American Economic Journal: Microeconomics 2021, 13(1): 374–398
https://doi.org/10.1257/mic.20180240

Prediction: The Long and the Short of It†

By Antony Millner and Daniel Heyen*

Commentators often lament forecasters' inability to provide precise predictions of the long-run behavior of complex economic and physical systems. Yet their concerns often conflate the presence of substantial long-run uncertainty with the need for long-run predictability; short-run predictions can partially substitute for long-run predictions if decision-makers can adjust their activities over time. So what is the relative importance of short- and long-run predictability? We study this question in a model of rational dynamic adjustment to a changing environment. Even if adjustment costs, discount factors, and long-run uncertainty are large, short-run predictability can be much more important than long-run predictability. (*JEL* D21, D81, D83)

Scientific progress over the past four hundred years has rendered a staggering range of phenomena more predictable. Atmospheric scientists forecast the weather, epidemiologists predict the spread of infectious diseases, macroeconomists forecast economic growth, and demographers predict population change. Yet despite many successes, reliable predictions of the long-run behavior of complex social or natural systems often remain elusive (Granger and Jeon 2007, Palmer and 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 (Grupe and Nitschke 2013). Concomitantly, improving long-run predictability is often a major goal of the scientific communities that produce forecasts. But just how important is it to be able to predict the distant future? Does substantial long-run uncertainty necessarily imply that accurate long-run predictions would be highly valuable? Or can long-run predictions be effectively substituted by short-run forecasts when decisions can be adjusted dynamically as new information arrives? This paper attempts to shed light on these questions.

* Millner: Department of Economics, North Hall, UC Santa Barbara, Santa Barbara, CA 93106 (email: amillner@econ.ucsb.edu); Heyen: Chair for Integrative Risk Management and Economics, ETH Zürich, Zürichbergstrasse 18, 8092 Zürich, Switzerland (email: dheyen@ethz.ch). Johannes Hörner was coeditor for this article. We are grateful to audiences at EEA, EAERE, Imperial College London, Heidelberg University, London School of Economics, University of Montpellier, UCSB, to Leo Simon, Larry Karp, Geoff Heal, and Derek Lemoine for valuable discussions, and to the CCCEP, Grantham Foundation, and the German Research Foundation (DFG) grant HE 7551/1 for support.
† Go to https://doi.org/10.1257/mic.20180240 to visit the article page for additional materials and author disclosure statement(s) or to comment in the online discussion forum.

It is not uncommon to find the presence of long-run uncertainty identified with the need for improved long-run predictions.[1] For example, a recent report by the National Academy of Sciences (2016, 17) on planned improvements in long-range weather forecasting suggests that “Enhancing the capability to forecast environmental conditions outside the well-developed weather timescale—for example, extending predictions out to several weeks and months in advance—could dramatically increase the societal value of environmental predictions, saving lives, protecting property, increasing economic vitality, protecting the environment, and informing policy choices.” Similarly, many commentators have suggested that the lack of reliable projections of the local impacts of climate change, most of which will occur many decades hence, is a significant barrier to effective adaptation planning. Füssel (2007, 265), for example, contends that “the effectiveness of pro-active adaptation to climate change often depends on the accuracy of [long run] regional climate and impact projections.” One can find a similar identification of the presence of long-run uncertainty with the importance of long-run predictions in economics. Lindh (2011, 585), for example, states that “Very long-run … forecasts of economic growth are required for many purposes in long-term planning. For example, estimates of the sustainability of pension systems need to be based on forecasts reaching several decades into the future.”

While one-shot decisions with fixed lead times between actions and outcomes (e.g., agricultural planting decisions) doubtless benefit from predictability at decision-relevant time scales, most long-run decision processes are at least partially flexible and can thus be adjusted over time. Firms or individuals who anticipate long-run changes in market conditions, regulation, or their physical environments will adjust their actions dynamically as new information becomes available. Similarly, governments concerned with policies that depend on conditions in the distant future (e.g., social security or adaptation to climate change) can alter the level of policy instruments dynamically as the future unfolds. The sequential nature of many long-run adjustment processes implies that there is no generic association between the presence of long-run uncertainty and the importance of long-run predictions. Since the long-run today will become the short-run tomorrow, short-run predictions can play an important role in informing decision-making, even when long-run uncertainty is large. Indeed, it is intuitive that short-run predictability is a perfect substitute for long-run predictability if adjustment is costless. In general, however, adjustment is costly, and large abrupt changes in response to short-run warnings are often significantly more costly than managed gradual transitions that may be informed by accurate long-run predictions. This suggests that long-run predictions could play an important role in informing anticipatory planning and avoiding excessive adjustment costs. It is, however, unclear a priori how the importance of

[1] An anecdote related by Arrow (1992) about his time as a military weather forecaster during World War II provides an extreme example:

> Some of my colleagues had the responsibility of preparing long-range weather forecasts, i.e., for the following month. The statisticians among us subjected these forecasts to verification and found they differed in no way from chance. The forecasters themselves were convinced and requested that the forecasts be discontinued. The reply read approximately like this: “The Commanding General is well aware that the forecasts are no good. However, he needs them for planning purposes.”

predictability at different lead times depends on the magnitude of adjustment costs. We develop a simple analytical model in which this question is answerable.

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

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

By exploring the dependence of V on its arguments, and the parameters of the decision problem, we quantify the value of predictability at different lead times. Our central finding is that if we account for sequential forecast updating and agents' ability to adjust their activities over time, short-run predictability can be more important than long-run predictability, even if adjustment costs, discount factors, and long-run uncertainty are large.

Although there is a sizable literature on the value of information and its role in dynamic decision-making, as far as we know there are few direct antecedents to the questions we seek to address in this paper. The literature on the value of information began with the pathbreaking work of Blackwell (1953) and Marschak and Miyasawa (1968), who defined an incomplete ordering of the "informativeness" of arbitrary information structures. We share this work's micro-oriented focus on the value of exogenous information sources for individual decision-makers but also differ from it in important respects. In order to ensure tractability, our model makes strong assumptions about the nature of agents' payoff functions and the

[2] Assuming convex adjustment costs is thus conservative with respect to adjudicating the importance of long-run predictability as this assumption *favors* long-run predictions. See the text following equation (21) for further discussion.

forecasts they receive. The return for this specificity is that we are able to study a much richer set of dynamic decisions than is typically used in this literature. Our focus on the dynamic characteristics of predictions, i.e., their accuracy as a function of lead time, is absent from this literature and necessitates a pared-down approach.

