AMBIGUITY, INFORMATION QUALITY AND ASSET PRICING*

Larry G. Epstein Martin Schneider

May 19, 2004

Abstract

When ambiguity averse investors process news of uncertain quality, they act as if they take a worst-case assessment of quality. As a result, they react more strongly to bad news than to good news. They also dislike assets for which information quality is poor, especially when the underlying fundamentals are volatile. These effects induce negative skewness in asset returns, increase price volatility and induce ambiguity premia that depend on idiosyncratic risk in fundamentals. Moreover, shocks to information quality can have persistent negative effects on prices even if fundamentals do not change. This helps to explain the reaction of markets to events like 9/11/2001.

1 INTRODUCTION

When individuals make decisions under uncertainty, they receive information, or signals, about unknown parameters. Most economic models assume that signals are processed the Bayesian way: an individual is assumed to form subjective probabilistic beliefs about the parameters, and he then incorporates signal realizations into beliefs by applying Bayes' Rule, given a likelihood function that relates signals to parameters. The Ellsberg Paradox highlights one limitation of the Bayesian model: it cannot accommodate ambiguity, where the individual is not confident enough in his *prior* understanding of the environment to commit to a single (subjective) probability measure. This paper is concerned with a second feature of complicated environments that is also beyond the scope of the Bayesian approach - it is often difficult to know the reliability of signals. We incorporate *ambiguous information* into a choice-theoretic framework. We then derive its implications for behavior and study its role in financial markets.

*Department of Economics, U. Rochester, Rochester, NY, 14627, lepn@troi.cc.rochester.edu, and Department of Economics, NYU, New York, 10003, ms1927@nyu.edu. We thank Monika Piazzesi and Pietro Veronesi for comments and suggestions.

Our model of information is illustrated most easily with an example. Let θ denote a parameter that the agent wants to learn. We assume that a signal s is related to the parameter by a *family of likelihoods*:

$$s = \theta + \epsilon, \qquad \epsilon \sim \mathcal{N}\left(0, \sigma_s^2\right), \qquad \sigma_s^2 \in \left[\underline{\sigma}_s^2, \overline{\sigma}_s^2\right]. \tag{1}$$

Bayesian models focus on the special case of a single likelihood, $\underline{\sigma}_s^2 = \overline{\sigma}_s^2$, and measure the quality of information provided by the signal s via the precision $1/\sigma_\varepsilon^2$. In our model, information quality is captured by the *range of precisions* $[1/\overline{\sigma}_s^2, 1/\underline{\sigma}_s^2]$. Information quality therefore has two dimensions: the overall location of the interval determines how quickly an agent expects uncertainty to be resolved, while its width measures (lack of) confidence in the reliability of the signal. The latter dimension is unique to the case of ambiguity. In this paper, we argue that to capture the effects of, say, an improvement in information quality, it is important to distinguish these two dimensions: the economic consequences of a rightward shift in the interval of precisions differ qualitatively from those of a reduction in the interval width.

To model preferences (as opposed to merely beliefs), we use recursive multiple-priors utility, axiomatized in Epstein and Schneider [17]. The axioms describe behavior that is consistent with experimental evidence typified by the Ellsberg Paradox. They imply that an ambiguity averse agent behaves *as if* he maximizes, every period, expected utility under a worst-case belief that is chosen from a *set* of conditional probabilities. In the present paper, the set of conditional probabilities incorporates ambiguous information. These probabilities are updated by applying Bayes' Rule to the whole family of likelihoods that describe a signal. Through updating, ambiguity in signals affects the set from which worst-case conditionals can be drawn, and hence behavior.

Ambiguous information has two key effects on behavior. First, *after* a signal has arrived, agents respond asymmetrically: bad news affect conditional actions – such as portfolio decisions – more than good news. This is because agents evaluate any action using the conditional probability that minimizes utility from that action. If an ambiguous signal conveys good (bad) news, the worst case is that the signal is unreliable (very reliable). Second, even *before* an ambiguous signal arrives, agents anticipate the arrival of low quality information and discount consumption plans for which this information may be relevant.

