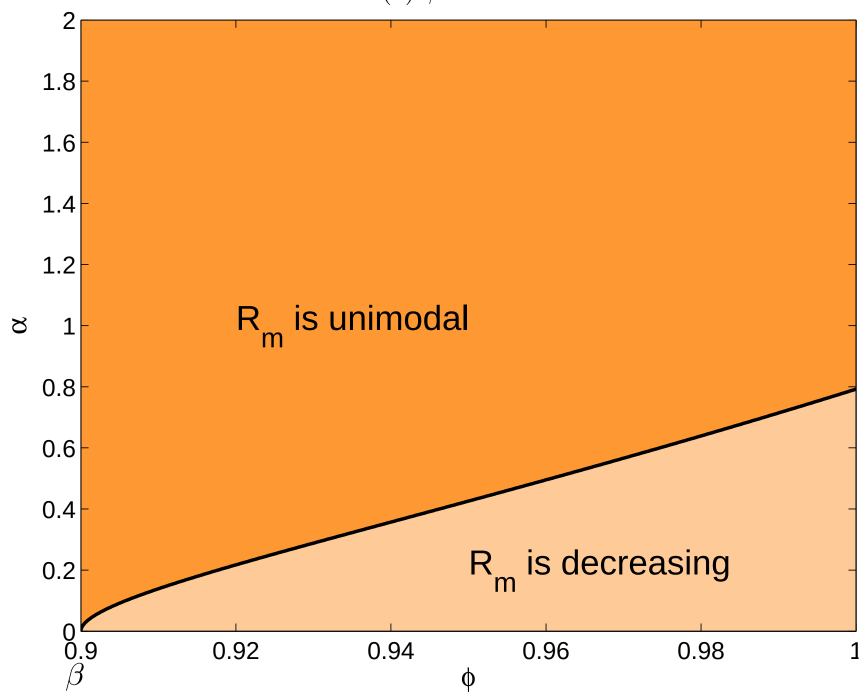
Some simple analysis (see Appendix D) shows that R_m can exhibit only two kinds of qualitative behaviour. Either it decreases monotonically with m, or it is a unimodal function with a global maximum at some $m \geq 2$. Figure 3 plots R_m for several values of the parameters. For low α, R_m is a declining function of m, indicating that agents who face low adjustment costs value short-run predictability more than long-run predictability. When α is sufficiently large however, R_m becomes a unimodal function.

Appendix D characterizes the regions of parameter space under which these two qualitative behaviours occur, demonstrating that the qualitative findings in Figure 3 extend to the entire parameter space of the model. 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. Appendix D derives analytic expressions for $\hat{\beta}$ and $\hat{\alpha}$. These 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 4 demonstrates these results graphically.

If R_m is decreasing, the most valuable forecast lead time is $m = 1$. Figure 4 shows that this occurs in a large region of parameter space, associated with lower values of β and α. Thus, in many decision environments, short-run predictability is more important than long-run predictability. The intuition for this finding is clear. When α is small the agent can



(a) $\phi = 1$



(b) $\beta = 0.9$

Figure 4: Qualitative behaviour of R_m in different regions of parameter space.

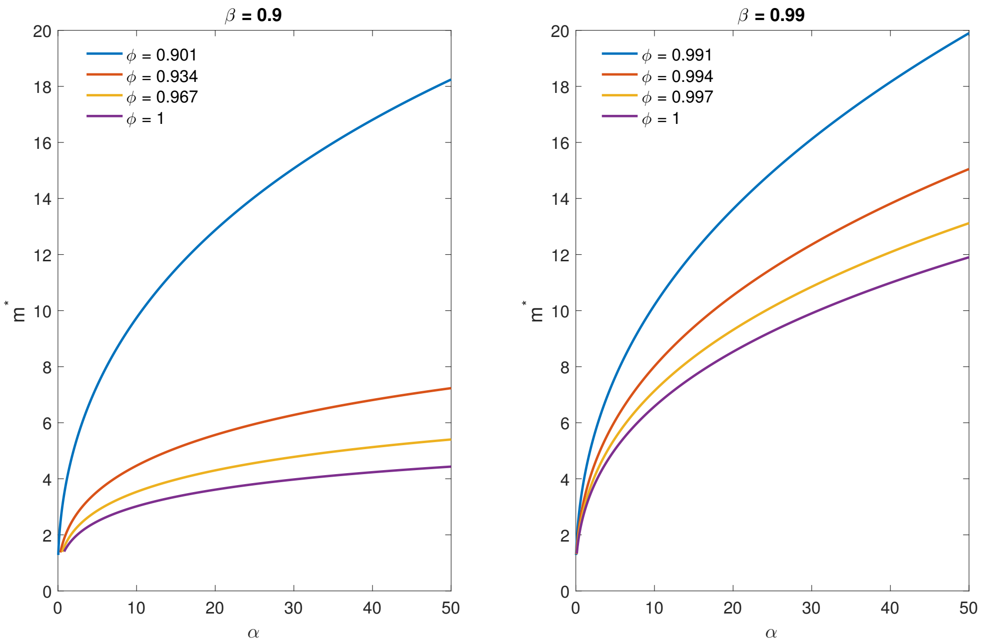


Figure 5: m^*, the most valuable forecast lead time when forecast precisions are small.

cheaply adjust her strategy to capitalize on current conditions. Since short run predictability improves her ability to capitalize on current conditions, she values this highly, but places comparatively little weight on long-run predictability.