Work on the role of information in optimal dynamic decision-making falls into two categories: two-period models that examine the effect of second-period learning on optimal first-period decisions (e.g., Arrow and Fisher 1974; Epstein 1980; Gollier, Jullien, and Treich 2000), or infinite horizon models that involve learning about the realizations of a stochastic state variable (e.g., Merton 1971), or a parameter of a structural dynamic-stochastic model (e.g., Ljungqvist and Sargent 2004). Neither of these standard approaches can capture the effects we study here. Two-period models cannot capture the repeatedly updated nature of prediction and the dependence of predictability on lead time, both essential features of our model. Finite horizon models also suffer from an inherent bias toward short-run forecasts, as in a model with horizon H there will be H lead time 1 forecasts but only one lead time H forecast. On the other hand, models based on familiar stochastic processes, or learning about parameters of structural models, do not allow the accuracy of predictions at different lead times to be controlled independently, meaning that it is impossible to ask questions about the relative importance of short- and long-run predictability (see the discussion on p. 381 for an elaboration of this point). We thus need a different approach if we are to define a model that is tractable, unbiased, disentangles lead times, and nevertheless retains coarse features of dynamic prediction.

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

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

I. The Model

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

A. *Model Statement*

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

The decision-maker may adjust X in each period at a cost that is convex in the magnitude of the adjustment. Since large, abrupt changes in activities are more costly than gradual incremental shifts of equal magnitude, the decision-maker has an incentive to engage in anticipatory planning. The state of the world at time n, denoted by $\tilde{\theta}^n \in \mathbb{R}$, is the loss-minimizing action in that period, and we assume that the $\tilde{\theta}^n$ are independently (but *not* identically) distributed (we discuss this important assumption on p. 8). Throughout the paper, an n superscript on any variable always denotes a time index rather than an exponent. Values of X that are closer to $\tilde{\theta}^n$ are better adapted to conditions at time n and give rise to higher period payoffs. The decision-maker's choices must achieve a balance between exploiting current conditions (i.e., choosing X close to the current expected value of $\tilde{\theta}$) and preparing for future conditions (i.e., shifting X toward expected future values of $\tilde{\theta}$), thus avoiding excessively large and costly adjustments later on.

For any time n, let $\theta_t^n = \tilde{\theta}^{n+t}$ for $t \geq 0$, i.e., θ_t^n is the value of the loss-minimizing decision that will be realized t time steps in the future, in period n (see Figure 1 for

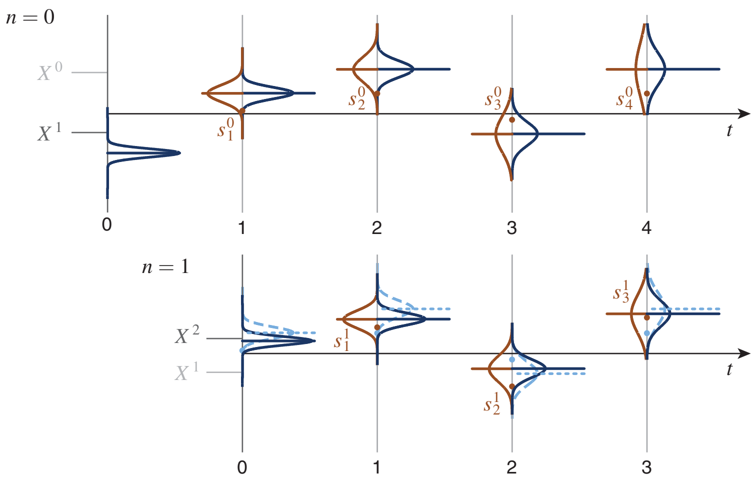

Figure 1. Illustration of the Model Setup

Notes: The figure depicts the agent's beliefs about future states (dark blue distributions), choices (location of X on the vertical axis), and the signals provided by the prediction system (brown dots), in the first two periods $n = 0, 1$. The current choice variable X^{n+1} is in black, and the inherited location X^n is in gray. The brown distributions at each t capture the agent's expectations about the signals she will receive at the end of the current period (i.e., $q_t^n(s_t^n; Y^n)$ in (7)). The dark blue distributions in the bottom line are updated beliefs at the beginning of period 1, after $S^0 = (s_t^0)_{t \geq 1}$ has been observed. The dashed light blue distributions and dots in the bottom line are the agent's previous beliefs and signals in period 0, for comparison. Smaller values of forecast precision 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 toward the realized signal.

an illustration). We denote the agent's beliefs about θ_t^n by $p_t^n(\theta_t^n)$. At $n = 0$, the agents' prior beliefs about the future values θ_t^0 are captured by an infinite sequence of normal distributions with means μ_t^0 and precisions (i.e., inverse variance) λ_t^0:

$$(1) \qquad \theta_t^0 \sim \mathcal{N}\left(\mu_t^0, 1/\lambda_t^0\right).$$

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

Let X^n be the value of the decision variable X that the agent inherits *at the beginning of period n*. At the beginning of the period, the agent chooses a new value for X, i.e., X^{n+1}. This is the value of X that will affect payoffs in the current period and be passed forward to the next period. The cost of modifying the decision variable from X^n to X^{n+1} is $(1/2)\alpha\left(X^{n+1} - X^n\right)^2$, where $\alpha \geq 0$ is a parameter that captures the magnitude of the adjustment costs the agent faces. After the choice of X^{n+1} is made, the agent experiences the realization of $\tilde{\theta}^n = \theta_0^n$ and sustains a loss equal to half the squared distance between X^{n+1} and θ_0^n. Thus, the expected period payoff at the beginning of the current period is given by

$$(2) \quad W\left(X^{n+1}, X^n, p_0^n(\theta_0^n)\right) = -\frac{1}{2}\left[\int_{-\infty}^{\infty}\left(X^{n+1} - \theta_0^n\right)^2 p_0^n(\theta_0^n)\, d\theta_0^n + \alpha\left(X^{n+1} - X^n\right)^2\right].$$

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

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

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

where

$$\epsilon_t^n \sim \mathcal{N}\left(0, 1/\tau_t\right), \qquad \operatorname{cov}\left(\epsilon_t^n, \epsilon_{t+k}^n\right) = 0 \quad \text{if } k \geq 1. \tag{4}$$

