Emergency Preparation and Uncertainty Persistence*

Savitar Sundaresan†

November 30, 2022

Abstract

Unusual events trigger persistent spikes in uncertainty. Standard models cannot match these dynamic patterns. This paper presents a unified framework, motivated by the literature on inattention. Agents *choose whether and how to prepare* for different possible states of the world by collecting information. Agents optimally ignore sufficiently unlikely events, so the occurrence of such events does not resolve, but rather increases, uncertainty. Uncertain agents have dispersed beliefs, making it harder to focus future preparation. Thus, uncertainty begets uncertainty for an inattentive agent, endogenously persisting. In a financial application, this framework matches patterns in volatility, volume of trade, belief dispersion, and spreads.

JEL: D82, D83, G10, G14, G20

*Thanks to Ricardo Reis, Michael Woodford, Larry Glosten, and Jose Scheinkman for invaluable guidance, assistance, and advice; to Adrian Buss, Nick Bloom, Patrick Bolton, Pedro Bordalo, Andrew Caplin, Chris Hansman, Victoria Ivashina, Rajkamal Iyer, Marcin Kacperczyk, Paymon Khorrami, Filip Matejka, Kristoffer Nimark, Elise Payzan-Lenestour, Joel Peress, Ansgar Walther, Stijn van Nieuwerburgh and Laura Veldkamp for discussions and comments that helped shape the content of this work; to seminar and conference participants at SLOAN-NOMIS, SITE, Columbia, Emory, Imperial, UBC, UC Davis, Northwestern Kellogg, London Business School, Johns Hopkins Carey, the Federal Reserve Board, SFS Cavalcade and Alex Chinco (discussant), CEPR and David Schreindorfer (discussant); and to an anonymous associate editor and two referees. All remaining errors are my own.

†Business School, Imperial College London; Email: s.sundaresan@imperial.ac.uk, Phone: +44 (0) 20 7594 6449

Do not fear everything, nor take account of all alike; If you wanted to take everything equally into account on every occasion that happens, you would never do anything.

\- Herodotus, *The Histories 7.50.1*

1 Introduction

Rare or unexpected events repeatedly cause measures of political and financial uncertainty to spike up *and stay persistently high* for weeks or even months. For example, the VIX increased after the terrorist attacks of 9/11 and the collapse of Lehman Brothers, policy uncertainty jumped after Brexit and Trump's election in 2016, and persistence of both variables is high, as shown in Figure 1.[1] In the midst of the global COVID-19 Pandemic, uncertainty has dictated how economic agents make consumption, investment, pricing, and portfolio allocation decisions, so understanding the source and nature of uncertainty persistence is crucial in both macroeconomics and finance.

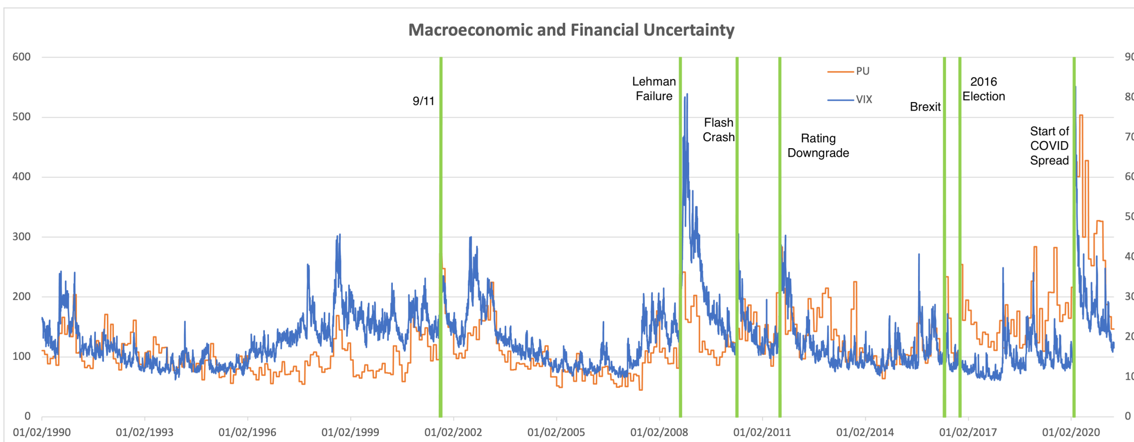


Figure 1: Monthly values of the Policy Uncertainty Index (PU) and daily values of the CBOE's Volatility Index (VIX). VIX measures expected volatility in the S&P 500 as calculated from options prices. PU measures macroeconomic uncertainty as calculated from news articles, policy expectations, and analyst disagreement.

Standard models cannot generate these patterns in uncertainty dynamics. The workhorse model of uncertainty dynamics is *parameter learning*,[2] where agents do not know the value of a parameter (say, the variance) of a particular process, but can learn about it through observation. However, because learning agents learn very quickly, parameter-learning rational-expectations models need a lot of assumptions on the data (e.g., a stochastic volatility process) or on the agents (e.g., a short memory or behavioral biases), to generate empirical patterns such as uncertainty spiking

1. As a rough measure, the VIX and PU exhibit autocorrelations of about 98% and 81%, respectively.
2. See Orlik and Veldkamp (2014), Kozlowski, Veldkamp, and Venkateswaran (2020), Nimark (2014) and Collin-Dufresne, Johannes, and Lochstoer (2016) for persistence in a macrofinancial setting.

repeatedly, and only jumping up instead of down. [3]

In the rational inattention literature, information acquisition is usually conducted contemporaneously instead of preparatively.[4] To fix ideas, suppose that an agent were interested in preparing for the next US Presidential election. In standard rational inattention models, that would involve collecting information to get an accurate signal about who would win that election. Uncertainty over the outcome could be resolved with sufficient attention - making persistence in uncertainty quite difficult to generate. This paper departs from the traditional inattention paradigm by instead using the concept of preparative attention: agents take their beliefs over who might win the election as given, and instead use their attention to try to increase the precision of future election outcome-contingent probability distributions.[5] In this paper, agents select *and pay for* future state-contingent certainty, today.

This paper's objective, contributing to a recent literature on dynamic inattention and learning,[6] is to develop a theory that can parsimoniously explain patterns of repeated, persistent spikes in uncertainty. It requires minimal assumptions on the data-generating process and a cost of attention as its only friction. In applying the model to a financial market, this paper is able to match patterns in bid-ask spreads, expected volatility, dispersion of beliefs, and volume of trade. The model provides a novel framework through which to microfound the autoregressive patterns seen in financial variables - usually modeled by exogenously specified ARCH and GARCH processes.

The Mechanism and Results

In this model, there is an underlying stationary process. Agents can choose to exercise attention to collect information about future states of the world one period in advance (taking their current beliefs over those future states as given). Given that information collection is costly, they cannot attend to all future states. Which states will they choose to learn about? Agents devote their resources towards states of the world that are, risk-weighted, more likely to occur, and ignore states of the world that are, risk-weighted, less likely to occur.[7] If the underlying process is

3. Parameter learning models and their different predictions are also discussed in more detail in the appendix.

4. See, for example, Van Nieuwerburgh and Veldkamp (2009)

5. In reality, agents perform both functions: learning about which outcome might occur, *and* preparing for what follows different outcomes. Because the first function has already been carefully analyzed, this paper abstracts from it, and focuses on the second. A more thorough discussion of the distinctions between the two frameworks is in the appendix, while a model that combines the two functions while preserving this paper's results is discussed in section 3.

6. See, for example: Banerjee and Breon-Drish (2017), Banerjee and Kremer (2010)

7. Throughout this paper, I assume distributions are risk-weighted. Unlikely events that lead to high levels of marginal utility could attract more attention, but the point of this paper's model is that *at some point* the unlikeliness

itself nonstationary, persistence is easier to obtain, so stationarity is assumed to showcase the mechanism's power.

Using preparation as the key mechanism to explain uncertainty dynamics, this paper's model delivers two results. The first result is that *agents prepare more for events they deem likely and less for events they deem unlikely.*[8] This result is also well established in the experimental literature.[9] When an anticipated event occurs, preparation pays off and the precision of agents' beliefs increases; whereas when an unanticipated event occurs, the precision of agents' beliefs decreases.

The second result is to show that *if agents are uncertain today, they will continue to be uncertain tomorrow, because they cannot prepare well enough for any event.* One unexpected or rare event will trigger a high level of uncertainty because agents will not have prepared for it. But agents' incentives to prepare change when they are uncertain. Uncertain agents assign low probabilities to a wide variety of states, and adequate preparation would require examining each of these wide variety of states. Given that preparation is costly, uncertain agents will acquire little information about each future contingency. Conversely, an expected or likely event results in a low level of uncertainty because agents will have prepared for it. Low uncertainty makes preparation easy because beliefs are tight, so agents can focus their preparations on the few, likely states in their beliefs. Therefore, uncertainty (and certainty) persists endogenously. This result differs from standard rational inattention models, which place constraints or costs on mutual information. In those models, costs are proportional to the probabilities of events, so ex-post uncertainty is identical in every state (no matter how likely that state was to occur). There could, in fact, be no uncertainty dynamics in such a setting.

I also show that the model's mechanism is robust to some natural extensions. First, suppose agents are also allowed to collect contemporaneous information which improves the precision of their beliefs today over states of the world tomorrow. In this setting, agents can 'catch up'. If an ex-ante low-probability state occurs, and their preparatory information collection was poor, they can make up for it ex-post by collecting contemporaneous information. I show that agents will top up their preparatory information with contemporaneous information, but only up to a point. Agents will therefore have an upper bound on how uncertain they can become, but the qualitative dynamics of the model continue to hold. Second, while the baseline model is presented in a two-state setting, I provide sufficient conditions for the mechanism to be robust to an N-state setting.

of an event outweighs the marginal utility. For example, we do not typically worry about meteors hitting the earth.

8. This result was shown first by Maćkowiak and Wiederholt (2018) in a static model.

9. See Shaw and Shaw (1977), Posner, Snyder, and Davidson (1980), Eriksen and James (1986)

Finally, I embed the model in a financial market to show how it can match empirical patterns. I allow a continuum of atomistic, sophisticated agents to collect state-contingent information about the stationary and exogenously specified return-process of an asset. Agents' can trade on their information through a competitive market making sector that sets prices. Market makers do not collect information of their own, but merely set bids and asks and facilitate trade. To ensure non-degenerate equilibria, there are also a measure of atomistic noise traders, who are price-insensitive. Sophisticated agents will prepare for expected returns, and will not prepare for unlikely returns. Therefore, by the previously described mechanism, unexpected returns trigger deteriorations in the quality of private information. Such deterioration leads to higher levels of uncertainty among sophisticated agents, reduced volume of trade, increased dispersion of beliefs, and higher levels of expected volatility. The last result provides a microfoundation for the shape of the VIX, as well as an intuition for ARCH and GARCH patterns in volatility and expected-volatility.

This mechanism of preparation is motivated by the experimental literature. Cognitive experiments have shown that allocating attention in a preparatory manner is a part of our attentional makeup. In a seminal experiment by Shaw and Shaw (1977), agents were presented with visual stimuli. The subjects were told how likely the stimuli were to occur in different parts of their visual field. When stimuli were then shown in parts they were told were likely, subjects were able to identify them accurately. When stimuli were shown in parts they were told were unlikely, subjects could not identify them accurately. The authors concluded that agents are able to arrange their peripheral attention to prepare for different events and that they prepare more for likely events than for unlikely events. To deliver this finding theoretically, I make one key assumption about the preparation process: *the marginal cost for preparation is the same for every state of the world*, no matter how likely to occur.[10] Such an assumption is not really necessary, and the results of the paper are robust to many intuitive alternatives. The cost structure that would weaken or even negate the findings of the paper is one where it is *harder* to prepare for common events than it is to prepare for rare events.[11]

This mechanism is further supported by the fact that *economic and financial agents do, in fact, prepare for various events.* For example, central banks have instructed financial institutions to undergo stress-testing in anticipation of another financial crisis. Martin and Pindyck (2015) consider the mitigating properties of such preparation for rare events. Stress-testing is, at least in

10. This assumption is a reduced form simplification of the channel capacity concept introduced to economics by Woodford (2012), which was motivated in part by the experimental evidence of Shaw and Shaw (1977).

11. It is, as shown by Woodford (2012), a relatively undesirable property of a constraint on mutual information.

part, useful to mitigate the uncertainty that would stem from a financial crisis, by having a plan in place to deal with some of its effects.

Additionally, this model supports the intuition that *agents invest more in preparation when they are uncertain*. Uncertain agents are sometimes modeled as wanting information more, or scrambling for information. As I show in one of the final results of the paper, despite the fact that high uncertainty leads to worse preparation *on average*, it also leads to more spending on preparation *in total* due to the wealth of possible states that are likely enough to warrant at least some preparation.

Understanding uncertainty dynamics is crucial for financial and economic analysis.[12] Because risk-aversion is a fundamental assumption in economics, reducing uncertainty provides first-order welfare benefits. Therefore, finding the mechanism by which uncertainty persists and propagates facilitates policies that reduce uncertainty and temper its effects. This paper addresses this inquiry by showing that tools from the cognitive and economics literature on attention yield a mechanism that causes uncertainty to spike and remain persistently high after rare events.