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

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

m^* summarizes the forecast lead times that are most important to the decision-maker. The larger is m^*, the more important is long-run predictability. It is straightforward to show that $\frac{\partial m^*}{\partial \alpha} > 0$, $\frac{\partial m^*}{\partial \beta} > 0$ and $\frac{\partial m^*}{\partial \phi} < 0$. Figure 5 plots m^* as a function of α for several values of ϕ and β. The figure shows that even when α is very large (recall that $\alpha = 3$ is already a large value, see Figure 2), m^* is often fairly small . Thus, even when the agent is highly inflexible, she will often value short to medium-term predictability more highly than long-term predictability.

2.2 Optimal predictability distributions

The preceding analysis provides a characterization of the dependence of the agent's value function on predictability at different lead times when forecast precisions are small. In this section we move beyond these results, aiming to summarize the dependence of $T(\vec{\tau})$ on $\vec{\tau}$ when $\vec{\tau}$ is non-marginal. When forecast precisions are non-marginal, the interactions between predictability at different lead times in the value function cannot be neglected. It is intuitively clear that these interactions are potentially 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 a marginal 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 (20) 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 $\vec{\tau}$, and the value of an additional unit of precision in forecasts of events at lead time m is smaller the larger is the precision of forecasts of events at lead time k. Capturing the effects of the interactions between forecast lead times in the value function in an analytically tractable and intuitively appealing way is however difficult, as the function $T(\vec{\tau})$ has a complex dependence on its arguments.

In order to make progress we focus on the following question: what would an optimal distribution of predictability across lead times look like? At first sight this question may seem to be ill-posed, since the function $T(\vec{\tau})$ is strictly increasing in each of the τ_m, so clearly 'optimal' predictability occurs when $\tau_m \to \infty$ for all m. We can however formulate a more enlightening version of it as follows. Suppose that the decision-maker has a fixed budget of forecast precisions across all lead times, i.e. $\sum_{m=1}^{\infty} \tau_m = B$ for some constant $B > 0$. Given this constraint, how would the agent prefer this predictability budget to be distributed across different lead times? Since $T(\vec{\tau})$ is a concave function a unique answer to this question exists for any $B > 0$. Now recall that $\vec{\lambda} = (\lambda_1, \lambda_2, \ldots)$ is the sequence of precisions of prior beliefs about future states, and consider the following quantity:

$$\vec{\sigma} \equiv \lim_{\vec{\lambda} \to \vec{0}} \frac{1}{B} \left(\text{argmax}_{\vec{\tau}} T(\vec{\tau}) \text{ s.t. } \sum_{m=1}^{\infty} \tau_m = B \right). \tag{27}$$

The m-th component of $\vec{\sigma}$, denoted σ_m, is the share of the total predictability budget B that the agent would optimally allocate to lead time m, in the limit as the precision of prior beliefs about all future periods approaches zero. By definition, $\forall m, \sigma_m \in [0, 1]$ and $\sum_{m=1}^{\infty} \sigma_m = 1$. $\vec{\lambda} \to \vec{0}$ corresponds to the limiting case in which the agent's beliefs about the future are *entirely* determined by the prediction system.[8] Inspection of the expression (20) shows that for any $k \in \mathbb{R}$,

$$\lim_{\vec{\lambda} \to \vec{0}} T(k\vec{\tau}) = \lim_{\vec{\lambda} \to \vec{0}} \frac{1}{k} T(\vec{\tau}).$$

This scale invariance property implies that $\vec{\sigma}$ is independent of the value of the constraint B. $\vec{\sigma}$ may thus be interpreted as the optimal distribution of predictability across lead times, when prior beliefs are arbitrarily diffuse. The fact that $T(\vec{\tau})$ is a well-behaved concave function

[8] If $\vec{\lambda} = \vec{0}$, the value function V is undefined, however the limit in (27) exists since the solution of the optimization problem is continuous in the parameters $\vec{\lambda}$.