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

$$\vec{\tau} = \left(\tau_1, \tau_2, \tau_3, \ldots\right).$$

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

Notice that the agent receives a new sequence of forecasts S^n at the end of every period n. However, the precision of the information she receives about a particular value $\tilde{\theta}^k$ that lies in her future changes as time progresses and she moves closer to time k. Since the agent's prior beliefs about the future values θ_t^0 in period $n = 0$ are normal, and the conditional distributions of signals s_t^0 given states are normal, her beliefs about the future values of the states will update according to the standard normal-normal Bayesian formula (see, e.g., DeGroot 1970). At any time n beliefs about future values θ_t^n will be normally distributed and characterized by a sequence of means μ_t^n and precisions λ_t^n. Moreover, the agent knows that the beliefs she currently holds about the future values θ_t^n will become her beliefs about θ_{t-1}^{n+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

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

write down the state equations that describe how the forecasting system changes the agent's beliefs about the values of the states θ_t^n from one period to the next:

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

$$\lambda_t^{n+1} = \lambda_{t+1}^n + \tau_{t+1}.$$

As is standard in the normal-normal Bayesian updating model, the posterior mean of beliefs about each future value θ_t^n 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 $\left(X^n, Y^n\right)$, where

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

collects together the infinitely many "belief" state variables, and X^n is the value of X the agent inherits at the beginning of the period. The dynamics of Y^n are given by (5), and the next value of X is chosen directly by the agent. Figure 1 provides a graphical summary of the model setup and the timing of events.

Before proceeding to the solution of the model, we now discuss some of its more unusual assumptions and how they relate to existing work. The payoff structure in the model, captured by (2), is formally identical to that in previous models of dynamic adjustment we alluded to at the beginning of this section. The novelty in our approach arises from our representation of the decision-maker's dynamic expectations, i.e., the updating process summarized in (5). In existing applications of the dynamic adjustment model, the loss-minimizing decisions $\tilde{\theta}^n$ are always modeled as realizations of a serially correlated stochastic process. Sargent (1978), for example, assumes that $\tilde{\theta}$ follows an AR(1) process. However, such models are not suited to the central question of this paper, i.e., understanding the relative importance of short- and long-run predictability. To understand why, suppose that the $\tilde{\theta}^n$ are serially correlated and that a prediction system provides information about some $\tilde{\theta}^n$. Then learning about $\tilde{\theta}^k$ means that we learn something about *all* values of $\tilde{\theta}^n$. It is thus not usually possible to associate a prediction about an event at a given lead time with a change in uncertainty at only that lead time in a serially correlated model.[4] Our central question is thus unanswerable in serially correlated models unless the correlation structure can be tuned to generate arbitrary patterns of predictability as a function of lead time.[5]

[4] To see the difficulty in an explicit example, consider an AR(1) model: $\tilde{\theta}^{n+1} = \rho\tilde{\theta}^n + \nu^n$, where $\nu^n \sim \mathcal{N}\left(0, \xi^2\right)$, $\rho \geq 0$. In this case, $\tilde{\theta}^{n+t} \sim \mathcal{N}\left(\rho^t\tilde{\theta}^n, \left((1-\rho^t)/(1-\rho)\right)\xi^2\right)$. The only thing that can be done to change the predictability of the environment in this model is to change ρ or ξ; however, those changes affect predictability at all lead times at once.

[5] This could be possible, for example, in models where the state equation depends on many lagged variables (e.g., AR(n) models); however, one needs an infinite number of lags to disentangle all lead times, the relationship between parameters of the state equation and predictability at different lead times is often complex, and such models do not admit a separation between prior uncertainty and predictability. Costello, Adams, and Polasky (1998) use

By contrast, our model possesses four desirable features for our purposes. First, and most importantly, it disentangles lead times. There is a unique parameter τ_t associated with the accuracy of forecasts of lead time t; changing that parameter changes the accuracy of forecasts at that lead time *only*. This is a consequence of our assumption that the states of the world θ_t^n and signals s_t^n at different times t are independent. The decision-relevant interaction between periods in the model arises from the structure of the agent's payoffs (i.e., the presence of nonlinear adjustment costs) and not from serial correlations in the environment. Second, the model allows us to separate the agent's prior uncertainty (captured by the sequence of initial precisions λ_t^0) from the accuracy of the forecasting system (captured by $\vec{\tau}$). This will allow us to study how the value of predictability at different lead times depends on the agent's initial uncertainty about short- and long-run events and would not be possible in a standard structural model. Third, the model is "unbiased": because the time horizon is infinite, the agent receives an equal number of forecasts at each lead time. Fourth, the model retains coarse features of dynamic prediction (i.e., sequentially updated expectations, and predictions whose accuracy depends on lead time), while maintaining tractability and parsimony. The analytical cost of this approach is that instead of being able to describe dynamic expectations with a single state variable as in standard models, we now require an infinite number of state variables, two for each independent belief about each future event. In sum, while the independence assumption we deploy may seem unusual to some readers, it is a very convenient expediency if we are to generate conceptual insights into the relative importance of predictability at different lead times. We now turn to the model's solution.

B. *Solution*

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

$$(7) \qquad q_t^n(s_t^n; Y^n) = \int_{-\infty}^{\infty} \Pr(s_t^n | \theta_t^n) \, p_t^n(\theta_t^n) \, d\theta_t^n$$

$$\Rightarrow s_t^n | Y^n \sim \mathcal{N}(\mu_t^n, 1/\lambda_t^n + 1/\tau_t),$$

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

$$(8) \qquad Q(S^n; Y^n) = \prod_{t=1}^{\infty} q_t^n(s_t^n; Y^n).$$

an independence assumption that is similar to ours, and sequential event forecasts of the kind we consider have also been studied in, e.g., Clements (1997), Selten (1998), Regnier (2017).

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,

$$V(X^n, Y^n) = \max_{X^{n+1}} W(X^{n+1}, X^n, Y^n) + \beta \int_{\mathbb{R}^\infty} V(X^{n+1}, F(Y^n, S^n)) Q(S^n; Y^n)\, dS^n, \tag{10}$$

where $dS^n = \prod_{t=1}^{\infty} ds_t^n$, $\beta \in (0,1)$ is the agent's discount factor, and we have changed notation slightly to emphasize the dependence of the period payoff function (2) on the belief state variables Y^n. Note that the dependence of the value function on the profile of forecast precisions $\vec{\tau}$ comes both through the updating rule $F(Y^n, S^n)$ (see equation (5)) and through the agent's expectations about the values of future forecast realizations (see equation (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^n. Despite the infinite dimensionality of the state space in our model, standard methods based on the Benveniste-Scheinkman condition (Benveniste and Scheinkman 1979) yield simple analytic solutions for the optimal control rule. We state this rule in some detail as it will help us to interpret the main results below. From now on, we suppress the superscript n time index when there is no possibility of confusion from doing so. All proofs can be found in the appendices.