Literature Review

Works such as Sims (2003) and Woodford (2012) argue that even if information is plentiful, attention is a scarce resource. This notion has been explored in numerous fruitful settings, where a cost is typically placed on the reduction in entropy agents obtain from receiving noisy information. See for example Matějka (2015a), Maćkowiak and Wiederholt (2009), Matejka and McKay (2015).

Bolton and Faure-Grimaud (2009) uses a dynamic information model to show that agents might not fully work through contingencies up front, postponing deliberations on low-probability or low-risk events. The focus of that paper lies less with dynamic spillovers and persistence of beliefs and more with the tradeoff between reacting quickly to an event and understanding future implications.

There is also a large literature that uses informational frictions to analyze the macroeconomic impacts of changes in beliefs. Such papers typically use a parameter-learning setting. Orlik and Veldkamp (2014) and Kozlowski, Veldkamp, and Venkateswaran (2020) study the impact of changes in tail beliefs. Nimark (2014) looks at the impact of extraordinary news events and generates persistence in volatility by showing that agents are more sensitive to information after such events.

This paper is part of an agenda in the inattention literature that is interested in exploring the *dynamics* of inattention. Specifically, this paper is placed in a sequence of papers that analyze

12. See the literature spawned by Bloom (2009)

how attending to the world *today* impacts one's ability and relative willingness to attend to the world later. Steiner, Stewart, and Matějka (2017) study sluggish responses to a slow-moving state. Maćkowiak, Matějka, and Wiederholt (2018) propose analytics for dynamic attention problems. Nimark and Sundaresan (2019) show that ex-ante identical agents can diverge to opposite ends of a belief spectrum due to rationally confirmatory and complacent behavior. Ilut and Valchev (2017) propose a dynamic framework where agents can pay attention to an entire policy *function*.

Papers such as Banerjee and Breon-Drish (2017), Banerjee and Kremer (2010), and Cujean and Praz (2015) consider dynamic implications of asymmetric information, information acquisition and disagreement on financial variables such as volume of trade, delayed entry, and correlation structures. Financial market interactions with information flows has been studied by Andrei and Cujean (2017) and Andrei and Hasler (2014) to explain patterns in volatility, momentum, reversal, and return predictability.

Because this paper emphasizes the importance of attention across many possible states, there is a natural parallel to the concept of salience. Salience in economics and finance was introduced by a sequence of papers: Bordalo, Gennaioli, and Shleifer (2012), Bordalo, Gennaioli, and Shleifer (2013a), Bordalo, Gennaioli, and Shleifer (2013b). They show that agents will choose to focus their attention on a subset of possible outcomes based on what they deem salient. This paper does not consider salience explicitly, but the notion that different states carry differing levels of importance, and therefore will command differing levels of attention, is crucial for this paper's results.

Sections 2 presents the model in a two-state setting. Section 3 extends the model to allow for contemporaneous information acquisition, and arbitrarily many (though still finite) states. Section 4 applies the model to a financial market, deriving implications for uncertainty, volatility, dispersion of beliefs, spreads, and trade. Section 5 concludes.

2 Simple Model

This section presents the simplest possible model in which the key mechanism of the paper can be described. This model has one agent, two possible states of nature, and an infinite number of periods. Using costly, state contingent preparation via information acquisition, the model delivers persistent spikes in uncertainty in response to unlikely events.

2.1 Model Structure

State Structure: There are two possible states s of the world each period, A and B. To fix intuition, suppose that one state of the world, A, is 'the market increases in value' while the other, B, is 'the market decreases in value'. The probability of the state taking the value A each period, unconditional on any information, is $\pi = P(s_{t+1} = A)$. For simplicity, I assume that $\pi = 1-\pi = \frac{1}{2}$. Relaxing this assumption would produce quantitative but not qualitative changes to the results.

Agent: There is a single agent. The agent enters period t with an information set I_t. Her beliefs in period t about period $t+1$'s state are given by $p_t = \max(P(s_{t+1} = A|I_t), P(s_{t+1} = B|I_t)) \geq \frac{1}{2}$. Note that p_t is *not* the probability of the market going up in value, but is a value indicating the likelihood of the (weakly) more likely state occurring. The likely state could either be the market increasing or decreasing. But one state has to be at least as likely as the other, so p_t refers to the probability of the more likely event. p_t can therefore be viewed as measuring the *precision* of the agent's beliefs. One can interpret these probabilities as being risk-weighted, or risk-neutral, to control for the possibility that the agent might prefer one state to another. I will index the more likely state for period $t+1$ with L in period t. I will index the less likely (or rare) state for period $t+1$ with R in period t. Put differently, $p_t = P(s_{t+1} = L) = 1 - P(s_{t+1} = R)$. Thus, hereafter we abandon the labelling of A and B, and use only L and R.

Preparation: Conditional on her beliefs p_t, the agent can choose in period t to *prepare* for each of the two possible states that could occur in period $t+1$. That is, the agent can *prepare* for the market either increasing or decreasing in value in the next period, by collecting information about what will happen conditional on each of those two possibilities. Preparation takes the form of picking information sets $I_{t+1,L}$ and $I_{t+1,R}$. When state L or state R occurs in period $t+1$, the agents beliefs will be:

$$
\begin{aligned}
p_L &= \max(P(s_{t+2} = A|s_{t+1} = L, I_{t+1,L}), P(s_{t+2} = B|s_{t+1} = L, I_{t+1,L})) \\
p_R &= \max(P(s_{t+2} = A|s_{t+1} = R, I_{t+1,R}), P(s_{t+2} = B|s_{t+1} = R, I_{t+1,R}))
\end{aligned}
$$

When one of the two states occurs, the agent's preparation for that state comes to fruition. Information forms the precision of her beliefs, conditional on the state. For simplicity, I model this decision as the agent directly selecting two possible values for p_{t+1} in period t: $p_L \geq \frac{1}{2}$ is the belief distribution she will have if the *likely* state occurs in $t+1$, and $p_R \geq \frac{1}{2}$ is the belief distribution she will have if the *rare* state occurs in $t+1$. In a binary distribution, the probability parameter p comoves with the precision of the distribution $(p(1-p))^{-1}$, so long as $p \geq \frac{1}{2}$, as has been assumed. Therefore, for the simplicity of the proofs, I will have the agent directly select p_L and p_R, which is isomorphic to the agent picking conditional precisions (a more standard assumption) $(p_L(1-p_L))^{-1}$ and $(p_R(1-p_R))^{-1}$ or conditional information sets $I_{t+1,L}$ and $I_{t+1,R}$.

The intuition behind this structure is as follows: an investor might have beliefs about whether the market will go up or down tomorrow. And she might have already placed trades to take advantage of those beliefs. But she can also prepare for what comes next. If the market goes up tomorrow, will it go up again the next day? If the market goes down tomorrow, will it rebound afterwards? Collecting information about those possibilities will allow the agent to react to each quickly and profitably. Therefore, she would want to spend resources to be better prepared. In the next section, I allow agents to collect information in period t about the events of period $t+1$ and show that all of the results continue to hold.

Agent's Costs: Improving the quality of preparation is costly, and the agent can choose how much to prepare, or whether to prepare at all. There is no fixed cost of preparing, but there is a cost function $c(p_{t+1})$ associated with increased precision of conditional beliefs. That cost function is the *same* for the likely state and the rare state.

Agent's Objective: The agent has a period-by-period utility function, which has one input: the precision of the agent's beliefs. As discussed above, under a binary distribution, as in this setup, the value p_t comoves with the precision of the distribution $(p_t(1-p_t))^{-1}$, so long as $p_t \geq \frac{1}{2}$, as has been assumed. Therefore, for simplicity, I assume that the agent's *utility function* is given by $U(p)$. The agent benefits from having precise beliefs. Such a benefit is a common feature in the utility function of any risk-averse agent. Note that the utility function assumes the agent is indifferent as to whether one state or another occurs, and cares only about the accuracy of her predictions going forward. The state itself is not an input into the utility function. Intuitionally, the agent does not care if the market rises or falls, and can trade beneficially long or short as long as her beliefs are precise. Relaxing this assumption would have quantitative but not qualitative effects on the results. One way to interpret the paper to allow for state-dependent payoffs is to think of all the probabilities describe here as *risk-neutral* or *risk-weighted* probabilities.

Graphical Interpretation: Figure 2 shows the setup of the model. An agent in period t takes her beliefs about the states in period $t+1$, p_t, as given. Then, using costly information acquisition, the agent chooses the conditional precision of her beliefs for period $t+1$ about period $t+2$, p_L and p_R. Now that the structure of the economy has been described, I can present the agent's problem.

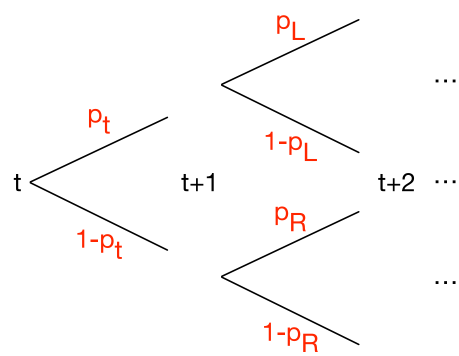


Figure 2: A sample of three periods in the binary version of the model. Agents take their beliefs in time t, $p_t, 1-p_t$, over the two states in period $t+1$ as given, and choose p_L and p_R, subject to a cost. One of those choices will form their beliefs in period $t+1$, depending on which state occurs, about the likelihood of the two states in $t+2$.

2.2 Statement of Problem

Formally, the agent's problem in period t is given by the following Bellman equation:

$$V(p_t) = \max_{p_L, p_R} U(p_t) - c(p_L) - c(p_R) + \beta p_t V(p_L) + \beta(1-p_t)V(p_R) + \mu_R \left(p_R - \frac{1}{2}\right) + \mu_L \left(p_L - \frac{1}{2}\right) \quad (1)$$

The agent begins period t with a certain set of beliefs p_t. The value associated with a particular set of beliefs can be broken up into several components. First, there is the *utility function*, U, a real-valued function on $\left[\frac{1}{2}, 1\right]$, that is increasing in the precision of the agent's beliefs. As discussed above, because $p_t \geq \frac{1}{2}$ comoves negatively with the variance of the distribution, having U be an increasing function of p_t mimics a common feature of almost any standard utility function. Then there is the *cost function*, c. The agent pays a cost to improve the precision of her future beliefs. That cost is additively separable across states, and is the same regardless of the likelihood of the state in question. These assumptions are strong for tractability, but not essential, as will be discussed further below. The agent discounts the future at a rate $\beta < 1$. With probability p_t, the *likely state* will occur, and the precision of the agent's conditional beliefs will be $(p_L(1-p_L))^{-1}$. With probability $1-p_t$, the *rare state* will occur, and the precision of agent's conditional beliefs will be $(p_R(1-p_R))^{-1}$. The agent's choices are subject to inequality constraints, without loss of generality, that her beliefs be weakly stronger than $\frac{1}{2}$. The Kuhn-Tucker conditions associated with those constraints are represented by the last two terms.[13] Intuitively, an agent values her beliefs

13. A proof that the Bellman equation 1 is a contraction is in the appendix. The fact that the Bellman is a contraction mapping implies that there is a unique solution to the optimal choice that can be found by iteration.

based on two things: (i) how uncertain those beliefs make her, today, and (ii) the relative difficulty those beliefs cause her in preparing for the next period's events.

Deviation from the standard: This Bellman formulation above differs in one crucial way from a standard dynamic optimization problem. The state variable this period has *two* purposes. The first, standard purpose is that it enters the utility function U. The second purpose, which is the innovation of this paper (as well as of Nimark and Sundaresan (2019), albeit under different circumstances) is that the state variable affects the *distribution of the continuation value.* Put differently, this latter channel shows that my information set today impacts how I want to collect information for the future. Therefore, my motives in collecting information today are twofold: first to maximize my utility next period, and second to change *how I want to collect information in the next period.* This second channel emphasizes the dynamic spillover of attention, and is the sole intertemporal aspect of the model.

2.3 Solution

Assumptions: To guarantee a solution, I need to impose some structure on the functions mentioned above. I assume that:

1. $U'(\cdot) > 0$. This assumption states that the agent prefers stronger beliefs to weaker ones.
2. $c'(\cdot) > 0$. This assumption states that stronger beliefs are more costly than weaker ones.
3. $c''(\cdot) > U''(\cdot)$. This assumption states that the marginal cost of stronger beliefs increases faster than the marginal benefit - a technical assumption to guarantee an interior solution.
4. $c'(1) > U'(1)$. This assumption states that complete certainty is too costly to ever be attained.

These assumptions are all standard, in that they are satisfied by many existing utility and costly information forms, and are sufficient to guarantee that interior solutions $p_L(p_t)$ and $p_R(p_t)$ exist.

2.4 Results

The first result is that the agent will choose to collect *more* information about the *likely* state, and *less* information about the *rare* state.