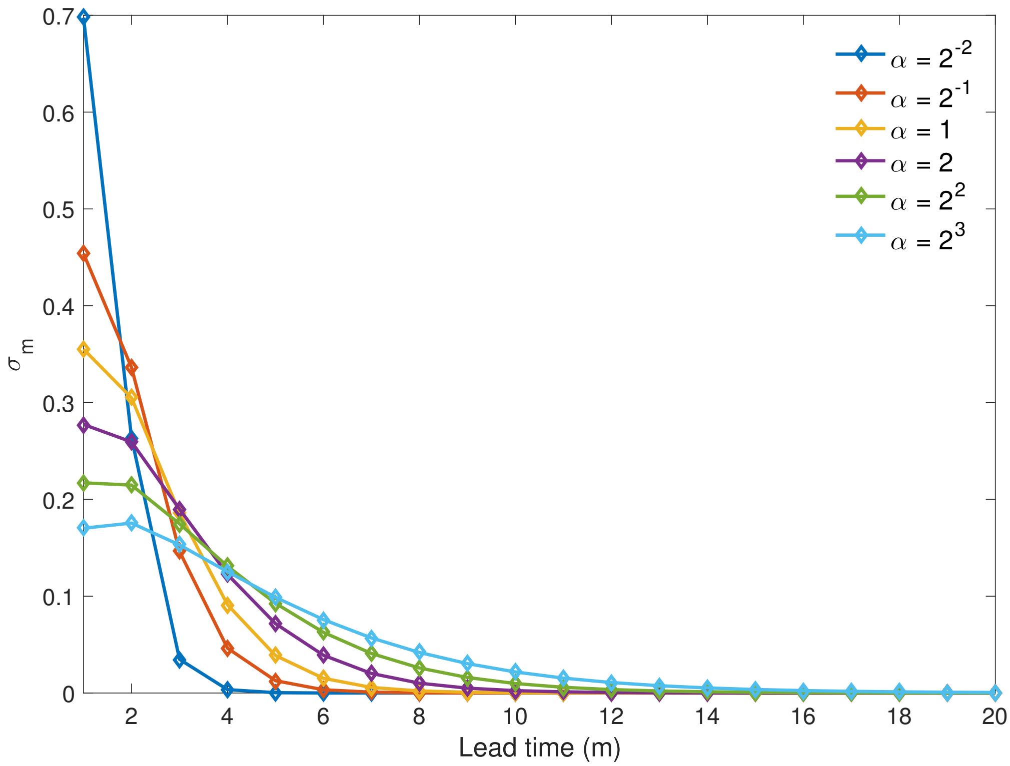


Figure 6: The optimal distribution of predictability across lead times $\vec{\sigma}$. $\beta = 0.95$ in this example.

means that it is straightforward to find an arbitrarily good approximation to $\vec{\sigma}$ using standard numerical constrained optimization routines. Figure 6 demonstrates the typical dependence of $\vec{\sigma}$ on α. Optimal predictability distributions either assign decreasing weight to larger lead times (small α), or are unimodal functions (large α), which are nevertheless skewed towards shorter lead times.

To understand how the distributions in Figure 6 capture the interactions between forecast lead times it is helpful to consider how this figure would change in a different limiting case of the model in which these interactions are effectively switched off. Suppose that instead of examining optimal budget allocations in the limit as $\vec{\lambda} \to \vec{0}$, we allowed $\vec{\lambda}$ to be unconstrained, but considered the limit as the predictability budget $B \to 0$. In this case, all forecast precisions would be infinitesimally small, implying that the value function is linear in τ_m for all m, and is given by the expression in (23). Since the value function is linear in the τ_m in this case, it is optimal to assign the entire budget B to the forecast lead time m for which the coefficient of τ_m, i.e. r_m in (23), is largest. Thus if the interactions between forecast lead times were negligible, the agent would focus all her attentions on the most valuable lead time. When interactions between lead times are non-negligible however, a more diffuse distribution of predictability across all lead times becomes desirable. The distributions in Figure 6 may thus be thought of as a non-marginal generalization of the most valuable lead time m^* depicted in Figure 5.

Figure 6 demonstrates that although the marginal analysis in Section 2.1 provides a tractable first order estimate of the value of predictability as a function of lead time, interactions between forecast lead times are likely to be quantitatively important, and accounting for them leads to an even greater emphasis on short-run predictability. While Figure 3 might seem to suggest that predictability at lead times of 10-20 time units still carries a good amount of value, these estimates neglect interaction effects. The optimal predictability distributions in Figure 6 account for these interactions, leading them to place very little weight on lead times beyond 10 time units for a wide range of adjustment costs. This finding is easily understood. Since short-run predictability can partially substitute for long-run predictability, in addition to providing information relevant to near-term conditions, accounting for the interactions between lead times tips the balance further towards the short run.

3 Discussion

The model we have developed is a conceptual tool that allows us to compute decision-maker's induced preferences over alternative prediction systems, i.e. forecasting products with a fixed profile 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 once it becomes available. The essential novel feature of our model is that it allows us to compute the contribution of predictive accuracy at each lead time to the overall value of a forecasting product, enabling a study of the relative importance of short- and long-run predictability that is, we believe, novel in the literature.

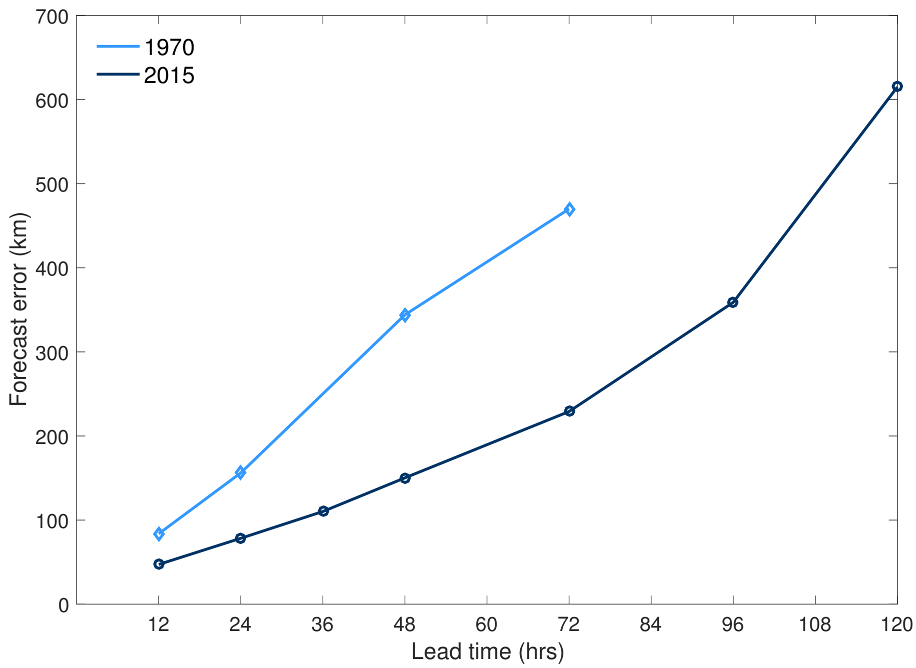
The model provides potentially important conceptual lessons for decision-makers. The tendency towards excessive emphasis on the importance of predicting the long-run is well illustrated by an anecdote related by Kenneth Arrow about his time as a military weather forecaster during World War Two (Arrow, 1991):

'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.'