PROPOSITION 1: *The optimal policy* $X^{n+1} = \pi(X^n, Y^n)$ *is given by*

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

where

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

$$b_t = \frac{a}{\alpha}(a\beta)^t.$$

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

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

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

$$\lim_{\alpha \to 0} b_t = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{if } t > 0, \end{cases} \quad \lim_{\alpha \to \infty} b_t = 0,$$

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

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

The coefficients of the policy rule have an intuitive dependence on the adjustment cost 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 $a + \sum_{t=0}^{\infty} b_t = 1$, consider the case in which $\mu_t = X$ for all t. In this case, the agent believes that her choice is perfectly adapted to conditions now and in the future, and she should thus not want to change X. This occurs if $aX + \sum_{t=0}^{\infty} b_t X = X$.

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

Value Function.—In order to understand the effect of the precision sequence $\vec{\tau}$ on the agent's expected payoffs, we need to compute the value function. This would seem to be difficult as the model's state space is infinite dimensional, the period payoff depends nonquadratically[6] on the precision state variables λ_t, and we need to take expectations of the value function over an infinite sequence of signals, the distribution of which depends on all the belief state variables. Nevertheless, it is possible to obtain an analytic expression for the value function, which enables the remainder of our analysis.

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

$$\Delta(\vec{Z}) = \Delta((z_0, z_1, z_2, \ldots)) \equiv (z_1, z_2, z_3, \ldots). \tag{12}$$

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

$$F(\vec{\lambda}) = \Delta(\vec{\lambda}) + \vec{\tau}, \tag{13}$$

where $F(\vec{\lambda})$ denotes the $\vec{\lambda}$ components of the updating rule in (5). Let $F^{(k)}(\cdot)$ be the kth iterate of F, and define $F_t^k(\vec{\lambda})$ to be the $(t+1)$th element of $F^{(k)}(\vec{\lambda})$.[7] In addition, recall that b_t is the coefficient of μ_t in the optimal control rule (11). Then, the following result holds.

[6] From (2) we have

$$W(X^{n+1}, X^n, Y^n) = -\frac{1}{2}\Big[(1+\alpha)(X^{n+1})^2 + \alpha(X^n)^2 - 2X^{n+1}(\mu_0^n + \alpha X^n) + \frac{1}{\lambda_0^n} + (\mu_0^n)^2\Big].$$

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

[7] For example,

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

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

PROPOSITION 2: *The value function* $V(X,Y)$ *is given by*

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

where

$$(14) \quad T(\vec{\tau}) = -\frac{1}{2}b_0\left[\sum_{k=1}^{\infty}\beta^k\sum_{t=0}^{\infty}\left(\frac{b_t}{b_0}\right)^2\frac{1}{F_t^k(\vec{\lambda})}\right]$$

$$\propto -\Bigg[\left(\frac{1}{\lambda_1+\tau_1}+(a^2\beta^2)\frac{1}{\lambda_2+\tau_2}+(a^2\beta^2)^2\frac{1}{\lambda_3+\tau_3}+\cdots\right)$$

$$+\beta\Bigg(\frac{1}{(\lambda_2+\tau_2)+\tau_1}+(a^2\beta^2)\frac{1}{(\lambda_3+\tau_3)+\tau_2}$$

$$+(a^2\beta^2)^2\frac{1}{(\lambda_4+\tau_4)+\tau_3}+\cdots\Bigg)$$

$$+\beta^2\Bigg(\frac{1}{(\lambda_3+\tau_3+\tau_2)+\tau_1}+(a^2\beta^2)\frac{1}{(\lambda_4+\tau_4+\tau_3)+\tau_2}$$

$$+(a^2\beta^2)^2\frac{1}{(\lambda_5+\tau_5+\tau_4)+\tau_3}+\cdots\Bigg)+\mathcal{O}(\beta^3)\Bigg].$$

To interpret this result, notice that the term $\sum_{t=0}^{\infty}(b_t/b_0)^2\left(1/F_t^k(\vec{\lambda})\right)$ in (14) represents the contribution to the value function from the uncertainty the agent faces when she takes a decision k time steps in the future. The $1/F_t^k(\vec{\lambda})$ term is the agent's uncertainty about events that are t time steps in the future, in period k. The exponentially declining factor $(b_t/b_0)^2 = (a^2\beta^2)^t$ captures the importance of uncertainty about events at temporal distance t for decision-making as can be seen from the optimal policy rule (11). Thus, $T(\vec{\tau})$ is 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 t time steps in the future in a period k time steps from now is reduced by forecasts of precision $\tau_{k+t}, \tau_{k+t-1}, \ldots, \tau_t$.

II. The Relative Value of Short- and Long-Run Predictability

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

A. *Marginal Predictability*

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

Begin by defining the function

$$g(m) \equiv \sum_{l=0}^{\infty} \frac{\beta^l}{\lambda_{m+l}^2} \tag{15}$$

and assuming that

$$\lim_{l\to\infty} \frac{\lambda_{l+1}^2}{\lambda_l^2} > \beta, \tag{16}$$

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

PROPOSITION 3: *If the interactions between forecast lead times are neglected, the increase in the value function due to the prediction system is* (*to first order*)

$$dV = T(d\vec{\tau}) - T(\vec{0}) \approx \frac{a}{\alpha(1 - a^2\beta)} \sum_{m=1}^{\infty} r_m d\tau_m, \tag{17}$$

where

$$r_m \equiv g(m)\,\beta^m\Big(1 - \big(a^2\beta\big)^m\Big). \tag{18}$$

To understand the intuition behind this result, we now derive it heuristically. 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 (14) makes clear, the value of a reduction in uncertainty about events t time steps in the future is proportional to $(b_t/b_0)^2 = (a^2\beta^2)^t$. 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

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

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

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

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., $\beta(1/\lambda_{m+1}^2)$. 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

$$\frac{1}{\lambda_m^2}\left(\beta^m\sum_{t=0}^{m-1}(a^2\beta)^t\right) + \frac{\beta}{\lambda_{m+1}^2}\left(\beta^m\sum_{t=0}^{m-1}(a^2\beta)^t\right) + \frac{\beta^2}{\lambda_{m+2}^2}\left(\beta^m\sum_{t=0}^{m-1}(a^2\beta)^t\right) + \cdots$$

$$\propto \left(\sum_{l=0}^{\infty}\frac{\beta^l}{\lambda_{m+l}^2}\right)\beta^m\left(1-(a^2\beta)^m\right).$$