Proposition 1. $p'_R(p_t) \leq 0$ *and* $p'_L(p_t) \geq 0$*.*

All proofs are in the appendix. To see the result more starkly, notice that Proposition 1 implies that

$$p_L(p_t) \geq p_L\left(\frac{1}{2}\right) = p_R\left(\frac{1}{2}\right) \geq p_R(p_t)$$

The value of preparing for an event is proportional to its likelihood, but the cost is independent. Therefore, an agent will focus her preparation on the state of the world she believes likely to occur, as it has a good chance of proving useful. The converse of this result is that the agent will tend to prepare less for states of the world she believes less likely to occur, as such information would have a small chance of proving useful. A version of this result for a static decision is shown in Maćkowiak and Wiederholt (2018). This result provides the first glance at how the occurrence of rare events can cause uncertainty to spike - due to lowered levels of attention ex-ante.

The second result is that if the agent's beliefs are sufficiently precise, she will not collect *any* information about the *rare* state.

Lemma 1. *If* $\frac{c'\left(\frac{1}{2}\right)}{U'\left(\frac{1}{2}\right)\beta} > 1 - p_t$*, then* $p_R = \frac{1}{2}$.

If the marginal benefit of collecting information about the rare state is lower than the marginal cost of collecting information, *when the agent hasn't collected any information yet*, the agent won't bother collecting any information about the rare state. Because $U'(\cdot)$ and $c'(\cdot)$ are positive everywhere, there is always a set of beliefs about the likely state that are precise enough such that the rare state will be ignored. This is obviously true in reality - for example, most people don't prepare for a meteor hitting their home.

Agents completely ignore *sufficiently* rare events. A dynamic implication of this result comes from noticing that the term *sufficiently* is relative to functional forms: The third result shows that under certain conditions, the agent will not collect any information about either state when her beliefs are the unconditional distribution.

Corollary 1. *(High uncertainty can be permanent). If* $\frac{c'\left(\frac{1}{2}\right)}{U'\left(\frac{1}{2}\right)\beta} \geq \frac{1}{2}$ *then* $\exists t$ *such that* $p_s = \pi_s$ $\forall s > t$.

This follows immediately from Lemma 1. If the agent doesn't collect information, her beliefs will be the unconditional probability distribution. But if the unconditional probabilities are *sufficiently* unlikely according to the above definition, that problem becomes permanent. If they don't collect information under the unconditional probability distribution, they will then again face the unconditional probability distribution in the subsequent period. Therefore, if the agent, for any reason, chooses not to acquire information for a particular state, and then that state occurs, the agent will opt *never to collect information again.* Such a condition could be satisfied if the marginal benefit of information is quite low at the unconditional distribution, or if the marginal cost is quite high.

But the corollary is stronger than that: it also implies that if the conditions are met, an uncertainty trap is *inevitable.* If the agent doesn't collect information for the unconditional probability distribution, the agent won't collect information for any state less likely than the unconditional 50-50 distribution, by Proposition 1. But $1 - p_t$, the probability of the rare state is weakly less than $\frac{1}{2}$. Therefore, if the above conditions hold, the agent will not collect information for the rare state, *ever.* The rare state must occur *at some point*, and when it does, the agent will *permanently* cease information collection.

The final and most important result of this section shows that there is a 'steady state' level of uncertainty to which the agent will converge.

Proposition 2. *(Uncertainty is persistent) If* $\frac{c'(\frac{1}{2})}{U'(\frac{1}{2})\beta} < \frac{1}{2}$*, then there is a 'convergent' fixed point* $\underline{p}$ *where* $\underline{p} = p_L = p_t$*.*

There is a level of uncertainty at which the agent's beliefs will remain *conditional on the likely state continuing to occur.* If the rare state occurs, uncertainty will spike up, and will only decline as the agent collects information again.

The best way to see the implications of these results is graphically. Consider Figure 3. On the x-axis of this figure is the agent's beliefs today. The right-hand side of the axis is p_t, while the left-hand side of the axis is $1 - p_t$. The kinked, curved line is the *policy function* - the agent's information collection as a function of her beliefs today. The left-hand side of the curve (the part that corresponds to the left-hand side of the x-axis) is p_R, and the right-hand side of the curve (the part that corresponds to the right-hand side of the x-axis) is p_L. The straight, solid line, is a *45-degree line* - the set of points where beliefs in the next period have the same distribution as beliefs today. There is an intersection between the policy function and the 45-degree line at the green dot. Call the point of intersection $\tilde{p}$. If $p_t = \tilde{p}$, then $p_L(\tilde{p}) = \tilde{p}$, and as long as the likely state occurs, the agent's beliefs will always be $\tilde{p}$. With beliefs of $\tilde{p}$, the agent's choices of information collection will be at the green dot and the red dot. The reason there is a kink in the curved line is because of Lemma 1 to the left of the kink, the rare state is sufficiently unlikely, and the agent will not prepare for it - hence $p_R = \frac{1}{2}$. To the right of the kink, the rare state is sufficiently likely, and the agent will prepare for it - hence $p_R > \frac{1}{2}$.

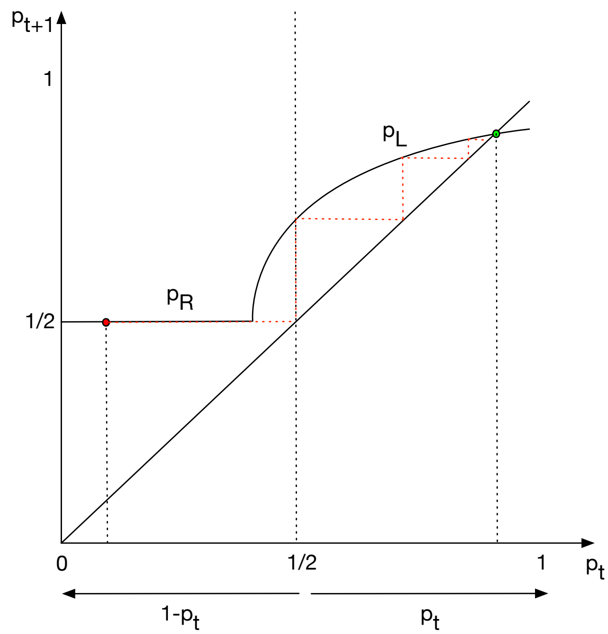


Figure 3: Plotted is the policy function of the precision of beliefs in period $t+1$ as a function of the probability of a particular event occurring in period t. The kinked-curved line is the policy function, while the straight line is a 45-degree line. The red dashed line shows a possible evolution of beliefs between the red-rare event level of precision and the green-steady state level of precision.

If the rare state occurs, as is indicated by the red dot, the agent will not have collected any information, and will therefore be very uncertain. The policy function, as illuminated by the red-dashed line, shows that when the agent is uncertain, she will collect information about both states (as they are equiprobable). The red dashed line, shows the evolution of the agent's belief precisions in subsequent periods as the likely state continues to occur. If the rare state occurred at some point before the agent's beliefs had converged back to $\tilde{p}$, the process would be set back, but would then start to converge again. One can think of this exercise as being a two-state version of an impulse response function. Given one 'shock' or rare event, followed by no other deviations, what happens to beliefs? An illustration of the evolution of the agent's beliefs can be seen in Figure 4.

This graph shows what happens to the agent's uncertainty if the likely state happens for several periods, the rare state happens once, and then the likely state happens thereafter. Uncertainty remains low while the likely state occurs. There is a spike when the rare state occurs, because the agent did not collect any information about it beforehand. Due to the increase in uncertainty, the agent will start to collect information again, and as the likely state continues to occur, the agent's uncertainty will start to decline. Notice the similarity of the dynamics of this exercise with the plots in Figure 1 - there is persistence in the expected variance of future price movements!

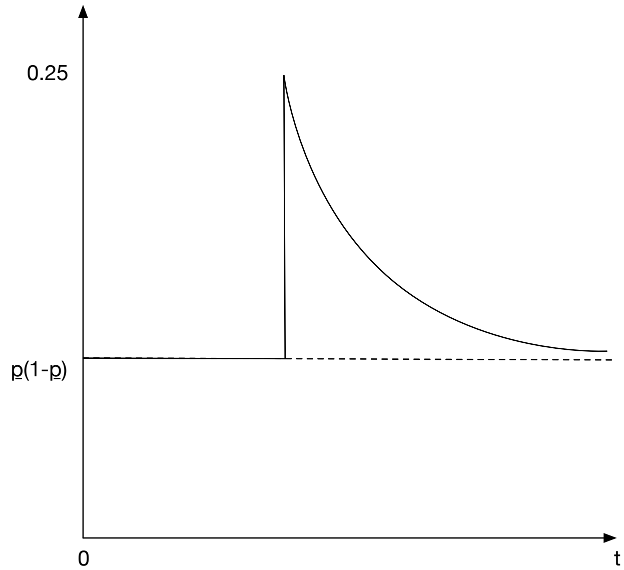


Figure 4: Plotted is a standard impulse response exercise of uncertainty as a function of time. After several periods of steady-state certainty, a rare event occurs. Upon such an occurrence, uncertainty spikes up, and beliefs will trace out the path of the red-dashed line in the previous figure, assuming that only likely states occur subsequently.

Uniqueness The solution does not guarantee *uniqueness*, yet uniqueness is not required for the results. The agent's belief precision can only move in two ways. Uncertainty can suddenly jump up if the rare state occurs. Uncertainty can decline slowly, according to the policy function, as the likely state occurs. As the likely state occurs, belief precisions will converge to the first intersection between the policy function and the 45-degree line seen in Figure 3. If there are other, additional intersections for higher values of p_t, they can never be reached, as belief precisions converge to the first intersection, and cannot move above it. If an agent's beliefs start at a steady state of higher precision, they will eventually become uncertain with probability 1, as the unconditional process is i.i.d, which means that they will converge to the first intersection.

The key link between periods - in fact, the *only* link between periods, is that collecting information in one period *affects how information can be collected* in the next. It is the sole mechanism that delivers persistence in information collection, and therefore, persistence in uncertainty.

2.5 Discussion of Key Assumptions

There are two key assumptions in this model. The first is the structure of information collection and revelation. The second is the cost structure of information. Both are important, but the model's results can withstand loosenings of both.

Information Structure: The big assumption in the structure of information collection is that the agent takes the precision of her beliefs in period t about the likelihood of states in period $t+1$ as *given*. This precludes the possibility that the agent can collect information about $t+1$ in period t - she can only collect information about period $t+2$ in period t, even though that information will only be revealed in $t+1$. As I show in the next section, this assumption is for tractability purposes - the results will hold even if it is relaxed (as long as she can still collect information about period $t+2$ in period t as in the baseline).

From a *theoretical* standpoint, this paper is interested in *contingent* information acquisition, so this is the simplest setting to *isolate* that channel. Relaxing this restriction, and allowing the agent to collect information in time t about the state in $t+1$ as well as contingent information about $t+2$, would weaken the propositions quantitatively, but not qualitatively. Ultimately, contingent information about $t+2$ can only be collected once beliefs about $t+1$ are established. If the agent can collect information about $t+1$ at time t, she will merely start from a higher baseline for her contingent information collection.

From an *intuitional* standpoint, it is not unreasonable to think that agents will try to form contingent plans (be they investment plans, spending plans, etc), taking their beliefs about the future as given. In planning for the results of elections, referendums, etc, investors make plans for any outcome. They may try to reduce their uncertainty about the event, but they will allocate their resources in conditional information collection based on their beliefs about the outcome itself.

Cost Structure: The main assumption in the cost of information collection is that it is equally costly to collect information about either future state of the world. From a *theoretical* standpoint, this assumption improves tractability. This assumption is also justified by the more detailed framework of Woodford (2012). There is, as was mentioned earlier, also substantial support in the cognitive experimental literature for the static results that obtain from such an assumption. From an *intuitional* standpoint, one might think that it is actualy *costlier* to collect information about low-probability states, and *cheaper* to collect information about high-probability states. Such an alteration of the cost structure, would actually strengthen the results by making it even less likely that the agent would collect information about rare states. The cost structure that would weaken, or even negate the results of the paper is one where it is *costlier* to collect information about high-probability states.

3 Extensions

In this section, I consider two extensions to the baseline model to show that the mechanism is robust to alternative formulations. The first extension allows agents to collect contemporaneous information; the second allows for more than 2 states of the world per period.

3.1 Extension to Contemporaneous Information Acquisition

Throughout the paper thus far I have assumed that agents can only prepare for future states, but cannot reduce their uncertainty today over states of the world in the next period. This assumption is obviously a strong one, which I made to highlight the preparation mechanism that is central to the paper's results. In this section, I relax that assumption, and consider the possibility that, in addition to preparatory information collection, an agent in period t can change their beliefs about the states of the world in period $t+1$ by collecting information (contemporaneous information collection).