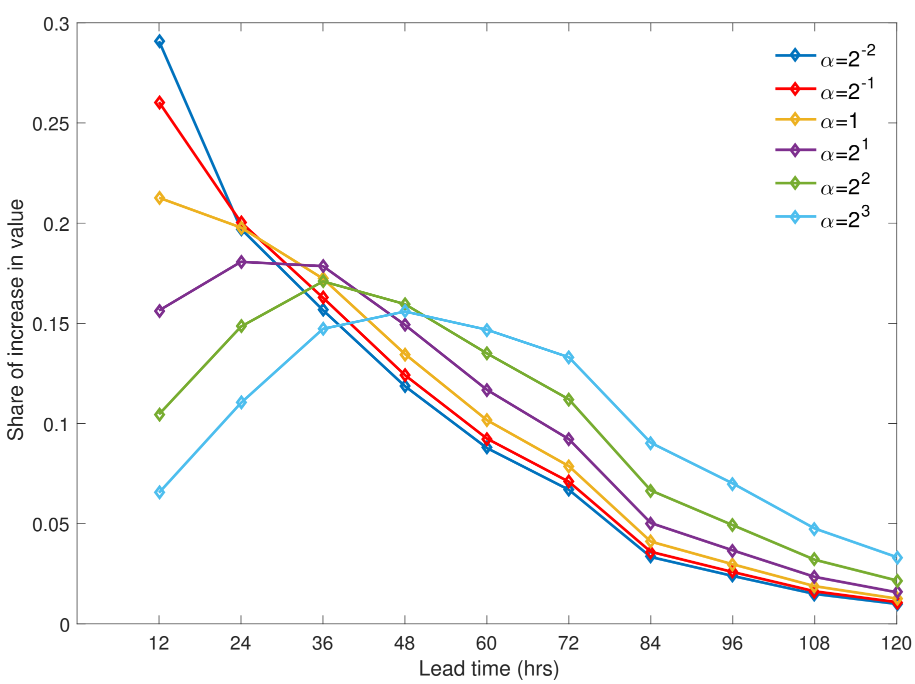
The General's implicit conflation of the presence of long-run uncertainty with the need for long-run predictions – no matter how imperfect – can also be found in current policy discussions. In December 2016 the US congress passed the Weather Research and Forecasting Innovating Act, which "aims to bolster the capacity of the National Oceanic and Atmospheric Administration (NOAA) to make seasonal weather predictions between 2 weeks and 2 years out" (Voosen, 2016). While improved seasonal predictions (i.e. 3 months to a year out) could doubtless improve decisions with a natural seasonal time scale (e.g. planting decisions), it is less clear *a priori* what the benefits of improving the predictability of weather events say 2-4 weeks out will be. We illustrate this point by examining historical advances in predictability on these temporal scales.

Figure 7a depicts advances in the accuracy of hurricane track predictions between 1970 and 2015. The figure shows there has been remarkable scientific progress in this domain, with predictability increasing at all lead times, but especially at longer lead times. To attribute changes in the value of hurricane forecasts to error reductions at each lead time we use the data in Figure 7a to calibrate the values of two forecast precision vectors, $\vec{\tau}_{2015}$ and $\vec{\tau}_{1970}$. The contribution of changes in each component of $\vec{\tau}$ to the change in the value of the prediction system, i.e. $T(\vec{\tau}_{2015}) - T(\vec{\tau}_{1970})$, can be estimated using an analogue of the Shapley value from cooperative game theory (Shapley, 1953; Shorrocks, 2013, see Appendix E for details). Figure 7b plots an estimate of the share of the change in value attributable to error reductions at each lead time between 12 and 120 hours. Although error reductions are significantly larger for longer lead times (e.g. 48 hours and beyond), much of the increase in value is attributable to comparatively minor changes in short run predictability when decision-makers are moderately flexible (e.g. $\alpha \leq 1$). One needs large values of α (e.g. $\alpha > 4$) for improvements in predictability at lead times greater that 48 hours to play a significant role, and even then the substantial improvements beyond 96 hours account for less than 15% of the increase in value. If these large improvements at lead times of 5 days contribute little to the change in value, how much impact can we expect even very significant changes in our ability to forecast events 2-4 weeks out to have?

The assumption that long-run predictability is necessary for successful decision-making is also prevalent in discussions of policies that aim to facilitate adaptation to climate change. 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 anticipatory planning. Füssel (2007) contends that 'the effectiveness of pro-active adaptation to climate change often depends on the accuracy of regional climate and impact projections, which are subject to substantial uncertainty', 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 extreme events than are



(a) Increase in the accuracy of Atlantic hurricane track forecasts, 1970-2015. Forecasts with lead times greater than 72 hours were introduced for the first time in 2001. Data from the National Hurricane Centre at NOAA, see Appendix E.