This is exactly the expression we obtained in (18). Notice how the derivation of this expression makes it clear that sequential updating of forecasts is not a major determinant of the value of a *marginal* unit of predictability. The fact that forecasts are updated sequentially gives rise to the factor $\left(\sum_{l=0}^{\infty}\beta^l/\lambda_{m+l}^2\right) = g(m)$ in (18), but if only a *single* marginal forecast of lead time m were received in the first period, the expression in (18) would look very similar, with this factor simply replaced by $1/\lambda_m^2$. Thus, neglecting the interactions between lead times is qualitatively similar to neglecting sequential forecast updating itself (we make this analogy exact in a special case below).

Equation (18) makes it clear that the dependence of prior uncertainty on lead time can have an important influence on the value of marginal predictability at different lead times through the function $g(m)$. To understand these effects in a parsimonious way, we will focus on a simple parametric model of prior beliefs. We suppose that the precisions of prior beliefs about the locations of θ_t are given by

$$(\lambda_t)^2 = \phi^t(\lambda_0)^2 + (1-\phi^t)(\lambda_\infty)^2, \tag{19}$$

where $\phi \in (0,1]$ and $0 < (\lambda_\infty)^2 < (\lambda_0)^2$. In this model, the squared precision of prior beliefs about events decays geometrically from $(\lambda_0)^2$ for the current period to $(\lambda_\infty)^2$ for events in the infinite future. It is straightforward to verify that (16) is always satisfied in this case as long as $\lambda_\infty > 0$. Moreover, notice that if beliefs about the infinitely distant future are arbitrarily uncertain, i.e., $(\lambda_\infty)^2 \to 0$, we have

$$\lim_{(\lambda_\infty)^2 \to 0^+} \frac{g(m)}{g(1)} = \lim_{(\lambda_\infty)^2 \to 0^+} \frac{\sum_{l=0}^{\infty} \frac{\beta^l}{(\lambda_{m+l})^2}}{\sum_{l=0}^{\infty} \frac{\beta^l}{(\lambda_{1+l})^2}} \tag{20}$$

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

Thus, in this (not implausible) limit, the ratio $g(m)/g(1)$ takes an especially simple form. The limiting ratio in (20) is well defined for all $\phi \in (0,1]$, even though $g(m)$ itself diverges if $(\lambda_\infty)^2 = 0$ and $\phi < \beta$.

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

$$R_m \equiv \frac{r_m}{r_1} = \underbrace{\beta^{m-1}}_{\text{Discounting}} \underbrace{\left(\frac{1}{\phi}\right)^{m-1}}_{\text{Uncertainty}} \underbrace{\left[\frac{1-(a^2\beta)^m}{1-a^2\beta}\right]}_{\text{Early warning}}. \tag{21}$$

Using this expression, the relative value of the predictability of events at different lead times may be computed as a function of the three parameters α, β, and ϕ. These parameters characterize the decision-maker's flexibility, impatience, and prior uncertainty about the future, respectively. Aside from being simple to analyze, the choice of priors in (19) makes the formulas for r_m/r_1 for updating from a single forecast versus updating from sequential forecasts coincide exactly in the limit as $(\lambda_\infty)^2 \to 0$ since $(1/(\lambda_m)^2)/(1/(\lambda_1)^2) = g(m)/g(1)$ in this case. To a first approximation, there is thus no difference between once-off and sequential forecasting in this model of priors. This is thus an especially good baseline from which to assess how accounting for the interactions between lead times alters the balance between short- and long-run predictability.

The formula (21) shows that there are three effects that determine the relative value of marginal predictability at different lead times. First, since forecasts at larger lead times relate to more distant payoffs, they are more heavily discounted. This gives rise to the first term in (21), which is *decreasing* in m. Second, since the prior precision of beliefs is smaller for larger lead times (i.e., uncertainty increases with the time horizon), and payoffs are concave in precisions, the effect of a marginal increase in the precision of beliefs is increasing in lead times. This leads to the second term in (21), which is *increasing* in m, reflecting the fact that the long run is more uncertain than the short run. Finally, the third term captures the cumulative effect of an early warning about events at lead time m on all subsequent decisions that are made until that event is realized. Since warnings of lead time m give rise to improved decision-making for $m - 1$ subsequent adjustment decisions, longer lead times are associated with greater cost savings. Thus, the third term in (21) is *increasing* in m, reflecting the fact that earlier warnings give rise to cheaper adjustments (since adjustment costs are convex). It is moreover straightforward to verify that

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

when $m \geq 1$. Thus, the larger are adjustment costs α, the faster the third term in (21) increases with m. This is intuitive since the more costly adjustments are, the more important it is to get early warning of the need for them (again due to convexity). This term thus demonstrates how our assumption of convex adjustment costs *favors* long-run predictions.[8]

The overall dependence of R_m on m depends on the relative rates of increase and decrease of the three terms in (21). Some simple analysis (see online Appendix F) shows that R_m can exhibit only three kinds of qualitative behavior. First, if $\phi \leq \beta$, R_m is an increasing function of m. In this case, the benefits of reducing large long-run uncertainties outweigh the effects of discounting, making long-run predictability more important than short-run predictability (when interactions between lead times are neglected). If $\phi > \beta$, R_m either decreases monotonically with m or is a unimodal function with a global maximum at some $m \geq 2$. Online Appendix F characterizes the regions of parameter space where these two qualitative behaviors occur. In general, when β is sufficiently small, R_m will be declining in m for all values of α. However, when β exceeds some critical value $\hat{\beta}$, there exists an $\hat{\alpha} > 0$ such that for all $\alpha > \hat{\alpha}$, R_m is unimodal. Analytic expressions for $\hat{\beta}$ and $\hat{\alpha}$ show that the faster prior uncertainty increases with lead time (i.e., the lower is ϕ), the lower are $\hat{\beta}$ and $\hat{\alpha}$. Figure 2 plots R_m for several values of the parameters.