Financial market participants absorb a large amount of news every day. To illustrate the role of ambiguous information in financial markets, we consider a representative agent asset pricing model. The agent's information consists of (i) past prices and dividends and (ii) an additional, ambiguous, signal that is informative about future dividends. Our setup thus distinguishes between *tangible* information – here dividends, and also past prices – that lends itself to econometric analysis, and *intangible* information – such as news reports – that is hard to quantify, yet important for market participants' decisions. We assume that intangible information is ambiguous while tangible information is not. We derive some general properties of prices and clarify the relationship between prices and information quality under ambiguity. Then we explore the response to sudden changes in information quality by calibrating a model to the aftermath of 9/11. This also illustrates how belief parameters can be identified from observed prices.

Reaction to ambiguous news affects the first three moments of asset prices. First, the anticipation of ambiguous information increases measured average excess returns. As with consumption plans, agents dislike assets about which ambiguous information is expected to arrive. They are willing to buy such assets only at a discount, so that a premium for low information quality emerges in equilibrium. Second, the volatility of prices and returns can be much larger than the volatility of fundamentals. Volatility depends on how much the worst-case conditional expectation of fundamentals fluctuates. If the range of precisions contemplated by ambiguity averse agents is large, they will often attach more weight to a signal than agents who know the true precision. Intangible information can thus cause large price fluctuations.

Third, the asymmetric response to ambiguous information implies that returns should be negatively skewed at high frequencies. Negative skewness here is not derived from asymmetries in the distribution of signals or fundamentals – it is simply due to agents' processing of signals under ambiguity. The model also suggests that skewness should be more pronounced for stocks for which the relevant amount of intangible information is larger. This is consistent with the fact that stocks that are "in the news" more – such as glamour stocks, stocks that have recently experienced a runup in prices, and stocks of large firms – exhibit more negative skewness.

Premia due to ambiguous information are distinct from risk premia. Under ambiguity, an asset can command a high premium even if it is uncorrelated with all other assets in the market and makes up a small share of the representative agent's portfolio. For ambiguity averse investors, uncertainty is a first-order concern: if information quality drops, an asset is treated *as if* its mean payoff has fallen. This lowers prices (and increases expected excess returns) regardless of covariance with the market. In addition, premia due to ambiguous information depend not only on the quality of information about that asset, but also on the volatility of its fundamentals. In markets where fundamentals do not move much, it is irrelevant whether information quality is high or low. In contrast, when fundamentals are volatile, information quality is much more of a concern and the premium for low quality is higher. Importantly, ambiguity averse investors fear not only market-wide, but also asset-specific ambiguous information. The model thus predicts the existence of a premium for idiosyncratic volatility.

Moreover, ambiguity premia are anticipatory: the prospect of lower information quality, perhaps triggered by an announcement or other event, is sufficient to lower asset prices. In contrast, in a Bayesian model changes in future information quality are irrelevant for current prices. This result suggests that conclusions commonly drawn from event studies should be interpreted with caution. For example, a negative abnormal return after a merger announcement need not imply that the market views the merger as a bad idea. Instead it might simply reflect the market's discomfort in the face of the upcoming period of ambiguous information. Importantly, a discount due to low future *idiosyncratic* information quality cannot be captured by the Bayesian model: if lower information quality is captured by higher risk, then it should be diversified away.

The importance of ambiguous news varies not only in the cross-section, but also over time. Ambiguous information is particularly prevalent when payoff-relevant news are unfamiliar to market participants. This leads us to consider shocks to information quality. Asset markets often witness events that simultaneously (i) increase uncertainty about fundamentals and (ii) change the nature of signals relevant for forecasting fundamentals. One example is the terrorist attack of September 11, 2001. This shock both increased uncertainty about future growth *and* shifted the focus to hitherto "unfamiliar" news about foreign policy and terrorism. Since the shock increased uncertainty, it marked the start of a learning process that affected prices. Since the news were unfamiliar, it is natural to model this process as learning from ambiguous signals.