3.1.1 Setup With Contemporaneous Acquisition

The agent now has the ability to collect contemporaneous information *and* prepare for future events, so the agent's problem has two sets of choice variables. The agent enters period t with belief p_t about the states of the world in $t+1$, but can now further adjust their precision of beliefs about the state of the world in $t+1$ by improving it directly. The agent picks the improvement $s_t \geq 0$ on the precision of future beliefs. The new and improved beliefs, $(p_t + s_t)$, would come at a cost $c_s(p_t + s_t) - c_s(p_t)$. Prior to any contemporaneous information acquisition, the agents' precision of beliefs will be p_t. I assume that improving the precision of beliefs is costlier the more precise beliefs already are, which is why the cost does not take the form $c_s(s_t)$. Another way to interpret this cost function is that the cost to the agent is not over the amount of information collected, but on the marginal improvement in beliefs. I further assume that $c_s(p_t + s_t) \to \infty$ as $s_t \to 1 - p_t$ so that full contemporaneous learning is never optimal and that $c_s'' > U''$ everywhere. Conditional on a level of contemporaneous information acquisition, s_t, the agent's problem is as before - the agent selects a set of beliefs $\{p_R, p_L\}$ now subject to the results of the contemporaneous information collection.

The agent's problem is therefore:

$$
\begin{aligned}
V(p_t) \quad = \quad & \max_{s_t, p_L, p_R} U(p_t + s_t) - c(p_L) - c(p_R) - (c_s(p_t + s_t) - c_s(p_t)) \\
& + \beta(p_t + s_t)V(p_L) + \beta(1 - p_t - s_t)V(p_R) + \mu_R \left(p_R - \frac{1}{2} \right) + \mu_L \left(p_L - \frac{1}{2} \right) + \lambda s_t
\end{aligned}
$$

The utility of the agent now depends on her improved beliefs after contemporaneous information collection; the cost of preparing for future states remains the same, but there is an additional cost for contemporaneous information collection; the continuation value also depends on the improved precision, while there is an additional non-negativity constraint on information collection in period t.[14]

3.1.2 Results

The first result echoes the static results of the previous section.

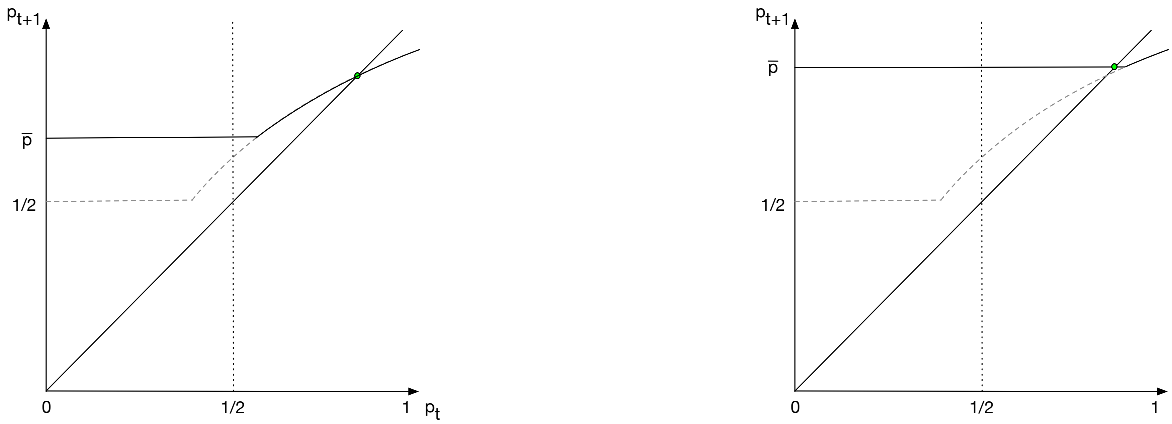
Proposition 3. *Conditional on a level of s_t, $p'_R(p_t) \leq 0$ and $p'_L(p_t) \geq 0$.*

Conditional on a level of contemporaneous learning, s_t, the static results of the previous section continue to hold: it will still be the case that the agent prepares more for the likely state, and less for the unlikely state. In order for the dynamic results of the model to hold, it must be the case that $p_t + s_t^*(p_t)$ is a weakly increasing function. That is, it must be that if an agent enters a period with more less precise beliefs, then the agent can't "make up for it after the fact" and obtain more precise beliefs than an agent who started the period with better information. The second result shows how contemporaneous information acquisition does not erase the mechanism of the previous section.

Proposition 4. *There exists a $\bar{p}$ such that for all $p_t < \bar{p}$, $s_t = \bar{p} - p_t$ and for all $p_t \geq \bar{p}$, $s_t = 0$.*

Essentially what this proposition says is that contemporaneous information collection is used to 'top-up' information when it's below a certain threshold, $\bar{p}$. However, if the agent is sufficiently certain (beliefs are stronger than $\bar{p}$), there will be no contemporaneous acquisition. Therefore, the two parts taken together imply that $p_t + s_t^*(p_t)$ is weakly increasing. Below $\bar{p}$ it is constant, and above $\bar{p}$ it is equal to p_t.

14. I have assumed that the agent can improve the quality of beliefs today. Implicitly, this assumes that the agent is improving the signal acquisition technology *prior* to receiving the signal that reduces her uncertainty. As this step of signal acquisition is elided in the baseline model, I continue to do so here.



(a) High cost of contemporaneous information acquisition.

(b) Low cost of contemporaneous information acquisition.

Figure 5: In this figure, the effects of different costs of contemporaneous information are shown. The panel on the left shows the model with a high cost of contemporaneous information acquisition, which leads to low levels of information top up, and high levels of uncertainty. The panel on the right shows the model with a low cost of contemporaneous information acquisition, leading to high levels of information top up, and low uncertainty.

Discussion

The intuition for this setting is best communicated visually. Recall Figure 3. The effect of contemporaneous information acquisition on that figure can be seen in Figure 5. Effectively, contemporaneous information acquisition raises the baseline of certainty for the agent. Certainty is always topped up to $\bar{p}$. Therefore the old equilibirum can still hold, if $\bar{p}$ is below the fixed point (panel A) or a new equilibrium will hold where $p_t = \bar{p}$ for all t (panel B).

The main impact of contemporaneous information acquisition is to weaken the impact of preparation. This takeaway is relatively unsurprising - if agents can collect and use information ex-post, they would typically prefer that to having to collect it ex-ante, when there is no guarantee that the information will be useful.

As long as contemporaneous information is not too cheap, the main results of the baseline model are preserved. Although agents "top up" their information acquisition contemporaneously, they won't do it to certainty, which means that above a threshold, it is still the case that agents will prepare more for likely states than for rare states.

3.2 Extension to Multiple States

Next we turn to an extension of the baseline model where there are multiple (more than two) possible states of the world each period. Suppose that in each period t there are n many possible states of the world that could occur in period $t+1$ with an unconditional distribution $p_1, ..., p_n$

(where $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$ for all i). In each period, the agent has to make n^2 many choices: for each of the n possible states that can realize in period $t+1$, (say state i) they must select n many probabilities which will determine the distribution over possible states in $t+2$ $(i1, ..., in)$. For the sake of simplicity, assume that the unconditional distribution is that $p_1 = ... = p_n$ - so the agent is maximally uncertain if no information is collected. The agent's problem then looks as follows:

$$\begin{aligned} V(p_1, ..., p_n) \quad = \quad & \max_{p_{11}, \ldots, p_{nn}} U(-H) - \sum_{j=1}^{n} \sum_{i=1}^{n} c(p_{ji}) + \beta \sum_{i=1}^{n} p_i V(p_{i1}, ..., p_{in}) \\ & + \lambda \left(\sum_{i=1}^{n} p_i - 1 \right) + \sum_{i=1}^{n} \mu_i \left(\sum_{j=1}^{n} p_{ij} - 1 \right) + \sum_{i=1}^{n} \sum_{j=1}^{n} \theta_{ij} p_{ij} \end{aligned}$$

Each period, the agent takes as given their probability distribution over states of the world in the next period. The agent gets some utility from the precision of that distribution, where precision is measured inversely by the distribution's entropy, $H \equiv -\sum_{i=1}^{n} p_i \ln(p_i)$. The agent can then, for each possible future state i, select the conditional probabilities for that future state $p_{i1}, ..., p_{in}$ subject to a cost function $c(p_{ij})$ which is the same for all i and j. These selections will also determine the state-conditional entropy, H_i. The choices are restricted (apart from their cost) by the fact that they must sum to one for each possible future state, and be weakly positive.

3.3 Results

Define $\bar{H} \equiv H\left(\frac{1}{n}, ..., \frac{1}{n}\right)$ to be the entropy of the unconditional distribution. We can then show the following:

Proposition 5. *For any concave utility function, there is a cost function c such that if the probability of a particular state increases, the conditional entropy in that state of the world weakly decreases.*

The results of the previous section continue to hold in this setting: namely, if the probability of a particular state i realizing in period $t+1$ increases, the entropy of the posterior distribution for the agent in state i decreases.

3.4 Discussion

Introducing multiple states makes the intuition slightly harder to grasp, but the fundamental mechanisms of the baseline model are still at play. The use of entropy[15] is convenient, as it provides a measure of precision that does not depend on state specific outcomes. Therefore, we can isolate the agent's concern about the precision of their beliefs without resorting to using second-moment based measures. If an agent can select the conditional precision (as measured by entropy) of her state-contingent beliefs, subject to a cost, she will choose to have a high degree of conditional precision (low entropy) for likely states of the world, and a low degree of conditional precision (high entropy) for unlikely states of the world.

4 Application

In this section we embed the the model of uncertainty persistence from section 2 into a financial market. In doing so, we can see the effects of persistence in uncertainty on financial and economic variables such as bid-ask spreads, volatility, dispersion of beliefs, and volume of trade. This application shows the ability of the mechanism to deliver patterns observed in data. However, this application is by no means the only one to which the mechanism is suited, and is meant to be illustrative.

4.1 Model Structure

State Structure: The correlate of the agent from previous sections is an *informed trader* in this section. She takes her beliefs in time t over the possible states of the world in $t+1$ as given. At time t, she can prepare by collecting information about period $t+2$ for each possible future state of the world in $t+1$.

Financial Market: The key to this application is that the states of the world are defined as possible fundamental values of an asset. The unconditional risk-weighted distribution of fundamental values at time $t+1$ is given by $F_{t+1} \sim f(F_t+1, F_t-1)$, where f is an equally weighted Bernoulli distribution. Therefore $P(F_{t+1} = F_t + 1) = P(F_{t+1} = F_t - 1) = \frac{1}{2}$. The unconditional volatility of fundamental values is constant across time, and therefore the unconditional process of the fundamental value is a random walk.

15. Or, really, any of the Hill numbers (Simpson's Index, or HHI would work well also).

Agents: Unlike the previous sections, there are multiple types of agents in this application. There are a continuum of perfectly competitive *market makers*, who observe public information, and set bids and asks. There are two types of trading agents who form a unit continuum. One fraction, comprising a measure $1 - T$, is an *informed trader*. This trader is the parallel of the agent of previous sections [16]. They are able to collect state-contingent information, and can trade with the benefit of that information. The remaining fraction, comprising a measure T, are *noise traders*. These traders are present to provide liquidity in the model and to preclude perfect price discovery - they participate without heeding public or private information and are insensitive to the price.[17] The last type of agent is a public entity who collects and broadcasts a public signal to all agents.

Trading: Each type of trader can buy one unit of the asset, sell one unit of the asset, or abstain from trade each period. By assumption half the noise traders will *always* buy one unit of the asset, and half the noise traders will *always* sell one unit of the asset. The informed traders base their trading decision on their private and public information and on prices which reflect public information. If their beliefs lie above the market makers' asks, they will buy one unit of the asset. If their beliefs lie below the market makers' bids, they will sell one unit of the asset. If their beliefs lie within the market makers' bid-ask spread, they will not trade.

Signals: There are two types of signals, public and private. The public signal, y_t is distributed $y_t \sim f_{pub,t}(F_{t+1})$ where $P_{pub,t}(y_t = F_{t+1}) = \phi_{pub,t} \geq \frac{1}{2}$. It is revealed to all agents (market makers and informed traders). Updated beliefs over F_{t+1} after seeing y_t can be characterized as: $f_{pub,post}(F_{t+1})$ where $\max P_{pub,post}(F_{t+1}) = \phi_{pub,t}$.

There is a private signal x_t, which is distributed $x_t \sim f_{priv,t}(F_{t+1})$ where $P_{priv,t}(x_t = F_{t+1}) = \phi_{priv,t} \geq \frac{1}{2}$. It is revealed only to informed traders. Updated beliefs over F_{t+1} after seeing y_t and x_t can be characterized as: $f_{pub,priv,post}(F_{t+1})$ where $\max P_{pub,priv,post}(F_{t+1}) = \phi_{pub,priv,t}$.

4.2 Statements of Traders' problems

There are two sets of decisions in the model. There are *pricing and trading decisions* that need to be made after all information has been revealed to the relevant agents. There are also *information decisions* that need to be made beforehand. I will proceed backwards through these decisions.

16. Here we treat the continuum of informed traders as a representative trader, who trades in a block. Alternatively, one could think of this as enforcing a symmetric equilibrium. The two alternatives are, ex-ante, identical.

17. See Chinco and Fos (2021) for an elegant endogenizing of noisy demand.

4.3 Pricing and Trading Decisions

There are two stage 1 decisions - a pricing decision by market makers, and a trading decision by traders.