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

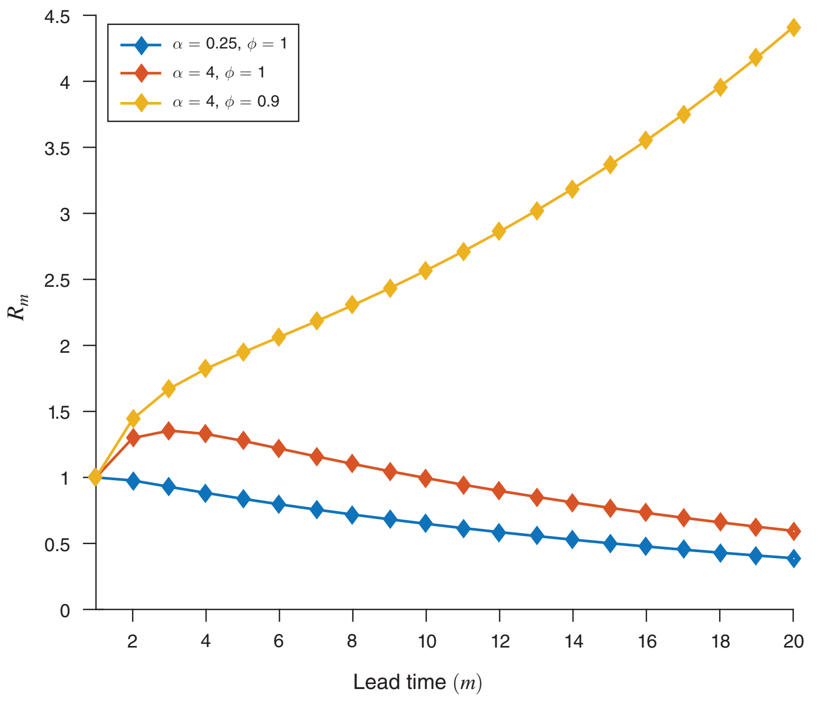

FIGURE 2

Notes: Typical dependence of R_m on adjustment costs (α) and prior uncertainty (ϕ). In these examples, $\beta = 0.95$.

Taken at face value, the first-order analysis in this subsection would seem to suggest that long-run predictability is often significantly more important than short-run predictability. When $\phi < \beta$ in (21), i.e., when the long run is significantly more uncertain than the short run, the analysis of R_m suggests that long-run predictability has a greater effect on discounted expected payoffs than short-run predictability regardless of the adjustment cost parameter α. Moreover, even when $\phi > \beta$, it is possible to find values of the parameters for which R_m increases for a long time before declining.[9] However, as we have emphasized, the first-order analysis largely neglects the dynamic nature of decision-making; it does not account for the interactions between lead times and thus cannot reflect the substitution possibilities that are a consequence of sequential fore-

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

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

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

cast updating. This analysis thus defines a naïve baseline that is conceptually similar to a conflation of the presence of long-run uncertainty with the need for long-run predictions.

B. *Accounting for Interactions*

In this section, we move beyond first-order results, aiming to summarize the dependence of $T(\vec{\tau})$ on $\vec{\tau}$ in a manner that accounts for the interactions between lead times. It is intuitively clear that these interactions are important determinants of the overall value of a prediction system. If we are able to predict events at lead time m very accurately, the value of an improvement in the predictability of events at lead time $m + 1$ must surely be quite low. Indeed, inspection of the expression for $T(\vec{\tau})$ in (14) 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, predictabilities at different lead times are substitutes. Since $T(\vec{\tau})$ is a nonseparable function of the infinite sequence of parameters $\vec{\tau} = (\tau_m)_{m \geq 1}$, there is no unique way of computing the contribution of each individual τ_m to the value function. We will focus on a measure of the importance of different lead times that we find especially intuitive.

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

Formally, we are interested in computing the following quantity:

$$\text{(22)} \qquad \vec{\sigma} \equiv \frac{1}{B}\left(\underset{\vec{\tau}}{\arg\max}\, T(\vec{\tau}) \text{ subject to } \sum_{m=1}^{\infty} \tau_m = B\right).$$

The mth component of $\vec{\sigma}$, denoted σ_m, is the share of the total predictability budget B that the agent would like to allocate to lead time m. By definition, $\sigma_m \in [0,1]$ for all m, and $\sum_{m=1}^{\infty} \sigma_m = 1$. We prove that the optimization problem in (22) has a unique solution in the online Appendix. Although it is not possible to solve for $\vec{\sigma}$ analytically, it is straightforward to find an arbitrarily good approximation to the solution using standard numerical optimization routines.

In general, $\vec{\sigma}$ depends on the vector of prior precisions $\vec{\lambda}$. To maintain consistency with the marginal analysis of the previous section, we assume that $\lambda_t = \phi^{t/2}\lambda_0$, corresponding to the $(\lambda_\infty)^2 \to 0$ limit of (19). The relative importance of prior beliefs and predictions in determining the expectations that enter the value function is captured by the ratio λ_0/B. To see this, notice from (14) that the optimization problem in (22) is equivalent to finding a sequence of values $\vec{\sigma}$ that maximizes

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

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

The results in Figure 3 depend on the choice of time step, which we can interpret as one year in this example. Decisions and beliefs are thus interpreted as updating annually (consider climate change adaptation, for example), and the agent's real discount rate is approximately 5 percent/year when $\beta = 0.95$.[10] Several features

[10] If we change the model's time step from Δt to $\Delta t' = k\Delta t$ for $k > 0$, the model parameters change to $\beta' = \beta^k$, $\phi' = \phi^k$, $B' = B/k$. The last of these transformations follows from the fact that $1/\tau_m$ measures the quantity of new information the agent receives about events at distance $m\Delta t$ *per unit time*. If we halve the time step to six months, for example, the agent should receive half as much information about future events, every six months. The change in the precision of beliefs about lead time $m\Delta t$ events from each forecast is $\tau_m/\Delta t$. Thus, if the