Shocks to information quality can have drawn out negative effects on prices even if fundamentals do not change. The initial drop in the stock market when it reopened on September 17 was followed by more losses over the following week, before a gradual rebound occurred. With hindsight, we know that no long term structural change occurred: the shock changed only information quality, not fundamentals. Thus a Bayesian model with known signal quality has problems explaining the initial slide in prices. Roughly, if signal precision is high, the arrival of enough bad news to explain the first week is highly unlikely. If signal precision is low, bad news will not be incorporated into prices in the first place. In our model, where signal precision is unknown, bad news are taken especially seriously and hence a much less extreme sequence of signals suffices to account for prices in the first week. In sum, ambiguous information can help to rationalize the delayed negative response observed after a shock to information quality.

Our assumption that tangible information is unambiguous is a simplification. More generally, one would expect that all information is at least somewhat ambiguous. However, it is plausible that intangible information is more ambiguous than tangible information. Econometric analysis can often help narrow down the range of precisions for tangible signals: for example, one can study the predictability of stock returns by the price dividend ratio. As discussed in Epstein and Schneider [19], learning can resolve ambiguity over time, so that one might expect a narrower range of precisions for tangible signals. In contrast, quantitative analysis that determines the true precision of news reports is almost impossible.

With ambiguous signals, the relationship between information quality and the volatility of prices and returns is very different from that with risky (or noisy) signals. With noisy signals, better information quality (that is, higher precision) simply means that information about future cash flows is revealed earlier, when the cash flows are discounted more heavily. As a result, changes in information quality affect the volatility of prices and returns in opposite directions. For example, with earlier release of information, prices fluctuate more, but returns fluctuate less. In our framework, higher information quality can also mean that news becomes less ambiguous, or easier to interpret. Rather than speed up the temporal resolution of uncertainty, this changes how shocks to fundamentals feed through to prices and returns at a point in time via the interpretation of signals. Changes in information quality then affect the volatility of prices and returns in the same direction. For example, a reduction of ambiguity – perhaps due to improvements in information technology – can increase the volatility of both prices and returns.

The paper is organized as follows. Section 2 presents a simple thought experiment – related to the Ellsberg Paradox – to clarify the concept of ambiguous information and our modeling approach. It also reviews recursive multiple-priors utility. Section 3 discusses a simple representative agent model and derives general properties. Here we also contrast the Bayesian and ambiguity aversion approaches to thinking about information quality and asset pricing. Section 4 considers the calibrated model of 9/11 as an example of shocks to information quality. Section 5 discusses related literature. Proofs are collected in an appendix.

2 AMBIGUOUS INFORMATION

In this section, we discuss two experiments that illustrate how ambiguity aversion can imply behavior that is both intuitive and inconsistent with the standard expected utility model. The first is the classic Ellsberg Paradox that was the starting point for the large experimental literature on ambiguity. A second (thought) experiment clarifies the concept of ambiguous information. We then discuss an example with normal distributions, already partly described in the introduction, that is the key tool for our applications. Finally, we discuss the axiomatic underpinnings of our approach as well as its connection to the more general model of learning under ambiguity introduced in Epstein and Schneider [19].

2.1 Experiments

Experiment 1 (Ellsberg Paradox). Consider two urns, each containing four balls that are either black or white. The agent is told that the first "risky" urn contains two balls of each color. For the second "ambiguous" urn, he is told only that it contains at least one ball of each color. It is announced that one ball will be drawn from each urn. The agent is invited to bet on their color. Any bet (on a ball of some color drawn from some urn) pays one dollar (or one util) if the ball has the desired color and zero otherwise.