Informed Trader Trading Decision: The informed trader buys one unit of the asset if their beliefs lie above the market makers' asks, and sells one unit if their beliefs lie above the market makers bids. Therefore, their trading problem is:

$$U(x_t, y_t, \phi_{pub,t}, \phi_{priv,t}) = \max\{(\text{bid}_t - s_t), (s_t - \text{ask}_t), 0\} \tag{2}$$

where $s_t = E[F_{t+1}|\text{public signal} = y_t, \text{private signal} = x_t]$.

Market Makers' Pricing Decisions: The market makers observe public information and set prices - a bid and an ask. They are modeled after the market makers in Glosten and Milgrom (1985). The bid is the price at which market makers buy the asset. The ask is the price at which market makers sell the asset. In order to understand their decision, we must understand the probability of receiving a buy or a sell order. The market makers' perceived probability of receiving a buy order conditional on the public signal being $y_t = F_t + 1$ is $\phi_{pub,t}\phi_{priv,t} + (1 - \phi_{pub,t})(1 - \phi_{priv,t})$ while the probability of a buy order conditional on the public signal being $y_t = F_t - 1$ is $(1 - \phi_{pub,t})\phi_{priv,t} + \phi_{pub,t}(1 - \phi_{priv,t})$. The probabilities of receiving sell orders are analogous. The market making sector is perfectly competitive, so each market maker sets prices to satisfy a zero-profit condition. These zero-profit conditions are written as follows for when $y_t = F_t + 1$ (analogous expressions capture the conditions when $y_t = F_t - 1$):

$$\begin{aligned} \Pi_{\text{ask}}|y_t = F_t + 1 &= (1 - T)(\phi_{pub,t}\phi_{priv,t} + (1 - \phi_{pub,t})(1 - \phi_{priv,t}))\,(ask - E[F_{t+1}|y_t = x_t]) \\ &\quad + \frac{T}{2}(ask - E[F_{t+1}|y_t = F_t + 1]) = 0 \qquad (3) \\ \Pi_{\text{bid}}|y_t = F_t + 1 &= (1 - T)((1 - \phi_{pub,t})\phi_{priv,t} + \phi_{pub,t}(1 - \phi_{priv,t}))(E[F_{t+1}|y_t \neq x_t] - bid) \\ &\quad + \frac{T}{2}(E[F_{t+1}|y_t = F_t + 1] - bid) = 0 \qquad (4) \end{aligned}$$

Because market makers set prices according to public information, prior to receiving any orders, there is no informational content in the price that is not already publicly available. Therefore, agents do not need to condition their orders on the price. This is a simplified setting of Glosten and Milgrom (1985) without sequential trade.[18]

18. If the model had a finite number of informed traders who traded sequentially, the results would still hold, albeit

4.4 Information Decisions

There are two stage 0 decisions - state-contingent information decisions made by informed traders and the public entity. The informed traders' decision in stage 0 of period t, given the quality of public signals, is to select a quality of a private signal on each possible realization of F_{t+1}. Formally, given the quality and realization of public and private information, the informed trader's decision is:

$$\begin{aligned}
& V(F_t, x_t, y_t, \phi_{pub,t}, \phi_{priv,t}) \\
= & \max_{\phi_{priv,t+1,F+1}, \phi_{priv,t+1,F-1}} U(x_t, y_t, \phi_{pub,t}, \phi_{priv,t}) + \beta E \\
& [P_{pub,priv,post,t}(F_{t+1} = F_t + 1) V(F_t + 1, x_{t+1,F+1}, y_{t+1,F+1}, \phi_{pub,t+1,F+1}, \phi_{priv,t+1,F+1}) \\
& + P_{pub,priv,post,t}(F_{t+1} = F_t - 1) V(F_t - 1, x_{t+1,F-1}, y_{t+1,F-1}, \phi_{pub,t+1,F-1}, \phi_{priv,t+1,F-1})] \\
& - \nu c(\phi_{priv,t+1,F-1}) - \nu c(\phi_{priv,t+1,F+1}) + \lambda_{+1} \left(\phi_{priv,t+1,F+1} - \frac{1}{2} \right) + \lambda_{-1} \left(\phi_{priv,t+1,F-1} - \frac{1}{2} \right)
\end{aligned} \quad (5)$$

where c is a convex increasing function, β is a discount factor, the subscript $F+1$ indexes the state where $F_{t+1} = F_t + 1$ and the subscript $F - 1$ indexes the state where $F_{t+1} = F_t - 1$. Essentially the value of a certain informational environment to the informed traders in time t consists of the profits they can earn by trading in the asset market today, coupled with the continuation value: the ability to increase the conditional precision of their beliefs for each of the two possible states that could occur in period $t + 1$.

The public entity chooses the quality of state-contingent public information. The public entity tries to maximize the traders' state-contingent posterior precision subject to a cost. The public entity works to select conditional signal quality, $\phi_{pub,t+1,i}(F_{t+1,i})$ for each potential value of F_t to maximize posterior beliefs:

$$\begin{aligned}
\max_{\phi_{pub,t+1,F+1}, \phi_{pub,t+1,F-1}} & E[\phi_{pub,priv,post,t+1}] - \nu_{pub} c_{pub}(\phi_{pub,t+1,F+1}) - \nu_{pub} c_{pub}(\phi_{pub,t+1,F-1}) \\
& + \mu_{+1} \left(\phi_{pub,t+1,F+1} - \frac{1}{2} \right) + \mu_{-1} \left(\phi_{pub,t+1,F-1} - \frac{1}{2} \right)
\end{aligned} \quad (6)$$

where c_{pub} is a convex function. The public entity's objective is to maximize conditional accuracy of traders' beliefs by selecting public signal precision for each of the two possible states in the next period. One could think of the public information as being reports or actions from public

somewhat less starkly.

institutions like the Federal Reserve or the government, or research reports published by financial institutions. There is good reason to believe that information producers would prepare for certain events in advance, so as to provide information as quickly and accurately as possible. For example, central banks have used stress testing to prepare for financial downturns. Such entities will prepare more for states that they consider to be (risk-weighted) more likely - recall that our distribution of fundamentals is risk-weighted normal.

4.4.1 Dynamic Equilibrium

Given values of $\{F_t, \phi_{pub,t}, \phi_{priv,t}, x_t, y_t\}$, a dynamic equilibrium is defined by choices of $\phi_{pub,t+1}(F_{t+1})$ by the public entity that solves equation 6 choices of $\phi_{priv,t+1}(F_{t+1})$ by the traders that satisfies equation 5, individual decisions to buy, sell, or abstain by traders to solve equation 2, and a choice of a bid and an ask by Market Makers to solve equations 3 and 4.

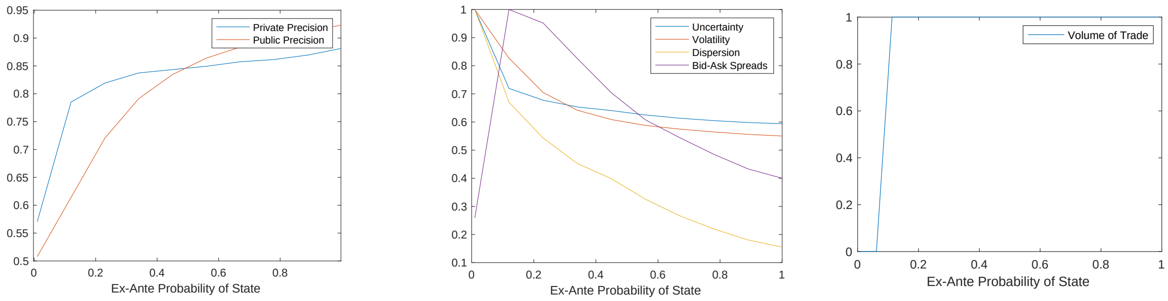
4.5 Predictions and Spillovers

I present the simulation of the model statically and dynamically.[19] First, statically, Figure 6 shows the values of several financially relevant variables as functions of the ex-ante probability of events. The x-axis on all three panels is the agent's ex-ante belief over the probability of a state. Panel (a) shows the precision of public and private signals, which are both, unsurprisingly, increasing in the ex-ante probability of the event: the more likely the event, the higher the quality of public and private information. This initial finding spills over to other variables. Panel (b) shows that Volatility,[20] Uncertainty,[21] and Dispersion are all lower after relatively likely events, than they are after rare events. Bid-Ask spreads follow a similar pattern by and large, but for very unlikely events, private information crashes so low that adverse selection in the market is quite small, leading to lower spreads. Finally, panel (c) shows that agents trade more aggressively after high probability events, because they are better prepared for what happens next, and can trade strongly on their relatively accurate beliefs. The volume results are more stark because of the market microstructure. Informed traders are always weakly better informed than market makers, because informed traders have public and private information, while market makers have only public information. Therefore, if informed traders have *any* private information, their beliefs will lie outside the bid-ask spread,

19. The parameter values here are $\nu_{pub} = 0.9$, $T = 0.9$, $\beta = 0.95$, $\nu = 0.0003$, $c(x) = \frac{1}{(x(1-x))^2}$ and $c_{pub}(x)\frac{1}{(x(1-x))^{0.1}} - 1$. This is not a calibration, but is meant only to be illustrative.

20. The expected variation of the fundamental conditional only on public information

21. The expected variation of the fundamental conditional on public *and* private information



(a) Informational Investment *(b) Financial Variables* *(c) Asset-Demand*

Figure 6: Plotted is a static snapshot of the applied model. The x-axis of all three plots is the ex-ante probability of an event. Figure (a) captures the precision of private and public signals. Figure (b) captures Uncertainty, Volatility, Bid-Ask Spreads and Dispersion. Figure (c) captures the volume of trade. The first two variables increase, and the next three variables decline the more 'expected' an event was. The last increases.

and they will all trade. If informed trders have *no* private information, they will not trade, as there is no informational advantage with which to trade.

These results all correlate to the static version of the model. Unlikely events have different properties than likely events *within a period*. The next presentation has to do with the dynamic implications of those properties. For this I present an impulse-response exercise. I suppose that an F_t takes an extremely unlikely value in period 2. Then I assume that the more likely value of F_t realizes until period 8.

The purpose of this exercise is to show how one perceived rare event in an unconditionally random walk process can trigger lasting effects in uncertainty and volatility. The results are shown in Figure 7. Unsurprisingly, given the previous graph all the variables spike up in period 2, which is when the unexpected change in the fundamental occurs. Between periods 2 and 8 any subsequent dynamics in these graphs is due to the natural processes of the mechanism of the model. Agents find themselves unable to prepare well for even the more likely events. However, with each passing period, they find that the event they prepared for more occurs, thus placing them in a slightly better position to prepare subsequently. This pattern continues as the agents converge to their steady state levels of uncertainty and risk. There are two results to note: first, total volume of trade, makes a momentary drop when traders are uninformed in period 2, but then fully recovers, as the informational advantage leads all informed traders to trade; second, bid ask spreads crash in period 2, as informed traders' information drops faster than public information, reducing adverse selection and spreads, but then starting in period 3, informed traders' information recovers faster as well, causing spreads to spike persistently until the steady state is achieved again.

A point of particular interest comes from the last panel. One hypothesis about human behavior

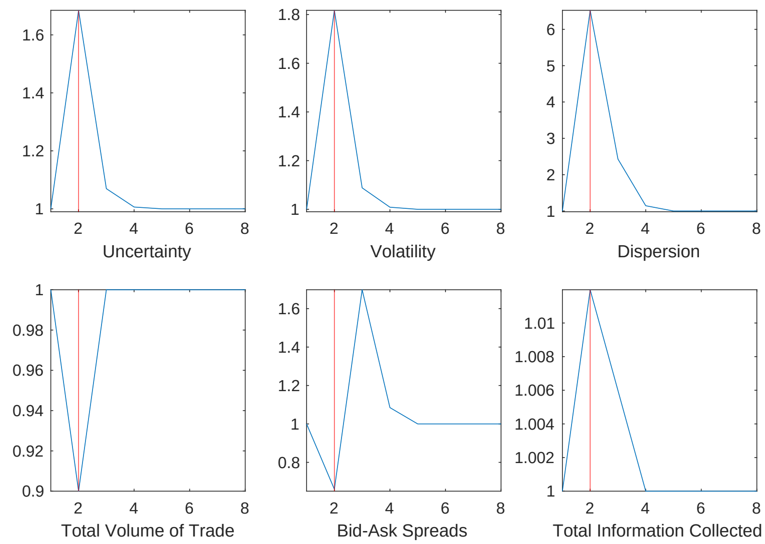


Figure 7: Plotted are five variables responding to an unexpected move in the fundamental of the asset and subsequent recovery. In period 2 *the fundamental makes the less likely movement. In periods 3 through 7, the fundamental makes the more likely movement.*

is that when we are uncertain, we try harder to learn. That intuition is borne out by this paper's mechanism. The last panel plots the total amount spent on information collection period by period. The point of emphasis thus far has been that when agents are uncertain, they don't know which states of the world to prepare for, so *on average* they are less prepared than when they are certain. However, when they are uncertain, they actually spend more *in sum* on information than when they are certain. This finding comes because agents can't ignore anything when they are uncertain - anything *could* happen! Therefore, they must prepare more *in total* than they would if they were certain, even though they are still less prepared *on average*.