(b) Share of increase in forecast value attributable to each lead time ($\beta = (0.95)^{\frac{1}{2\times 365}}$).

Figure 7: Valuing improvements in the predictability of Atlantic hurricanes

possible with the current generation of climate models' (Dessai et al., 2009). Taken together these quotes suggest that it is all too easy to conflate the presence of substantial long-run uncertainty with the need for long-run predictability. Since the long-run today will become the short-run in the future, successful adaptation to a changing climate is not necessarily contingent on our ability to predict the long run. Our model suggests that even if accurate long-run forecasts were available, accounting for the continually updated nature of decision-making and prediction means that in many circumstances long-run predictability contributes relatively little to decision-makers' expected payoffs.

While our analysis has focussed on the benefits of predictability, we expect our conclusions to be reinforced by an analysis of the costs of improving predictability at different lead times. The marginal cost of a unit increase in the predictability of events at lead time m is likely to be an increasing function of m in most cases. While short-run predictability may often be increased by reducing measurement errors in initial conditions (e.g. by expanding the observation network, in the case of weather) or other marginal statistical improvements, long-run uncertainty is often dominated by model misspecification errors, which reflect our lack of understanding of the structural dynamics of natural or social phenomena (e.g. Hawkins & Sutton, 2009). These deeper uncertainties often yield only to new scientific breakthroughs or methodologies, making reducing them both riskier, and more costly than reductions in short-run uncertainty. Although increasing long-run predictability may serve scientific goals, the instrumental consequences of such improvements for decision-makers are less clear, and depend strongly on the costs incurred when adapting activities to new information.

Although we believe that our model provides important conceptual insights into the determinants of agents' 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 payoff 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 some 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. Extending the model to more complex cases in which this symmetry is broken will be a technical challenge, since the infinite dimensional stochastic control problem we study will no longer be analytically tractable in this case. Nevertheless, this would naturally be of interest in future applications. The limitations of the quadratic payoff function are well-known, and we will not reprise them here. Suffice to say that this choice proves analytically convenient, and we believe it captures the essence of the problem we study.

Second the model does not consider correlations between signals at different lead times. This is by design. There is no clear notion of predictability as a function of lead time in a correlated model, as in this case all signals convey information about all future events. Indeed, we conceived of our model precisely so that we could overcome this problem. There is no great conceptual obstacle to allowing for correlated signals in our framework, however in practice this gives rise to analytical and interpretational difficulties. As in other things, we have opted for conceptual clarity over empirical comprehensiveness on this front.

Finally, we emphasize that the model is intended to represent a decision-maker who faces an *exogenously* changing environment. Thus, its conceptual lessons apply to individuals and firms, but less to large entities whose actions may 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

A. Angström (1919). 'Sannolikhet och praktisk väderleksprognos (Probability and practical weather forecasting)'. *Centraltryckeriet, Teknologföreningens Förlag* p. 11.

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.

S. Athey & J. Levin (1998). 'The Value Of Information In Monotone Decision Problems'. *MIT Department of Economics Working Paper No. 98-24* .

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.

A. Cabrales, et al. (2013). 'Entropy and the Value of Information for Investors'. *The American Economic Review* **103**(1):360–377.

M. P. Clements & D. F. Hendry (eds.) (2011). *The Oxford Handbook of Economic Forecasting*. OUP USA, New York.

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

M. Demers (1991). 'Investment under Uncertainty, Irreversibility and the Arrival of Information Over Time'. *The Review of Economic Studies* **58**(2):333.

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.

A. K. Dixit & R. S. Pindyck (1994). *Investment under uncertainty*. Princeton University Press, Princeton, N.J.

G. Elliot & A. Timmermann (eds.) (2013). *Handbook of Economic Forecasting*, vol. 2A-B. North-Holland.

G. Elliott, et al. (eds.) (2006). *Handbook of Economic Forecasting*, vol. 1. North-Holland.

G. Elliott & A. Timmermann (2008). 'Economic Forecasting'. *Journal of Economic Literature* **46**(1):3–56.

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

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.

C. W. J. Granger & M. J. Machina (2006). 'Forecasting and Decision Theory'. In G. Elliott, C. W. J. Granger, & A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 1, pp. 81–98. Elsevier.

E. Hawkins & R. Sutton (2009). 'The potential to narrow uncertainty in regional climate predictions'. *Bulletin of the American Meteorological Society* **90**:1095–1107.

R. Katz & A. Murphy (1997). *Economic value of weather and climate forecasts*. Cambridge University Press, Cambridge, United Kingdom, 1 edn.

E. L. Lehmann (1988). 'Comparing Location Experiments'. *The Annals of Statistics* **16**(2):521–533.

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.

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

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

R. R. Nelson & S. G. Winter (1964). 'A Case Study in the Economics of Information and Coordination: The Weather Forecasting System'. *The Quarterly Journal of Economics* **78**(3):420–441.

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

N. Persico (2000). 'Information Acquisition in Auctions'. *Econometrica* **68**(1):135–148.

L. S. Shapley (1953). 'A Value for n-person Games'. In *Contributions to the Theory of Games*, pp. 307–317. Annals of Mathematical Studies, Princeton University Press.

A. F. Shorrocks (2013). 'Decomposition procedures for distributional analysis: a unified framework based on the Shapley value'. *The Journal of Economic Inequality* **11**(1):99–126.