Intuitive behavior pointed to by Ellsberg, and subsequently documented in many experiments, is the preference to bet on drawing black from the risky urn as opposed to the ambiguous one, and a similar preference for white. The paradox is that decision-makers who form a single subjective probability over the composition of the ambiguous urn cannot exhibit such behavior. Indeed, strict preference for black from the risky urn reveals that the subjective probability of black for the ambiguous urn is less than $\frac{1}{2}$ (that is, the objective probability of black for the risky urn). At the same time, strict preference for white from the risky urn reveals that the subjective probability of black for the ambiguous urn is more than $\frac{1}{2}$, a contradiction. An alternative way to think about Ellsberg-type behavior is that a decision-maker forms a (subjective) *range* of probabilities about the composition of the ambiguous urn. He then evaluates bets by calculating the worst-case expected utility. For example, suppose that the range of probabilities of black is the interval $\left[\underline{p}, \bar{p}\right]$. The worst-case expected utility of a bet on black is then $\underline{p}$ for a bet on black, and $1 - \bar{p}$ for a bet on white. Since the objective probability of black from the risky urn is $\frac{1}{2}$, Ellsberg-type behavior follows whenever $\underline{p} < \frac{1}{2} < \bar{p}$.

To emphasize the relevance for asset pricing, it is helpful to view the Ellsberg Paradox as a simple portfolio choice problem under model uncertainty. Bets on black and white from the ambiguous urn are assets. The correct "model" of their payoffs is not known. A Bayesian investor treats all model uncertainty as risk – he decides on a prior over possible distributions and uses that prior to calculate conditional payoffs. Since information about the assets is symmetric here, a typical Bayesian would probably adopt a prior that respects symmetry. But this implies that under his conditional belief each asset pays one with probability one half, and zero otherwise, whatever the precise shape of the prior. As a result, the Bayesian would be indifferent between either bet on the ambiguous urn and a bet on the fair risky urn. The behavior pointed to by Ellsberg shows that people do not treat model uncertainty simply as risk. Instead, they behave as if they adjust the *mean* return on the ambiguous assets. This first order effect of model uncertainty will be an important theme below.

In sum, the Ellsberg Paradox arises because decision makers appear to feel more comfortable when the probabilities of uncertain events are objectively known. Gilboa and Schmeidler [22] have shown formally that this attitude can be captured by multiple priors. If S is a set of possible states of the world and $c : S \to \mathbb{R}$, is a consumption plan, they define utility by

$$U(c) = \min_{p \in \mathcal{P}} E^p \left[u(c)\right], \tag{2}$$

where u is a standard utility function, $\mathcal{P}$ is a set of probability measures on S, and E^p is the expectation under $p \in \mathcal{P}$. The model coincides with the expected utility model when beliefs are given by a single probability, that is, $\mathcal{P} = \{p\}$. More generally, the decision-maker behaves as if he evaluates the utility of a plan c under the "worst-case" probability in $\mathcal{P}$. Gilboa and Schmeidler prove that this is implied by "preference for objectively known probabilities", as described by their axioms.

Ambiguous Information

Experiment 2. Consider again a risky and an ambiguous urn. Instead of betting on the next draw, the agent is now invited to bet on the colors of two specific balls, called the "coin balls". For each urn, the color of the coin ball is determined by flipping a fair coin: it is black if the coin toss produces heads and white otherwise, where the coin tosses are independent across urns. In addition to the coin ball, each urn contains n "non-coin balls", of which exactly $\frac{n}{2}$ are black and $\frac{n}{2}$ are white. For the risky urn, it is known that $n = 4$: there are exactly two black and two white non-coin balls. In contrast, the number of non-coin balls in the ambiguous urn is unknown – there could be either $n = 2$ (one white and one black) or $n = 6$ (three white and three black) non-coin balls. The possibilities are illustrated in Figure 1.