One of the benefits of this type of application is that it contains a measure of uncertainty and expected volatility that follow an autoregressive pattern, which is borne out empirically by the VIX, while the underlying process itself follows a random walk. This paper's mechanism is able to wed those two seemingly conflicting concepts relatively tightly.

5 Conclusion

I have presented and solved a simple, tractable model that uses an attentional constraint to deliver observed patterns in uncertainty and volatility dynamics. Specifically, I employ the concept of preparation: I assume that agents take the probabilities of different states of the world occurring tomorrow as given, and instead use their resources and attention to prepare for what happens afterwards, should different states of the world occur. When unlikely or rare states occur, agents will not have not prepared for them beforehand and will face high levels of uncertainty. High uncertainty makes it harder for agents to prepare subsequently, as they do not know where to focus their preparations. Therefore, uncertainty begets itself, persisting endogenously, even when the underlying states of the world exhibit no persistence. To fix ideas, I first show this mechanism in a stylized two-state case, and show it is robust to some natural extensions. Finally, I embeded it in a financial model, and match observed patterns in volatility, dispersion of beliefs, bid-ask spreads and volume of trade. This mechanism provides a fundamentally new perspective from which to analyze uncertainty dynamics and its implications.

The mechanism in this paper is quite general and can be applied to other settings as well. For example, it can be used to think about the persistence and countercyclicality of price dispersion or wage dispersion. If firms prepare more for good times than for bad, then according to this paper's mechanism, firms will receive noisy signals on how to set prices and wages in bad times. Therefore, price dispersion and wage dispersion will be higher in bad times. By the mechanism of the paper, the dispersion will also be persistent. Persistence in performance could also be explained with this mechanism. Firms or individuals who had prepared for a given state of the world will perform well and be more certain when that state occurs. Thus, they would be better positioned to prepare for events in the next period, and on average their profitability should be persistent as well. Therefore, they would be more likely to face low levels of uncertainty subsequently, which would improve their performance as well. Firms that perform poorly would be unsure of how to proceed, and would continue to perform poorly. I leave questions like these for future research.

References

Andrei, Daniel, and Julien Cujean. 2017. "Information percolation, momentum and reversal." *Journal of Financial Economics* 123 (3): 617–645.

Andrei, Daniel, and Michael Hasler. 2014. "Investor Attention and Stock Market Volatility." *The Review of Financial Studies* 28 (1): 33–72.

———. 2020. "Dynamic attention behavior under return predictability." *Management Science* 66 (7): 2906–2928.

Banerjee, Snehal, and Bradyn M Breon-Drish. 2017. "Dynamic information acquisition and strategic trading."

Banerjee, Snehal, and Ilan Kremer. 2010. "Disagreement and learning: Dynamic patterns of trade." *The Journal of Finance* 65 (4): 1269–1302.

Bloom, Nicholas. 2009. "The impact of uncertainty shocks." *econometrica* 77 (3): 623–685.

Bolton, Patrick, and Antoine Faure-Grimaud. 2009. "Thinking ahead: the decision problem." *The Review of Economic Studies* 76 (4): 1205–1238.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. "Salience theory of choice under risk." *The Quarterly journal of economics* 127 (3): 1243–1285.

———. 2013a. "Salience and asset prices." *American Economic Review* 103 (3): 623–28.

———. 2013b. "Salience and consumer choice." *Journal of Political Economy* 121 (5): 803–843.

Chinco, Alex, and Vyacheslav Fos. 2021. "The sound of many funds rebalancing." *The Review of Asset Pricing Studies* 11 (3): 502–551.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A Lochstoer. 2016. "Parameter learning in general equilibrium: The asset pricing implications." *American Economic Review* 106 (3): 664–98.

Cujean, Julien, and Rémy Praz. 2015. "Asymmetric information and inventory concerns in over-the-counter markets." *Available at SSRN 2464399.*

Epstein, Larry G, Jawwad Noor, Alvaro Sandroni, et al. 2010. "Non-bayesian learning." *The BE Journal of Theoretical Economics* 10 (1): 1–20.

Eriksen, Charles W, and James D St James. 1986. "Visual attention within and around the field of focal attention: A zoom lens model." *Perception & psychophysics* 40 (4): 225–240.

Gershman, Samuel J, Eric J Horvitz, and Joshua B Tenenbaum. 2015. "Computational rationality: A converging paradigm for intelligence in brains, minds, and machines." *Science* 349 (6245): 273–278.

Glosten, Lawrence R, and Paul R Milgrom. 1985. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders." *Journal of financial economics* 14 (1): 71–100.

Godley, Alfred Denis, et al. 1920. "Histories." *Herodotus Complete Works* 7 (50.1).

Ilut, Cosmin, and Rosen Valchev. 2017. "Economic agents as imperfect problem solvers."

Kozlowski, Julian, Laura Veldkamp, and Venky Venkateswaran. 2020. "The tail that wags the economy: Beliefs and persistent stagnation." *Journal of Political Economy* 128 (8): 2839–2879.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2018. "Dynamic rational inattention: Analytical results." *Journal of Economic Theory* 176:650–692.

Maćkowiak, Bartosz, and Mirko Wiederholt. 2009. "Optimal sticky prices under rational inattention."

———. 2010. "Business cycle dynamics under rational inattention."

———. 2011. "Inattention to rare events."

———. 2018. "Lack of Preparation for Rare Events." *Journal of Monetary Economics.*

Martin, Ian WR, and Robert S Pindyck. 2015. "Averting catastrophes: the strange economics of Scylla and Charybdis." *American Economic Review* 105 (10): 2947–85.

Matejka, Filip, and Alisdair McKay. 2015. "Rational inattention to discrete choices: A new foundation for the multinomial logit model." *American Economic Review* 105 (1): 272–98.

Matějka, Filip. 2015a. "Rationally inattentive seller: Sales and discrete pricing." *The Review of Economic Studies,* rdv049.

———. 2015b. "Rigid pricing and rationally inattentive consumer." *Journal of Economic Theory* 158:656–678.

Nimark, Kristoffer. 2014. "Man-bites-dog business cycles." *The American Economic Review* 104 (8): 2320–2367.

Nimark, Kristoffer P, and Savitar Sundaresan. 2019. "Inattention and belief polarization." *Journal of Economic Theory* 180:203–228.

Orlik, Anna, and Laura Veldkamp. 2014. *Understanding uncertainty shocks and the role of black swans.* Technical report. National Bureau of Economic Research.

Peters, Achim, Bruce S McEwen, and Karl Friston. 2017. "Uncertainty and stress: Why it causes diseases and how it is mastered by the brain." *Progress in neurobiology* 156:164–188.

Posner, Michael I, Charles R Snyder, and Brian J Davidson. 1980. "Attention and the detection of signals." *Journal of experimental psychology: General* 109 (2): 160.

Posner, Michael I, Charles R Snyder, and R Solso. 2004. "Attention and cognitive control." *Cognitive psychology: Key readings* 205.

Shaw, Marilyn L, and Peter Shaw. 1977. "Optimal allocation of cognitive resources to spatial locations." *Journal of Experimental Psychology: Human Perception and Performance* 3 (2): 201.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of monetary Economics* 50 (3): 665–690.

Steiner, Jakub, Colin Stewart, and Filip Matějka. 2017. "Rational Inattention Dynamics: Inertia and Delay in Decision-Making." *Econometrica* 85 (2): 521–553.

Sundaresan, Savitar, and Enrico Zanardo. 2019. "An Axiomatically Defined Cost of Information." *Unpublished Manuscript.*

Tenenbaum, Joshua B, Charles Kemp, Thomas L Griffiths, and Noah D Goodman. 2011. "How to grow a mind: Statistics, structure, and abstraction." *science* 331 (6022): 1279–1285.

Van Nieuwerburgh, Stijn, and Laura Veldkamp. 2009. "Information immobility and the home bias puzzle." *The Journal of Finance* 64 (3): 1187–1215.

———. 2010. "Information acquisition and under-diversification." *The Review of Economic Studies* 77 (2): 779–805.

Veronesi, Pietro. 1999. "Stock market overreactions to bad news in good times: a rational expectations equilibrium model." *The Review of Financial Studies* 12 (5): 975–1007.

Woodford, Michael. 2012. "Inattentive valuation and reference-dependent choice." *Unpublished Manuscript, Columbia University.*

A Proofs

Proof. Proof that the Bellman is a contraction. The Bellman equation is:

$$V(p_t) = \max_{p_L, p_R \in \left[\frac{1}{2}, 1\right]} U(p_t) - c(p_L) - c(p_R) + \beta p_t V(p_L) + \beta(1-p_t)V(p_R)$$

I show here that the equation satisfies Blackwell's sufficient conditions and thus describes a contraction mapping. The value function V is bounded, as $\frac{1}{2} \leq p_t \leq 1$, and U is a real-valued function. The Value function is a mapping from $\left[\frac{1}{2}, 1\right]$ to $\left[\frac{U\left(\frac{1}{2}\right)}{1-\beta}, \frac{U(1)}{1-\beta}\right]$. Therefore the Bellman equation describes a self-map on the space of bounded functions $B(X)$. To show that the mapping T from space of functions onto itself is a contraction, we must show that:

1. Monotonicity: if $f, g \in B(X)$, and $f(x) < g(x)$ for all $x \in X$, then $Tf(x) \leq Tg(x)$ for all $x \in X$.

2. Discounting: For $\gamma \in \mathbb{R}$, there exists a β such that for all $f \in B(X)$, and all $x \in X$, $T(f+\gamma)(x) \leq Tf(x) + \beta\gamma$.

First, monotonicity:

$$\begin{aligned} Tf(p_t) &= \max_{p_L, p_R \inf\left[\frac{1}{2}, 1\right]} U(p_t) - c(p_{L,f}) - c(p_{R,f}) + \beta p_t f(p_{L,f}) + \beta(1-p_t) f(p_{R,f}) \\ &\leq \max_{p_L, p_R \inf\left[\frac{1}{2}, 1\right]} U(p_t) - c(p_{L,f}) - c(p_{R,f}) + \beta p_t g(p_{L,f}) + \beta(1-p_t) g(p_{R,f}) \\ &\leq \max_{p_L, p_R \inf\left[\frac{1}{2}, 1\right]} U(p_t) - c(p_{L,g}) - c(p_{R,g}) + \beta p_t g(p_{L,g}) + \beta(1-p_t) g(p_{R,g}) \\ &= Tg(p_t) \end{aligned}$$

where the first equality is a definition, the second follows from the fact that $f < g$ everywhere, and the third follows from the the agent's optimization. The last equality is again a definition.

Second, discounting:

$$\begin{aligned} T(f+\gamma)(p_t) &= \max_{p_L, p_R \inf\left[\frac{1}{2}, 1\right]} U(p_t) - c(p_{L,f}) - c(p_{R,f}) + \beta p_t f(p_{L,f}) + \beta(1-p_t) f(p_{R,f}) + \beta\gamma \\ &= Tf(p_t) + \beta\gamma \end{aligned}$$

□

Proof. Proposition(1)

The first order conditions for the agent's problem are:

$$\begin{aligned}
V'(p_t) &= U'(p_t) + \beta(V(p_L) - V(p_R)) \\
c'(p_L) &= \beta p_t V'(p_L) + \mu_L \\
c'(p_R) &= \beta(1-p_t)V'(p_R) + \mu_R \\
\mu_L\left(p_L - \frac{1}{2}\right) &= 0 \\
\mu_R\left(p_R - \frac{1}{2}\right) &= 0
\end{aligned}$$

Substituting, we get that:

$$\begin{aligned}
\frac{c'(p_L) - \mu_L}{\beta p_t} - U'(p_L) - \beta(V(p_L^*(p_L)) - V(p_R^*(p_L))) &= 0 \\
\frac{c'(p_R) - \mu_R}{\beta(1-p_t)} - U'(p_R) - \beta(V(p_L^*(p_R)) - V(p_R^*(p_R))) &= 0
\end{aligned}$$

If the solutions are not interior and μ_L or μ_R are not zero, then the constraints bind. Let us therefore, consider the cases where $\mu_L = 0$ and $\mu_R = 0$ in turn. Deriving both sides of the first equation with respect to p_t yields the following (by the envelope condition the Vs are already optimized, and so do not change):

$$-\frac{c'(p_L)}{\beta p_t^2} + \frac{c''(p_L)p_L'(p_t)}{\beta p_t} - U''(p_L)p_L'(p_t) = 0$$

Rearranging we get:

$$p_L'(p_t) = \frac{-c'(p_L)/p_t}{U''(p_L)\beta p_t - c''(p_L)} > 0$$

Because $c'' > U''$ everywhere by assumption. Following a similar logic for the rare state:

$$\frac{c'(p_R)}{\beta(1-p_t)^2} + \frac{c''(p_R)p_R'(p_t)}{\beta(1-p_t)} - U''(p_R)p_R'(p_t) = 0$$