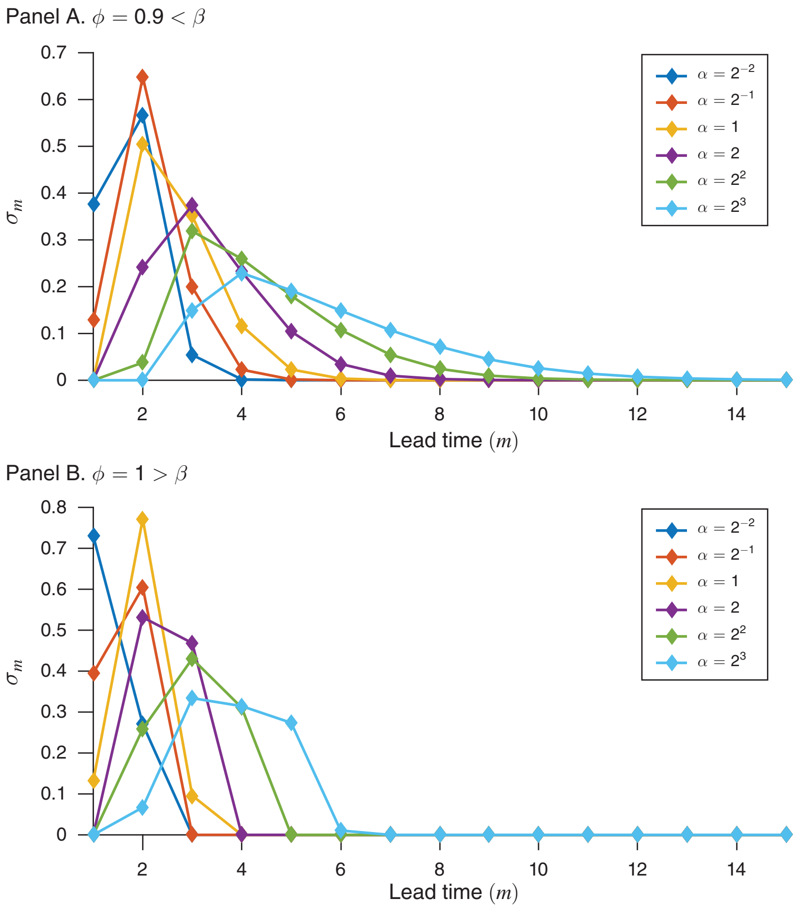

FIGURE 3

Notes: Budget share σ_m allocated to lead time m in the optimization problem in (22). $\beta = 0.95$, $\lambda_0/B = 1$.

of the figure deserve highlighting. First, the budget allocations in this figure tell a very different story from the marginal analysis in Figure 2. Even when $\phi < \beta$ (i.e., the top panel in Figure 3) so that the long run is significantly more uncertain than the short run, the decision-maker would like to allocate most of her predictability budget to short lead times, i.e., 1–4 time steps ahead. By contrast, the analysis in Figure 2 showed that the value of a marginal unit of predictability is *increasing* in lead time when $\phi < \beta$; at face value, this would seem to suggest that the decision-maker should allocate her entire budget to long-run prediction. The results in Figure 3 are very different because $\vec{\sigma}$ accounts for the substitution possibilities between lead times that arise from sequential forecast updating. These substitution effects favor the short run because accurate short-run forecasts can compensate for any errors in long-run forecasts once the long-run events in question come nearer to

total predictability budget in (22) is $B = \sum_m \tau_m/\Delta t$ when the time step is Δt, this is equivalent to a budget of B/k when the time step is $\Delta t'$.

the present, in addition to providing information about current near-term conditions. If adjustment is highly costly (i.e., α is large), it is clearly more costly to react to short-run warnings, and accurate early warnings become more important. This is reflected in the fact that $\vec{\sigma}$ places larger weight on long lead times as α increases. However, perhaps surprisingly, even for large values of α substitution effects cause the agent to allocate very small budget shares to lead times larger than ten time steps. The presence of predictability at these shorter lead times renders long-run predictability essentially irrelevant. Figure 3 also shows that prior uncertainty does not have nearly as large an effect on how the decision-maker would like to allocate her predictability budget as the marginal analysis suggested it might. The two panels of Figure 3, which correspond to values of ϕ below and above β respectively, differ in their details, with lower values of ϕ (i.e., greater long-run uncertainty) giving rise to greater weight on longer lead times. But in both cases, very little weight is given to long lead times. In contrast, the marginal analysis suggested that these cases should give rise to very different behavior, with long-run forecasts always being more valuable than short-run forecasts when $\phi < \beta$. This indicates that substitution effects dominate the role of the prior in determining the budget allocation, despite the precision of the prior being comparable to or greater than that of predictions.

While Figure 3 pertains to the baseline case $\lambda_0/B = 1$, it is also interesting to examine $\vec{\sigma}$ when $\lambda_0/B \to 0$. In this limit, the agent's beliefs about all future periods are *entirely* determined by the prediction system; the prior plays no role. In this case, $\vec{\sigma}$ captures the "pure" effects of substitution between lead times, unadulterated by the prior (which favors long lead times). Results for this case are depicted in Figure 4, when the agent's discount factor $\beta \to 1$, i.e., in the case that is *most* generous to long-run forecasts. This figure demonstrates how substitution between lead times can lead the decision-maker to place most of her budget on short-run predictability, even when she is very farsighted and adjustment costs are large. An equivalent figure for the lower value $\beta = 0.95$ is available in the online Appendix; short-run forecasts are of course substantially more strongly favored in this case.

III. Conclusions

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

Our results point to potentially important lessons for decision-makers and for efforts to improve the social value of forecasts. As observed in the introduction, it is not uncommon to find the presence of long-run uncertainty conflated with a need for

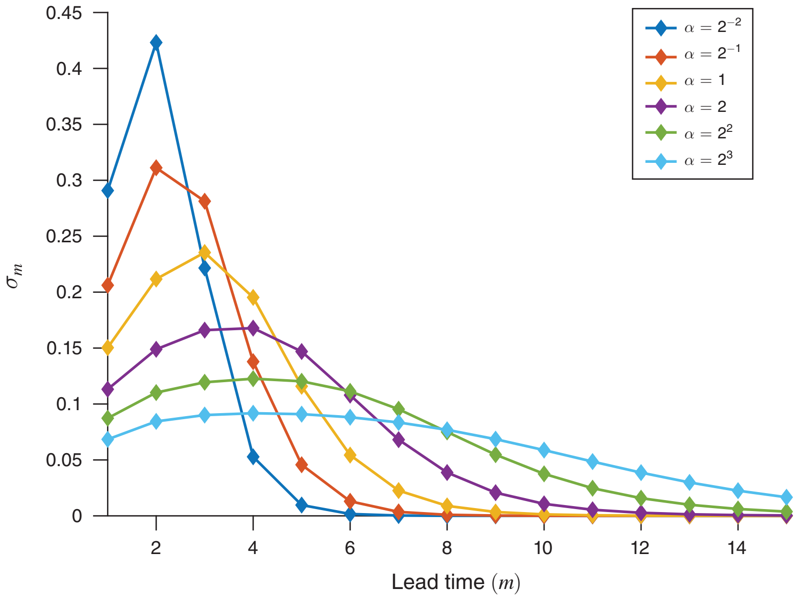

Figure 4

Notes: Budget share σ_m allocated to lead time m in the optimization problem in (22), when $\beta \to 1, \lambda_0/B \to 0$. This figure illustrates the "pure" effect of substitution between lead times when priors play no role in the analysis and in the case that is most generous toward long-run forecasts, i.e., $\beta \to 1$.