N. N. Taleb (2008). *The Black Swan: The Impact of the Highly Improbable*. Penguin, London.

P. Voosen (2016). 'U.S. Congress wants to know the weather weeks ahead'. *Science* **354**(6317):1212–1212.

A Proof of Proposition 1

We use the Bellman equation (10) to solve for the optimal policy function $X_{n+1} = \pi(X_n, Y_n)$. Let primed variables denote next period quantities, and unprimed variables denote current period quantities, i.e. $W = W(X', X, Y)$ and $V = V(X, Y)$. Then the first order condition for X_{n+1} is

$$\frac{\partial W}{\partial X'}(\pi(X_n, Y_n), X_n, Y_n) + \beta \int \frac{\partial V}{\partial X}(\pi(X_n, Y_m), F(Y_n, S))Q(S; Y_n)dS = 0. \tag{28}$$

By the envelope theorem,

$$\frac{\partial V}{\partial X}(X_n, Y_n) = \frac{\partial W}{\partial X}(\pi(X_n, Y_n), X_n, Y_n) \tag{29}$$

From (6), we have

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

Substituting into (28), we find that the policy rule must satisfy

$$\mu_0 + \alpha X_n - (1+\alpha)\pi(X_n, Y_n) + \beta \int \left[\alpha(\pi(\pi(X_n, Y_n), F(Y_n, S)) - \pi(X_n, Y_n))\right] Q(S; Y_n)dS = 0. \tag{30}$$

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 (30) we find:

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

Since $E_{s_t}\mu_t'(s_{t+1}) = \mu_{t+1}$, we can simplify this to:

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

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:

$$\alpha\beta a^2 - (1+\alpha(1+\beta))a + \alpha = 0 \tag{31}$$
$$\Rightarrow a = \frac{1+\alpha(1+\beta) \pm \sqrt{(1+\alpha(1+\beta))^2 - 4\alpha^2\beta}}{2\alpha\beta} \tag{32}$$

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

$$a = a(\alpha, \beta) = \frac{1+\alpha(1+\beta) - \sqrt{(1+\alpha(1+\beta))^2 - 4\alpha^2\beta}}{2\alpha\beta} \tag{33}$$

The equation for b_0 is:

$$1 - (1+\alpha)b_0 + a\beta\alpha b_0 - \alpha\beta b_0 = 0$$
$$\Rightarrow b_0 = \frac{1}{1+\alpha+\alpha\beta(1-a)}. \tag{34}$$

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

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

Thus for all $t \geq 0$,

$$b_t = \frac{1}{1+\alpha+\alpha\beta(1-a)}\left[\frac{\alpha\beta}{1+\alpha+\alpha\beta(1-a)}\right]^t \tag{35}$$

We can simplify this further by using the equation for a in (31). Define

$$\Lambda \equiv 1+\alpha+\alpha\beta(1-a) \tag{36}$$

From (31) 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

$$b_t = \frac{a}{\alpha}\left(a\beta\right)^t. \tag{37}$$

We now prove the properties of the coefficients a, b_t stated in 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 (33) we have

$$\frac{\partial a}{\partial \alpha} = -\frac{1}{2}\frac{-\alpha\beta+\sqrt{\alpha^2(1-\beta)^2+2\alpha(1+\beta)+1}-\alpha-1}{\alpha^2\beta\sqrt{\alpha^2(1-\beta)^2+2\alpha(1+\beta)+1}}\ . \tag{38}$$

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 (37) 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 (31).

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 (37) 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.

B 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

$$V(X,Y) = kX^2 + \sum_{t=0}^{\infty} c_t \mu_t X + \sum_{t=0}^{\infty}\sum_{p=t+1}^{\infty} D_{t,p}\mu_t\mu_p + \sum_{t=0}^{\infty} d_t \mu_t^2 + \sum_{t=0}^{\infty}\sum_{i=0}^{\infty} \frac{f_{i,t}}{\lambda_t + h_{i,t}}. \qquad (39)$$

All except the last term of this expression are straightforward to guess simply by inspection of the formula for the period payoff in (6). 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 $(\tau_t)_{t\geq 0}$.

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_{t+1}^2 in the second term will cancel to 1 (as occurs in the case $t \neq p$), and we are left with

$$\mathbf{E}_{s_{t+1}} \mu'_t(s_{t+1}) \mu'_t(s_{t+1}) = \frac{\tau_{t+1}}{\lambda_{t+1}(\lambda_{t+1} + \tau_{t+1})} + \mu_{t+1}^2. \tag{40}$$

Hence, in summary:

$$\mathbf{E}_{s_{t+1},s_{p+1}} \mu'_t(s_{t+1}) \mu'_p(s_{p+1}) = \begin{cases} \mu_{t+1} \mu_{p+1} & t \neq p \\ \frac{\tau_{t+1}}{\lambda_{t+1}(\lambda_{t+1}+\tau_{t+1})} + \mu_{t+1}^2 & t = p. \end{cases} \tag{41}$$

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

$$\frac{\tau_{t+1}}{\lambda_{t+1}(\lambda_{t+1}+\tau_{t+1})} = \frac{1}{\lambda_{t+1}} - \frac{1}{\lambda_{t+1}+\tau_{t+1}}. \tag{42}$$