Rearranging we get:

$$p_R'(p_t) = \frac{c'(p_L)/(1-p_t)}{U''(p_L)\beta(1-p_t) - c''(p_L)} < 0$$

Because $c'' > U''$ everywhere. □

Proof. It is straightforward to see that $p_R\left(\frac{1}{2}\right) = p_L\left(\frac{1}{2}\right)$. Then the first-order conditions with respect to the choice variabless can be written as:

$$\frac{c'\left(\frac{1}{2}\right)}{\beta(1-p)} = U'\left(\frac{1}{2}\right) + \mu_R$$

The value p^* for which $\mu_R = 0$, is the point past which the agent stops collecting information. For all values of p larger than p^*, p_R is constrained at $\frac{1}{2}$. Therefore,

$$\frac{c'\left(\frac{1}{2}\right)}{\beta(1-p)} \geq U'\left(\frac{1}{2}\right)$$

□

Proof. Proposition (2) Suppose $p^* > \frac{1}{2}$. We know that $p'_R(p_t) < 0$. Therefore $p_R\left(\frac{1}{2}\right) = p_L\left(\frac{1}{2}\right) > \frac{1}{2}$. We also know that $p'_L(p_t) > 0$ and $p_L(1) < 1$ due to the limit conditions on c and U. Therefore, there must exist a point $\tilde{p}$ such that $p_L(\tilde{p}) = \tilde{p}$. □

Proof. Proposition (3) The first order conditions of the agent's problem are:

$$\begin{aligned} V'(p_t) &= U'(p_t + s_t) + \beta V(p_L) - \beta V(p_R) - c'_s(p_t + s_t) + c'_s(p_t) \\ 0 &= U'(p_t + s_t) + \beta V(p_L) - \beta V(p_R) - c'_s(p_t + s_t) + \lambda \\ 0 &= -c'(p_L) + \beta p_t V'(p_L) + \mu_L \\ 0 &= -c'(p_R) + \beta(1 - p_t)V'(p_R) + \mu_R \end{aligned}$$

Substituting, as before, we get:

$$\begin{aligned} c'_s(p_L(p_t + s_t)) - \lambda &= \frac{c'(p_L(p_t + s_t)) - \mu_L}{\beta(p_t + s_t)} \\ c'_s(p_R(p_t + s_t)) - \lambda &= \frac{c'(p_R(p_t + s_t)) - \mu_L}{\beta(1 - p_t - s_t)} \\ U'(p_t + s_t) + \beta V(p_L) - \beta V(p_R) + \lambda &= c'_s(p_t + s_t) \end{aligned}$$

The first two of these conditions are identical to those from the previous section, which means that conditional on a level of s_t, the same static results will hold. □

Proof. Proposition (4) There exists a $\bar{p}$ such that for all $p_t < \bar{p}$, $s_t = \bar{p} - p_t$.

This desiderata is satisfied by the condition

$$U'(p_t + s_t) + \beta V(p_L) - \beta V(p_R) + \lambda = c'_s(p_t + s_t)$$

from the previous proof. If $\lambda = 0$, then a marginal change in p_t yields the following equality (again the Vs drop out via the envelope condition):

$$U''(p_t + s_t)(1 + s'_t(p_t)) = c''_s(p_t + s_t)(1 + s'_t(p_t))$$

Because $c''_s > U''$ everywhere by assumption, this equation is uniquely satisfied by $s'_t(p_t) = -1$. If $\lambda > 0$, then λ adjusts to satisfy the equation because s_t is bounded below by zero by assumption. □

Proof. Proposition (5) The first order conditions are:

$$\begin{aligned}
V_1(p_1, ..., p_n) &= (ln(p_1) + 1)\, U'(-H) + \beta V(p_{1,1}, ..., p_{1,n}) + \lambda \\
&\vdots \\
V_n(p_1, ..., p_n) &= (ln(p_n) + 1)\, U'(-H) + \beta V(p_{n,1}, ..., p_{n,n}) + \lambda \\
c'(p_{i,1}) &= \beta p_i V_1(p_{i,1}, ..., p_{i,n}) + \mu_i \\
&\vdots \\
c'(p_{i,n}) &= \beta p_i V_n(p_{i,1}, ...p_{i,n}) + \mu_i
\end{aligned}$$

Summing across first order conditions we get the following:

$$\sum_{j \neq k}^{n} c'(p_{i,j}) - \beta p_i V_j(p_{i,1}, ..., p_{i,n}) = (n-1)c'(p_{i,k}) - \beta(n-1)p_i V_k(p_{i,1}, ..., p_{i,n})$$

And then substituting in for V_j we get:

$$\begin{aligned}
&\sum_{j \neq k}^{n} \left[c'(p_{i,j}) - \beta p_i \left((ln(p_{i,j}) + 1)\, U'(-H_i) + \beta V(p_{ij,1}, ..., p_{ij,n}) + \lambda\right)\right] \\
= \; & (n-1)c'(p_{i,k}) - \beta(n-1)p_i \left((ln(p_{i,k}) + 1)\, U'(-H_i) + \beta V(p_{ik,1}, ..., p_{ik,n}) + \lambda\right)
\end{aligned}$$

Rearranging terms:

$$\begin{aligned}
\sum_{j\neq k}^{n} c'(p_{i,j}) - (n-1)c'(p_{i,k}) &= \sum_{j\neq k}^{n} \beta p_i \left((ln(p_{i,j})+1)\, U'\left(-H_i\right) + \beta V(p_{ij,1},...,p_{ij,n}) + \lambda \right) \\
&\quad -\beta(n-1)p_i \left((ln(p_{i,k})+1)\, U'\left(-H_i\right) + \beta V(p_{ik,1},...,p_{ik,n}) + \lambda \right) \\
\sum_{j\neq k}^{n} c'(p_{i,j}) - (n-1)c'(p_{i,k}) &= \beta p_i \left[\sum_{j\neq k}^{n} (ln(p_{i,j})+1) - (n-1)\left(ln(p_{i,k})+1\right) \right] U'\left(-H_i\right) \\
&\quad +\beta^2 p_i \left(\sum_{j\neq k}^{n} V(p_{ij,1},...,p_{ij,n}) - (n-1)V(p_{ik,1},...,p_{ik,n}) \right) \\
\frac{\sum_{j\neq k}^{n} c'(p_{i,j}) - (n-1)c'(p_{i,k})}{\beta p_i} &= \left[\sum_{j\neq k}^{n} ln(p_{i,j}) - (n-1)ln(p_{i,k}) \right] U'\left(-H_i\right) \\
&\quad +\beta \left(\sum_{j\neq k}^{n} V(p_{ij,1},...,p_{ij,n}) - (n-1)V(p_{ik,1},...,p_{ik,n}) \right)
\end{aligned}$$

There are now n many conditions of this form in this version of the model. We will consider states of the world where at least some information has been collected, otherwise the boundary condition is hit. Let us pick c such that $c'(x) = K\ln(x)$. And let us define $X_k(\{p_{il}\}) \equiv \sum_{j\neq k}^{n} ln(p_{i,j}) - (n-1)ln(p_{i,k})$. Then taking the derivative of both sides with respect to p_i (again by the envelope condition the Vs are unaffected), we get:

$$-\frac{KX_k(\{p_{il}\})}{\beta p_i^2} + \frac{KX_k'(\{p_{il}\})}{\beta p_i} = X_k'(\{p_{il}\})U'(-H_i) - X_k(\{p_{il}\})U''(-H_i)H_i'(p_i)$$

Rearranging we get:

$$H_i'(p_i) = \frac{1}{U''(-H_i)} \left(-\frac{X_k'(\{p_{il}\})}{X_k(\{p_{il}\})} \left(\frac{K}{\beta p_i} - U'(-H_i) \right) + \frac{K}{\beta p_i^2} \right)$$

Call the righthand side of this expression Y. While the exact value of $-\frac{X_k'(\{p_{il}\})}{X_k(\{p_{il}\})}$ depends on the manner in which entropy is adjusted, I will show sufficiency by assuming that if one chooses to reduce entropy, one increases the largest probability, $p_{i,max}$ and proportionally decreases all other probabilities. In such a case $p_{ij} = \frac{1-p_{i,max}}{n-1}$ and $p_{ij}' = \frac{-p_{i,max}'}{n-1}$. Therefore, for any k, one can show that:

$$-\frac{X_k'(\{p_{il}\})}{X_k(\{p_{il}\})} = \frac{p_{i,max}' \left(\frac{1}{1-p_{i,max}} + \frac{1}{p_{i,max}} \right)}{\ln\left(\frac{1-p_{i,max}}{p_{i,max}(n-1)} \right)}$$

The denominator of this expression is always negative, because $p_{i,max} \times n > 1$. The numerator can be positive or negative depending on whether or not one wants to increase or decrease entropy. Because $U'' < 0$ by assumption, there is always a $K > 0$ such that $\left(\frac{K}{\beta p_i} - U'(-H_i)\right)$ is negative, which in turn would guarantee that $H_i'(p_i) = Y < 0$ and $p_{i,max}' > 0$ solves the first order condition. $\square$

B Comparison to Other Models

In this section, I consider two existing approaches to modeling uncertainty persistence, which are (i) a parameter-learning Bayesian agent in a regime-switching model, and (ii) a contemporaneous rational inattention model.

B.1 Regime-Switching and Parameter Learning

State Structure: As before, suppose that there are two possible states s of the world each period A and B. However, unlike before, the unconditional distribution of the states is not perfectly known, and can take one of two values. $\pi_H = P(s_{t+1} = A) > 0.5$ is the low-volatility regime, where more often than not, the state of the world is A. $\pi_L = P(s_{t+1} = A) = 0.5$ is the high-volatility regime, where each state of the world occurs with equal probability. The regime-switching process is governed by a Markov process with symmetric transition matrix P:

$$P = \begin{pmatrix} q & 1-q \\ 1-q & q \end{pmatrix}$$

where $1 > q >> 0.5$, so regimes are likely to persist.

Agent: As before, there is one Bayesian agent. In this setup, the agent faces no decision problem, but merely observes the state of the world each period (A or B) and updates her beliefs over the regime π_t. In a given period t, the agent has a belief $\tilde{\pi}_t = P(\pi_t = \pi_H)$. At any given point in time, an agent enters a period with belief $\tilde{\pi}_t$, about the regime, and beliefs $\tilde{p}_t$ about the state of the world. Necessarily, this implies that there are two ways to define or think about uncertainty, unlike the model of the previous section. The first is uncertainty over the data-generating process, which we can call *Knightian uncertainty*, and the second is uncertainty over tomorrow's state of the world which we can call *risk*.

Updating Solution

If an agent observes the value $s_t = A$, she updates her beliefs about π using Bayes' rule:

$$\tilde{\pi}_t = \frac{\pi_H(\tilde{\pi}_{t-1}q + (1 - \tilde{\pi}_{t-1})(1 - q))}{\pi_H(\tilde{\pi}_{t-1}q + (1 - \tilde{\pi}_{t-1})(1 - q)) + \pi_L(\tilde{\pi}_{t-1}(1 - q) + (1 - \tilde{\pi}_{t-1})q)} \tag{7}$$

A similar equation holds for how the agent updates when she observes the value $s_t = B$. Therefore, her belief over the value of the state in period $t + 1$ is given by:

$$\tilde{p}_t = \tilde{\pi}_t \pi_H + (1 - \tilde{\pi}_t)\pi_L$$

Results

It is straightforward to see that if $\tilde{\pi}_{t-1}$ is very close to 1, then the realization of $s_t = B$ will cause a sharp spike in $\tilde{\pi}_t$, which will persist, even if $s_{t+k} = A$ for all $k \geq 1$.

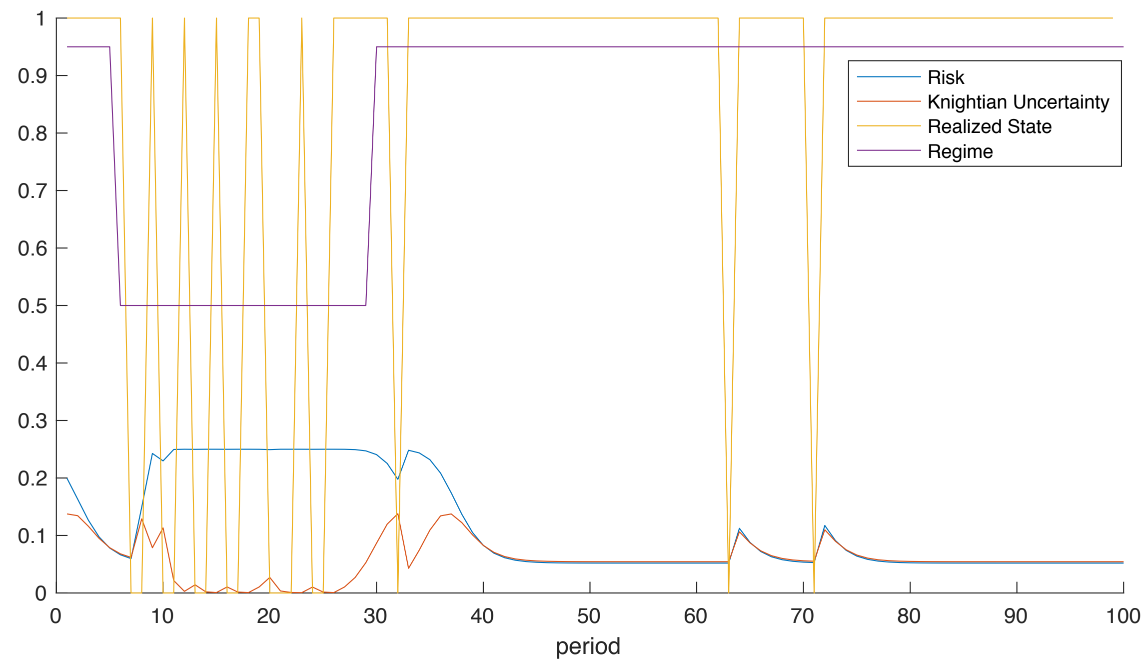
Lemma 2. *Knightian Uncertainty and risk do not comove.*

This proof is trivial - for high values of $\tilde{\pi}_t$, the two measures will comove, as agents will be very sure as to which regime they are in, as well as being very sure of the next period's state, because the regime itself is very low-risk. However, for low values of $\tilde{\pi}_t$, agents again are very sure of which regime they are in (so Knightian Uncertainty is low), but the regime itself is very high-risk, so perceptions of risk remain high.