A priori, before any draw is observed, one should be indifferent between bets on the coin ball from either urn - all these bets amount to betting on a fair coin. Suppose now

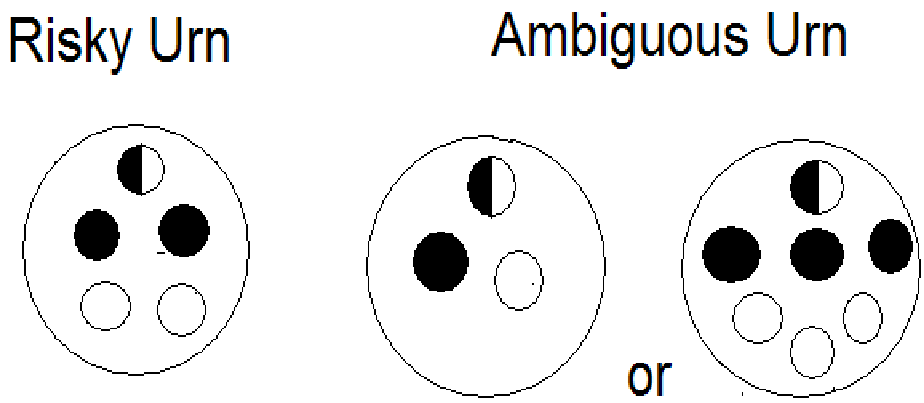

Figure 1: Risky and ambiguous urns for Experiment 2. The coin balls are drawn as half black. The ambiguous urn contains *either* $n = 2$ *or* $n = 6$ non-coin balls.

that one draw from each urn is observed and that both balls drawn are black. For the risky urn, it is straightforward to calculate the conditional probability of a black coin ball. Let n denote the number of non-coin balls. Since the unconditional probability of a black coin ball is equal to that of a black draw (both are equal to $\frac{1}{2}$), we have

$$\Pr\left(\text{coin ball black}|\text{black draw}\right) = \Pr\left(\text{black draw}|\text{coin ball black}\right) = \frac{n/2+1}{n+1},$$

and with $n = 4$ for the risky urn, the result is $\frac{3}{5}$.

The draw from the ambiguous urn is also informative about the coin ball, but there is a difference between the information provided about the two urns. In particular, it is intuitive that one would prefer to bet on a black coin ball in the risky urn rather than in the ambiguous urn. The reasoning here could be something like "if I see a black ball from the risky urn, I know that the probability of the coin ball being black is exactly $\frac{3}{5}$. On the other hand, I'm not sure how to interpret the draw of a black ball from the ambiguous urn. It would be a strong indicator of a black coin ball if $n = 2$, but it could also be a much weaker indicator, since there might be $n = 6$ non-coin balls. Thus the posterior probability of the coin ball being black could be anywhere between $\frac{6/2+1}{6+1} = \frac{4}{7} \approx .57$ and $\frac{2/2+1}{2+1} = \frac{2}{3}$. So I'd rather bet on the risky urn." By similar reasoning, it is intuitive that one would prefer to bet on a white coin ball in the risky urn rather than in the ambiguous urn. One might say "I know that the probability of the coin ball being white is exactly $\frac{2}{5}$. However, the posterior probability of the coin ball being white could be anywhere between $\frac{1}{3}$ and $\frac{3}{7} \approx .43$. Again I'd rather bet on the risky urn."

Could a Bayesian agent exhibit these choices? In principle, it is possible to construct a subjective probability belief about the composition of the ambiguous urn to rationalize the choices. However, any such belief must imply that the number of non-coin balls in the ambiguous urn depend on the color of the coin ball, contradicting the description of the experiment. To see this, assume independence and let p denote the subjective probability that $n = 2$. The posterior probability of a black coin ball given a black draw is

$$\tfrac{2}{3}p + \tfrac{4}{7}(1-p).$$

Strict preference for a bet on a black coin ball in the risky urn requires that this posterior probability be greater than $\frac{3}{5}$ and thus reveals that $p > \frac{3}{10}$. At the same time, strict preference for a bet on a white coin ball in the risky urn reveals that $p < \frac{3}{10}$, a contradiction.