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

$$\sum_{i=0}^{\infty}\sum_{t=0}^{\infty}\frac{f_{i,t}}{\lambda_t + h_{i,t}} = -\frac{1}{2}\frac{1}{\lambda_0} + \beta\left(\sum_{t=0}^{\infty} d_t\left[\frac{1}{\lambda_{t+1}} - \frac{1}{\lambda_{t+1}+\tau_{t+1}}\right] + \sum_{i=0}^{\infty}\sum_{t=0}^{\infty}\frac{f_{i,t}}{\lambda_{t+1}+\tau_{t+1}+h_{i,t}}\right) \tag{43}$$

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 (43) is to hold as an identity for all λ_t, τ_t we require

$$f_{0,0} = -\frac{1}{2} \tag{44}$$

$$f_{0,t} = \beta d_{t-1} \text{ for } t \geq 1. \tag{45}$$

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 (43). 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:

$$h_{0,t} = 0; \quad h_{i,t} = \tau_t + h_{i-1,t-1} \qquad i \geq 1 \tag{46}$$

$$f_{0,0} = -\frac{1}{2}; \quad f_{0,t} = \beta d_{t-1} \qquad t \geq 1 \tag{47}$$

$$f_{1,0} = 0; \quad f_{1,t} = \beta[-d_{t-1} + f_{0,t-1}] \qquad t \geq 1 \tag{48}$$

$$f_{i,0} = 0; \quad f_{i,t} = \beta f_{i-1,t-1}. \qquad n \geq 2, t \geq 1. \tag{49}$$

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:

$$\mathbf{f} = \begin{pmatrix} -\frac{1}{2} & \beta d_0 & \beta d_1 & \beta d_2 & \beta d_3 & \dots \\ 0 & -\beta(d_0 + \frac{1}{2}) & \beta(\beta d_0 - d_1) & \beta(\beta d_1 - d_2) & \beta(\beta d_2 - d_3) & \dots \\ 0 & 0 & -\beta^2(d_0 + \frac{1}{2}) & \beta^2(\beta d_0 - d_1) & \beta^2(\beta d_1 - d_2) & \dots \\ 0 & 0 & 0 & -\beta^3(d_0 + \frac{1}{2}) & \beta^3(\beta d_0 - d_1) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots \end{pmatrix} \tag{50}$$

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.[9]

Clearly $f_{i,t} = 0$ for any $i > t$. Thus the only parameters $h_{i,t}$ that are relevant have $0 \leq i \leq t$. It is straightforward to solve the recurrence relation (46) to find

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

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.

[9]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.

Define

$$\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{51}$$

Then the Bellman equation for the μ_t^2 terms yields

$$\begin{aligned} d_t &= -\frac{1}{2}[(\alpha+1)(b_t)^2 + (1-2b_0)\delta_{t,0}] + \beta(k(b_t)^2 + c_{t-1}b_t(1-\delta_{t,0}) + d_{t-1}(1-\delta_{t,0}) \\ &= \left(k\beta - \frac{1}{2}(\alpha+1)\right)(b_t)^2 - \frac{1}{2}(1-2b_0)\delta_{t,0} + \beta c_{t-1}b_t + \beta d_{t-1} \end{aligned} \tag{52}$$

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:

$$\begin{aligned} k &= -\frac{1}{2}((1+\alpha)a^2 + \alpha - 2a\alpha) + \beta(ka^2) \\ \Rightarrow k &= -\frac{1}{2}\left(\frac{(1+\alpha)a^2 + \alpha - 2a\alpha}{1-\beta a^2}\right), \end{aligned} \tag{53}$$

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 (31) 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 (31) 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

$$c_t = a(a\beta)^t \tag{54}$$

for all $t \geq 0$.

Equation (52) 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 (35) and the definition of Λ in (36) 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

$$d_t = \frac{1}{2}\beta^t\left[\frac{a}{\alpha} - 1 + \frac{a}{\alpha}\frac{a^2\beta}{1-a^2\beta}(1-(a^2\beta)^t)\right]. \tag{55}$$

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 $\mathbf{f}$.

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

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

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.

C Proof of Proposition 3

From 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 (46),

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

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 $\mathbf{f}$, 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 (55) 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.

D Behaviour of R_m

From the formula (26), 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

$$(a^2\beta)^m \ln\left(\frac{a^2\beta^2}{\phi}\right) = \ln\left(\frac{\beta}{\phi}\right). \tag{56}$$

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 occurs. 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 $R_2 > 1 \iff \Delta \equiv a^2\beta^2 + \beta - \phi > 0$. Since $a = 0$ at $\alpha = 0$, we know $\Delta = \beta - \phi < 0$ when $\alpha = 0$. Also, since a is increasing in α, so is Δ. Combining these facts we see that the set of parameters values for which $\Delta > 0$ 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 $\Delta = 0$. Solving the condition $\Delta = 0$ 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

$$\hat{\alpha}(\beta, \phi) = \frac{(\phi - \beta)(1 + \beta) + \phi\sqrt{\phi - \beta}}{\beta^2 + (\phi - \beta)^2 - (\phi - \beta)(1 + \beta^2)} \ . \tag{57}$$

Observe that $\hat{\alpha}(0, \phi) = \frac{\phi(1+\sqrt{\phi})}{\phi(\phi-1)} < 0$ so Δ 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

$$\hat{\beta}(\phi) = \frac{\sqrt{1 + 4\phi} - 1}{2} \ . \tag{58}$$

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.