Remark 1. *Risk is fully determined by the regime, so risk remains high, even when uncertainty over the regime is low.*

Figures

Consider the following simulation. Here $\pi_H = 0.9$, $\pi_L = 0.5$, $q = 0.99$. I initialize the model with $\tilde{\pi}_0 = 0.5$ and $\tilde{p}_0 = 0.725$. I simulate the model for 100 periods. The purple line shows the realized regime - low volatility for 5 periods, high volatility for 28 periods, and then low volatility for 63 periods. The yellow line shows the realized states, 1 for A and 0 for B. There are several points to note about this simulation, particularly in apposition to the model of the last section. When the agent has a high degree of confidence that she is in the low-volatility regime, uncertainty and risk comove, and look very similar to the results of the last section: uncommon realizations of state B cause both measures to rise, before slowly declining as A is observed more frequently.



However, when the agent is uncertain about which regime she is in, or when she is certain that she is in the high-volatility regime, the two measures no longer comove, and no longer exhibit the sharp increase and slow decrease we observed in the data and in the previous model. Instead, perceptions of risk (blue line) stay high, as the agent slowly starts to adjust her beliefs towards the high volatility regime. And once the agent is convinced she is in the high voltility regime, it takes several realizations of A for her to adjust her uncertainty (red line) slowly upwards and her risk slowly downwards.

Comparison to Baseline Model

The new theory laid out in this paper differs from the parameter-learning setup presented here in several ways. First, the model of section 2 *requires fewer assumptions about the data-generating process*. That model delivers the patterns of sharp increases and slow decreases in uncertainty for almost any unconditional state-distribution. For a parameter-learning model to do the same, we needed regime-switching, else agents would have learned everything quickly and never been uncertain again.

Second, section 2's formulation of *uncertainty can only jump up, and never down, which is a feature of the world*. When agents are uncertain in that model, they can only gradually regain certainty through preparation. Even if *realized* volatility vanishes when agents are uncertain (that is, the modal outcome always occurs), the modal outcome in a dispersed distribution is still not *very* likely ex-ante, so agents will not prepare for it *much* more than other outcomes. Put differently, in a normal distribution, there is no limit to how unlikely an event can be, but there are limits to how likely events can be, so there are limits as to how quickly preparation can recover. In the parameter-learning context presented in this section, a sudden drop in uncertainty can cause a sudden drop in the estimation of the volatility parameter. We do not typically observe sudden reductions in uncertainty, so this difference is a feature of this paper's model.

Third, the model of section 2 matches patterns in uncertainty and risk during *high volatility periods*. Sustained periods of volatility will cause initial spikes in such estimates that subside somewhat. The basic intuition for why is that uncertain agents in this paper will view future large shocks as being relatively more likely than when they were certain and will thus prepare for them slightly more when they are uncertain. Such a pattern is observable in Figure 1 even during periods of sustained volatility like the financial crisis. In the parameter-learning framework of this section, sustained periods of high volatility see sharp and continued increases in estimated volatility. The more high volatility parameter-learners observe, the more they believe that their estimate of volatility should be high. This pattern is typically not observed in the data.

Finally, the baseline model is testing a *fundamentally different type of uncertainty* from that generated by parameter-learning models. Parameter-learning agents are uncertain due to a lack of understanding of their regime. Therefore, uncertainty in such a model refers to variance in agents' beliefs about a parameter's value that changes each period as their information sets change. Put more formally, uncertainty in a parameter-learning model refers to $Var_t[X|I_t]$, the variance in agents' beliefs at time t, conditional on any and all information available to them in time t about a deep parameter X that will never be directly observed. By contrast, agents in the baseline model know exactly how the world in which they live works. There is only one regime. However, due to their lack of preparation for unexpected events, such events will make them very uncertain about *what happens next*. Uncertainty in this paper refers to the variance in agents' beliefs about a variable that will be revealed to them imminently. More formally, uncertainty in this paper refers to $Var_t[X_{t+1}|I_t]$, the variance in agents' beliefs at time t, conditional on any and all information available to them in time t about a variable X_{t+1} that will be revealed to them in the subsequent period. This latter definition of uncertainty conforms to how uncertainty is described in the uncertainty shocks literature, started by Bloom (2009).

The two approaches are complementary. The sustained increase in the level of the VIX post-COVID relative to pre-COVID could be attributed to attitudes regarding regimes or risk (which would be a parameter learning story), while the frequent jumps and gradual declines observed in the VIX at a higher frequency would be consistent with the mechanism of the baseline model.

Irrational Beliefs and Learning

Another way to deal with the speed of learning by Bayesian learners is to introduce an irrationality in the way in which they process information. For example, the agent could be predisposed to overreact to recent information. Consider, for example, an agent with the ability or prediliction to overreact or underreact to information relative to the Bayesian benchmark. I will model this hewing to the formulation of Epstein, Noor, Sandroni, et al. (2010). Suppose that the environment is the same as in B.1 - that is that there are two possible states of the world, and that the unconditional distribution is not perfectly known, and changes with a probability q. Then an agent's posterior beliefs after observing a signal would take the form:

$$\tilde{\pi}_{t+1} = (1 - \gamma_t)\tilde{\pi}_{t+1,BU} + \gamma_t \tilde{\pi}_t$$

where $\tilde{\pi}_t = P(\pi_t = \pi_H)$, $\tilde{\pi}_{t+1,BU}$ being the 'rational' updated beliefs after observing a signal (if the agent sees $s_t = A$, this expression is given by equation 7) and $\gamma_t < 1$ guides the degree to which the agent's posterior beliefs are influenced by their prior beliefs. If $\gamma_t > 0$, the agent underreacts to new information, and is biased towards their prior beliefs; if $\gamma_t < 0$ the agent overreacts to new information, and is biased away from their prior beliefs.

Epstein, Noor, Sandroni, et al. (2010) consider a stationary environment, and show that if $\gamma_t > 0$, then convergence to the truth is achieved, albeit more slowly than with a 'rational' Bayesian agent, but that if $\gamma_t < 0$ then convergence to incorrect beliefs is also possible. Here, I will consider two possibilities.

Case 1: $\gamma_t > 0$. In this case, the general qualitative patterns of uncertainty from the 'rational' model above remain, but the speed of learning is slowed (possibly significantly, depending on γ_t). For the sake of intuition, suppose that the economy stays for a long time in the 'high volatility' regime. The 'rational' Bayesian will, as was seen above, quickly learn the truth, and Knightian uncertainty will be low. The underreacting Bayesian will slowly learn the truth, but for that agent too, Knightian uncertainty will gradually decrease. However, while Knightian uncertainty is decreasing for both agents, risk perceptions will remain high (again, for both agents) as they will both acknowledge that volatility is, in fact, high.

Case 2: $\gamma_t < 0$**.** In this case, we need an additional restriction on γ_t, namely that γ_t must be small enough in absolute value such that $\tilde{\pi}_{t+1} \in (0,1)$. For sufficiently large overreactions, posterior beliefs could hit the boundaries of the belief space, which would be absorbing. In Epstein, Noor, Sandroni, et al. (2010) this is paralleled by assumption (A.3) in the appendix. If this condition on γ_t is satisfied, then the period-by-period volatility of beliefs will increase, as the agent will over-correct towards the high volatility regime every time there is state B, and the agent will over-correct towards the low volatility regime every time there is a state A. Again, for the sake of intuition if we suppose that the economy stays for a long time in the 'high volatility' regime, a negative γ_t agent could potentially converge to the correct or incorrect belief. Therefore, risk-perceptions will be poorer than in Case 1 - the agent could believe, correctly, that volatility is high, or incorrectly that volatility is low, while knightian uncertainty remains low. This contrasts with the agent of the baseline model who never has any knightian uncertainty, but is also never incorrect about expected volatility.

B.2 Dynamic Inattention Over the State of the Economy

Another way of modeling dynamic learning with inattention would be as follows. Suppose that there's a process $X_t \equiv s_t + \epsilon_{X,t}$, where $\epsilon_{X,t} \sim N(0, n_{t-1}^{-1})$, and suppose that the state s follows an AR(1) structure, so that $s_t = \lambda s_{t-1} + \sigma \epsilon_{s,t}$ where $\epsilon_{s,t}$ is a standard normal. Suppose further that an agent's information set can be summarized by the mean (μ) and precision ($\hat{\tau}$) of her beliefs after observing a signal. Then an agent could face the following problem[22]:

$$V(\mu_t, \hat{\tau}_t) = \max_{n_t \geq 0} \pi(\mu_t) - c(n_t) + \beta \int V(\mu_{t+1}, (n, x), \tau_{t+1}(n))\phi(x)dx$$

where n_t is the amount of attention paid today at a cost c, and π is a payoff function. An agent's payoff today is a function of the realization of the state today. The agent can improve the precision of her signals about the realization of the state tomorrow, subject to a cost function c. There are four key differences between the model described above and the model of this paper. As I will argue here, the four differences either do not significantly alter the spirit of the question, or are introduced to focus on a novel source of dynamic learning:

22. This subsection is inspired, with my thanks, by the careful analysis of an anonymous referee.

Comparison to Baseline Model

First, the utility of the agent in this paper depends on the precision of beliefs, not on the realized state. This assumption was introduced to focus on the effect that uncertainty has on information collection - allowing for state-dependent payoffs would result in an asymmetry in the results (holding constant the ex-ante likelihood of a state, an agent might want to learn more about "bad" states than "good" states), but the qualitative effect would remain in that there would be sufficiently unlikely bad states and sufficiently unlikely good states that an agent would not care to learn about. Formally (adapting this to a continuous state setup), this paper posits $\pi(\hat{\tau}_t)$ increasing, instead of $\pi(\mu_t)$ increasing.

Second, instead of improving the precision of a signal subject to a cost function, I assume the agent directly improves the precision of future beliefs subject to a cost function. The two methods are isomorphic under the binomial and normal settings shown in this paper - and I used the selection of precision of beliefs for notational convenience. Formally, $\max_{\tau_{t+1}\geq 0}$ instead of $\max_{n_t\geq 0}$.

Third, the agent in the problem above selects, in time t, the precision of her beliefs over states of the world in time $t+1$. The agent of this paper takes as given the precision of her beliefs in time t over states of the world in time $t+1$, and instead selects in time t, what the precision of her beliefs *will be* in time $t+1$ over states of the world in time $t+2$ *for each possible state of the world in time* $t+1$. This is a large, fundamental departure from the problem above, and is meant to reflect a realistic aspect of how agents make informational decisions. As was discussed in the introduction - agents of all types often prepare for different possible states of the world *before the states occur*: preparing for different financial scenarios in the aftermath of Brexit; preparing for different Democratic nominees in 2020; preparing for different contingencies in a shuttle launch. In all of these cases, *both* forms of learning are taking place: agents are trying to figure out how likely each scenario is (learning in time t about what will happen in time $t+1$) and are trying to figure out what to do should each scenario arise (setting, in time t, the precision of beliefs in $t+1$ about $t+2$ for each $t+1$ state). This paper shuts down the first form of learning to focus on the second, and this paper's results arise from this second form of state-contingent preparation. However, as I show in the appendix, reintroducing the first form does not eliminate this paper's results, *as long as the second form of learning is still present.* Formally, $\max_{\tau_{t+1}(x)\geq 0}$ instead of $\max_{n_t\geq 0}$.

Fourth, and finally - I assume that the agent pays for reduction in $t+1$ uncertainty at time t. Again, this is a departure from the model above, as typically, models assume that you pay at time t

for the precision of your beliefs in time t over states of the world in time $t+1$. But this departure is grounded in reality. This paper focuses on *preparation*, which entails collecting information today *that may or may not pay off*. Preparing for the possibility of Trump being reelected in 2020 required time and effort and money in 2019, even if the payoff of such preparation would not have occured until 2020, and indeed never did. Formally, $-\int c(\tau_{t+1}(x))dx$ instead of $-c(n_t)$.

Because the model presented here involves no state contingency, and because it only allows for information to be useful contemporaneously, it cannot match the predictions of this paper.