This limitation of the Bayesian model is similar to that exhibited in the Ellsberg Paradox above. However, the key difference is that the Ellsberg Paradox arises in a static context, while here ambiguity is only relevant *ex post,* after the signal has been observed.

Information Quality and Multiple Likelihoods

The preference to bet on the risky urn is intuitive because the ambiguous signal – the draw from the ambiguous urn – appears to be of lower quality than the noisy signal – the draw from the risky urn. A perception of low information quality arises because the distribution of the ambiguous signal is not objectively given. As a result, the standard Bayesian measure of information quality, precision, is not sufficient to adequately compare the two signals. The precision of the noisy signal is parametrized by the number of non-coin balls n: when there are few non-coin balls that add noise, precision is high. We have shown that a single number for precision (or, more generally, a single prior over n) cannot rationalize the intuitive choices. Instead, behavior is *as if* one is using different precisions depending on the bet that is evaluated.

Indeed, in the case of bets on a black coin ball, the choice is made as if the ambiguous signal is less precise than the noisy one, so that the available evidence of a black draw is a weaker indicator of a black coin ball. In other words, when the new evidence – the drawn black ball – is "good news" for the bet to be evaluated, the signal is viewed as relatively imprecise. In contrast, in the case of bets on white, the choice is made as if the ambiguous signal is more precise then the noisy one, so that the black draw is a stronger indicator of a black coin ball. Now the new evidence is "bad news" for the bet to be evaluated, and is viewed as relatively precise. The intuitive choices can thus be traced to an asymmetric response to ambiguous news. In our model, this is captured by combining worst-case evaluation as in Gilboa-Schmeidler with the description of an ambiguous signal by multiple likelihoods.

More formally, we can think of the decision-maker as trying to *learn* the colors of the two coin balls. His *prior* is the same for both urns and simply places probability $\frac{1}{2}$ on black. The draw from the risky urn is a noisy signal of the color of the coin ball. Its (objectively known) distribution is that black is drawn with probability $\frac{3}{5}$ if the coin ball is black, and $\frac{2}{5}$ if the coin ball is white. However, for the ambiguous urn, the signal distribution is unknown. If $n = 2$ or 6 is the unknown number of non-coin balls, then black is drawn with probability $\frac{n/2+1}{n+1}$ if the coin ball is black and $\frac{n/2}{5n+1}$ if it is white. Consider now updating about the ambiguous urn conditional on observing a black draw. Bayes' Rule applied in turn to the two possibilities for n gives rise to the posterior probabilities for a black coin ball of $\frac{4}{7}$ and $\frac{2}{3}$ respectively, which leads to the range of posterior probabilities $\left[\frac{4}{7}, \frac{2}{3}\right]$.[1] If bets on the ambiguous urn are again evaluated under worst-case probabilities, then the expected payoff on a bet on a black coin ball in the ambiguous urn is $\frac{4}{7}$, strictly less than $\frac{3}{5}$, the payoff from the corresponding bet on the risky urn. At the same time, the expected payoff on a bet on a black coin ball in the ambiguous urn is $\frac{1}{3}$, strictly less than the risky urn payoff of $\frac{2}{5}$.

Normal Distributions

To write down tractable models with ambiguous signals, it is convenient to use normal distributions. The following example features a normal ambiguous signal that inherits all the key features of the ambiguous urn from Experiment 2. This example is at the heart of our asset pricing applications below. Let θ denote a parameter that the agent wants to learn about. This might be some aspect of future asset payoffs. Assume that the agent has a unique normal prior over θ, that is $\theta \sim \mathcal{N}\left(m, \sigma_\theta^2\right)$ – there is no ambiguity ex ante. Assume further that an ambiguous signal s is described by the set of likelihoods (1) from the introduction. For comparison with Experiment 2, the parameter θ here is analogous to the color of the coin ball, while the variance σ_s^2 of the mean-zero shock ε plays the same role as the number of non-coin balls in the ambiguous urn.