long-run predictability in policy circles. In general, however, this is a logical fallacy as it neglects decision-makers' abilities to adjust activities over time in response to updated forecasts. Our analysis suggests that if adjustments in response to sequential forecast updates are accounted for, short-run predictability is often more valuable than long-run predictability, even if adjustment costs and long-run uncertainty are large. It is perhaps surprising just how effectively short-run predictability can substitute for long-run predictability in our model as the convexity of adjustment costs would seem to imply that accurate long-run forecasts would give rise to significant cost savings when adjustment costs are large. The fact that this result would be difficult to guess a priori (at least for us) points to the necessity of modeling approaches that aim to disentangle the contribution of predictions at each lead time to the overall value of forecasts. Such models could also provide forecast producers with valuable information about where they should focus their efforts at forecast improvement. While improvements in long-run predictions often require new scientific approaches that reduce model misspecification errors, short-run predictions can often be substantially improved by simply reducing measurement errors in initial conditions (i.e., increasing the quality of observations). Our results suggest that the latter activity may carry significant value for decision-makers concerned with adapting to long-run changes, even though such improvements will yield little new information about long-run conditions.

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

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

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

Third, the model does not consider serially correlated environments. While we have opted for analytical and conceptual clarity over empirical comprehensiveness on this front, it is clearly of interest to study models that admit such correlations, but nevertheless allow predictability at different lead times to be disentangled, in future work.

REFERENCES

Arrow, Kenneth J. 1992. "Kenneth J. Arrow—'I Know a Hawk from a Handsaw.'" In *Eminent Economists—Their Life Philosophies*, edited by Michael Szenberg, 42–50. Cambridge, UK: Cambridge University Press.

Arrow, Kenneth J., and Anthony C. Fisher. 1974. "Environmental Preservation, Uncertainty, and Irreversibility." *Quarterly Journal of Economics* 88 (2): 312–19.

Benveniste, L.M., and J.A. Scheinkman. 1979. "On the Differentiability of the Value Function in Dynamic Models of Economics." *Econometrica* 47 (3): 727–32.

Blackwell, David. 1953. "Equivalent Comparisons of Experiments." *Annals of Mathematical Statistics* 24 (2): 265–72.

Blanchard, Olivier Jean, and Stanley Fischer. 1989. *Lectures on Macroeconomics*. Cambridge, MA: MIT Press.

Clements, Michael P. 1997. "Evaluating the Rationality of Fixed-Event Forecasts." *Journal of Forecasting* 16 (4): 225–39.

Costello, Christopher J., Richard M. Adams, and Stephen Polasky. 1998. "The Value of El Niño Forecasts in the Management of Salmon: A Stochastic Dynamic Assessment." *American Journal of Agricultural Economics* 80 (4): 765–77.

Costello, Christopher, Stephen Polasky, and Andrew Solow. 2001. "Renewable Resource Management with Environmental Prediction." *Canadian Journal of Economics* 34 (1): 196–211.

DeGroot, Morris H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.
Epstein, Larry G. 1980. "Decision Making and the Temporal Resolution of Uncertainty." *International Economic Review* 21 (2): 269–83.
Füssel, H.-M. 2007. "Adaptation Planning for Climate Change: Concepts, Assessment Approaches, and Key Lessons." *Sustainability Science* 2 (2): 265–275.
Gollier, Christian, Bruno Jullien, and Nicolas Treich. 2000. "Scientific Progress and Irreversibility: An Economic Interpretation of the 'Precautionary Principle.'" *Journal of Public Economics* 75 (2): 229–53.

Granger, Clive W.J., and Yongil Jeon. 2007. "Long-Term Forecasting and Evaluation." *International Journal of Forecasting* 23 (4): 539–51.
Grupe, Dan W., and Jack B. Nitschke. 2013. "Uncertainty and Anticipation in Anxiety: An Integrated Neurobiological and Psychological Perspective." *Nature Reviews Neuroscience* 14: 488–501.
Lee, Ronald. 2011. "The Outlook for Population Growth." *Science* 333 (6042): 569–73.
Lindh, Thomas. 2011. "Long-Horizon Growth Forecasting and Demography." In *Oxford Handbook of Economic Forecasting*, edited by Michael P. Clements and David F. Hendry, 585–606. New York: Oxford University Press.
Ljungqvist, Lars, and Thomas J. Sargent. 2004. *Recursive Macroeconomic Theory*. 2nd ed. Cambridge, MA: MIT Press.
Marschak, Jacob, and Koichi Miyasawa. 1968. "Economic Comparability of Information Systems." *International Economic Review* 9 (2): 137–74.
Merton, Robert C. 1971. "Optimum Consumption and Portfolio Rules in a Continuous-Time Model." *Journal of Economic Theory* 3 (4): 373–413.
Myers, Monica F., D.J. Rogers, J. Cox, Antoine Flahault, and S.I. Hay. 2000. "Forecasting Disease Risk for Increased Epidemic Preparedness in Public Health." *Advances in Parasitology* 47: 309–30.
National Academy of Sciences. 2016. *Next Generation Earth System Prediction—Strategies for Subseasonal to Seasonal Forecasts*. Washington, DC: National Academies Press.
Palmer, Tim, and Renate Hagedorn, eds. 2006. *Predictability of Weather and Climate*. Cambridge, UK: Cambridge University Press.
Regnier, Eva. 2017. "Probability Forecasts Made at Multiple Lead Times." *Management Science* 64 (5): 2407–26.
Sargent, Thomas J. 1978. "Estimation of Dynamic Labor Demand Schedules under Rational Expectations." *Journal of Political Economy* 86 (6): 1009–44.
Selten, Reinhard. 1998. "Axiomatic Characterization of the Quadratic Scoring Rule." *Experimental Economics* 1: 43–61.
Yusuf, Moeed. 2009. "Prediction Proliferation: The History of the Future of Nuclear Weapons." Brookings Institution Foreign Policy Paper 11.