E Valuing improvements in the predictability of hurricane tracks

Figure 8 depicts data from the US National Hurricane Center on the history of Atlantic hurricane track forecast errors (`http://www.nhc.noaa.gov/verification/pdfs/1970-present_OFCL_ATL_annual_trk_errors_noTDs.pdf`). The dots represent the forecast errors for different lead times from 1970 to 2015 together with a linear trend for each lead time. 36 hr forecasts were first reported in 1988, and 96 and 120 hr forecasts were first reported in 2001. The forecast errors in 1970 and 2015 displayed in Figure 7a are calculated based on the linear trend at each lead time, and missing data at intermediate lead times in 1970 have been linearly interpolated from the 1970 estimates.

Determining the value $T(\vec{\tau})$ of a forecast system τ requires the sequence of precisions $\vec{\tau} = (\tau_t)_{t\geq 1}$. The time unit is 12 hrs. Precision sequences are calibrated based on the procedure described above, with precision levels beyond the forecast horizon (>72 hrs before 2001, >120 hrs after 2001) set to a minimum precision level $\tau_{\min}$. We choose $\tau_{\min} = 1/e_{\max}^2$, where $e_{\max} = 1276.4$ km is the highest forecast error in the data set. In the following we assume that the decision-maker's priors are arbitrarily diffuse, i.e. $\lambda_t \to 0$ for all t. This makes the function $T(\vec{\tau})$ scale invariant, meaning that the units in which we measure forecast errors are unimportant for computing relative changes in value. Our results apply when the loss minimizing decision $\tilde{\theta}$ is any linear function of the hurricane's physical location.

We use the Shapley value to estimate the contribution of error reductions at each lead

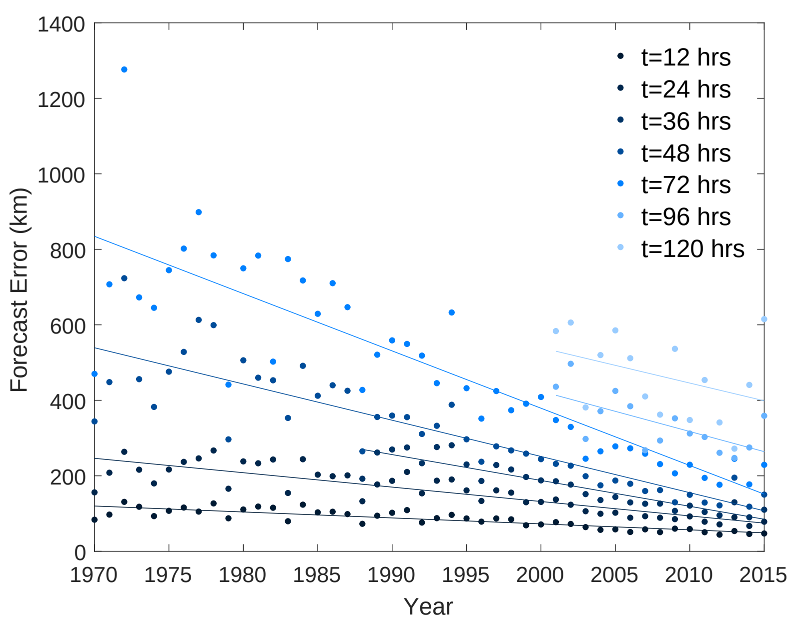


Figure 8: Historical errors in Atlantic hurricane track forecasts.

time to the overall change in value. Consider two vectors $\vec{\tau}$ and $\vec{\tau}'$ whose elements differ a finite number of times. Let the set of indices where the two vectors differ be A, with generic index i, and define the vector

$$\vec{\Delta} \equiv \vec{\tau}' - \vec{\tau}.$$

The ith non-zero element of $\vec{\Delta}$ is denoted by Δ_i. Without loss of generality, assume that $T(\vec{\tau}') > T(\vec{\tau})$. To disentangle the contribution of Δ_i to the overall change in forecast value, we view the vector $\vec{\tau}'$ as a 'coalition' of changes in the value of $\vec{\tau}$ at each lead time, which is assembled one by one. For any permutation π of the indices in A, let $\vec{\tau}_{\pi,i}$ be a precision vector in which the elements that precede i in the permutation π are given by the elements of $\vec{\tau}'$, with the remaining elements of $\vec{\tau}_{\pi,i}$ given by the corresponding elements of $\vec{\tau}$. Then the share of the change in value attributable to lead time i is defined to be

$$s_i \equiv \frac{1}{|A|!} \sum_{\pi \in P(A)} \frac{T(\vec{\tau}_{\pi,i} + \vec{\Delta} \cdot \mathbf{1}_i) - T(\vec{\tau}_{\pi,i})}{T(\vec{\tau}') - T(\vec{\tau})}, \tag{59}$$

where $P(A)$ is the set of permutations of the indices in A, and $\mathbf{1}_i$ is a unit vector with all elements equal to zero except element i. In words, we compute the marginal contribution of the change in $\vec{\tau}$ in component i in each possible way of assembling the new vector $\vec{\tau}'$ from discrete changes in the components of $\vec{\tau}$, and average the results. This procedure inherits the well-known properties of the Shapley value, and in particular is agnostic as to the order in which the changes in each component of $\vec{\tau}$ are applied so as to construct the new precision vector $\vec{\tau}'$.