To update the prior, apply Bayes' rule to all the likelihoods to obtain a family of posteriors:

$$\theta \sim \mathcal{N}\left(m + \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_s^2}\left(s - m\right), \frac{\sigma_s^2 \sigma_\theta^2}{\sigma_\theta^2 + \sigma_s^2}\right), \qquad \sigma_s^2 \in \left[\underline{\sigma}_s^2, \overline{\sigma}_s^2\right]$$

Even though there is a unique prior over θ, updating leads to a nondegenerate set of posteriors – the signal induces ambiguity about the parameter. Suppose further that in each period, choice is determined by maximization of expected utility under the worst-case belief chosen from the family of posteriors. Now it is easy to see that, after a signal has arrived, the agent responds asymmetrically. For example, when evaluating a bet, or asset, that depends positively on θ, he will use a posterior that has a low mean. Therefore, if the news about θ is good ($s > m$), he will act as if the signal is imprecise (σ_s^2 high), while if the news is bad ($s < m$), he will view the signal as reliable (σ_s^2 low). As a result, bad news affect conditional actions more than good news.

2.2 A Model of Learning under Ambiguity

Recursive multiple-priors utility, axiomatized by Epstein and Schneider [17], extends the Gilboa-Schmeidler model to an intertemporal setting. Suppose that S is a finite *period* state space. One element $s_t \in S$ is observed every period. At time t, the decision-maker's information consists of the history $s^t = (s_1, ..., s_t)$. Consumption plans are sequences $c = (c_t)$, where each c_t depends on the history s^t. Given a history, preferences over future

[1] Because the agent maximizes expected utility under the worst-case probability, his behavior is identical if he uses the entire interval of posterior probabilities or if he uses only its endpoints.

consumption are represented by a conditional utility function U_t, defined recursively by

$$U_t(c; s^t) = \min_{p_t \in \mathcal{P}_t(s^t)} E^{p_t} \left[u(c_t) + \beta U_{t+1}(c; s^t, s_{t+1}) \right], \tag{3}$$

where β and u satisfy the usual properties. The set $\mathcal{P}_t(s^t)$ of probability measures on S captures conditional beliefs about the next observation s_{t+1}. Thus beliefs are determined by the whole *process* of conditional one-step-ahead belief sets $\{\mathcal{P}_t(s^t)\}$.

To clarify the connection to the atemporal case (2), it is helpful to rewrite utility using discounted sums. Consider the collection of all sets $\mathcal{P}_t(s^t)$, as one varies over times and histories. This collection determines a unique set of probability measures $\mathcal{P}$ on S^∞ satisfying the regularity conditions specified in [17].[2] Thus one obtains the following equivalent and explicit formula for utility:

$$U_t(c; s^t) = \min_{P \in \mathcal{P}} E^P \left[\Sigma_{s \geq t} \beta^{s-t} u(c_s) \mid s^t \right].$$

This expression shows that each conditional ordering conforms to the multiple-priors model in Gilboa and Schmeidler [22], with the set of priors for time t determined by updating the set $\mathcal{P}$ measure-by-measure via Bayes' Rule.

Epstein and Schneider [19] propose a particular functional form for $\{\mathcal{P}_t(s^t)\}$ in order to capture learning from a sequence of conditionally independent signals. Let Θ denote a *parameter space* that represents features of the data that the decision maker tries to learn. Denote by $\mathcal{M}_0$ a set of probability measures on Θ that represents initial beliefs about the parameters, perhaps based on prior information. Taking $\mathcal{M}_0$ to be a set allows the decision-maker to view this initial information as ambiguous. The distribution of the signal s_t conditional on a parameter value θ is described by a *set* of likelihoods $\mathcal{L}$. Every parameter value $\theta \in \Theta$ is thus associated with a set of probability measures $\mathcal{L}(\cdot \mid \theta)$. The size of this set reflects the decision maker's (lack of) confidence in what an ambiguous signal means, given that the parameter is equal to θ. Signals are unambiguous only if there is a single likelihood, that is $\mathcal{L} = \{\ell\}$. Otherwise, the decision-maker feels unsure about how parameters are reflected in data. The set of normal likelihoods described in (1) are a tractable example of this that will important below.

Beliefs about every signal in the sequence $\{s_t\}$ are described by the same set $\mathcal{L}$. Moreover, for a given parameter value $\theta \in \Theta$, the signals are known to be independent over time. However, the decision-maker is not confident that the data are actually identically distributed over time. In contrast, he believes that any sequence of likelihoods $\ell^t = (\ell_1, .., \ell_t) \in \mathcal{L}^t$ could have generated a given sample s^t and any likelihood in $\mathcal{L}$ might underly the next observation. The set $\mathcal{L}$ represents factors that the agent perceives as being relevant but which he understands only poorly - they can vary across time in a way that he does not understand beyond the limitation imposed by $\mathcal{L}$. Accordingly, he has decided that he will not try to (or is not able to) learn about these factors. In contrast, because θ is fixed over time, he can try to learn the true θ.

[2] In the infinite horizon case, uniqueness obtains only if $\mathcal{P}$ is assumed also to be *regular* in a sense defined in Epstein and Schneider [18], generalizing to sets of priors the standard notion of regularity for a single prior.

Conditional independence implies that the sample s^t affects beliefs about future signals (such as s_{t+1}) only to the extent that they affect beliefs about the parameter. We can therefore construct beliefs $\{\mathcal{P}_t(s^t)\}$ in two steps. First, we define a set of posterior beliefs over the parameter. For any history s^t, prior $\mu_0 \in \mathcal{M}_0$ and sequence of likelihoods $\ell^t \in \mathcal{L}^t$, let $\mu_t\left(\cdot\,; s^t.\mu_0, \ell^t\right)$ denote the posterior obtained by updating μ_0 by Bayes rule if the sequence of likelihoods is known to be ℓ^t. Updating can be described recursively by

$$d\mu_t\left(\cdot\,; s^t, \mu_0, \ell^t\right) = \frac{\ell_t(s_t \mid \cdot)}{\int_\Theta \ell_t(s_t \mid \theta')\, d\mu_{t-1}(\theta'; s^{t-1}, \mu_0, \ell^{t-1})} d\mu_{t-1}(\cdot\,; s^t, \mu_0, \ell^{t-1}).$$

The set of posteriors $\mathcal{M}_t\left(s^t\right)$ now contains all posteriors that can be derived by varying over all μ_0 and ℓ^t:

$$\mathcal{M}_t(s^t) = \left\{\mu_t\left(s^t; \mu_0, \ell^t\right) :\ \mu_0 \in \mathcal{M}_0,\ \ell^t \in \mathcal{L}^t\right\}. \tag{4}$$

Second, we obtain one-step-ahead beliefs by integrating out the parameter. This is analogous to the Bayesian case. Indeed, if there were a single posterior μ_t and likelihood ℓ, the one-step-ahead belief after history s^t would be

$$p_t\left(\cdot | s^t\right) = \int_\Theta \ell(\cdot \mid \theta)\, d\mu_t(\theta | s^t).$$

With multiple posteriors and likelihoods, we define

$$\begin{aligned}
\mathcal{P}_t(s^t) &= \left\{p_t\left(\cdot\right) = \int_\Theta \ell_{t+1}(\cdot \mid \theta)\, d\mu_t(\theta) : \mu_t \in \mathcal{M}_t(s^t),\ \ell_{t+1} \in \mathcal{L}\right\} \\
&= \int_\Theta \mathcal{L}(\cdot \mid \theta)\, d\mathcal{M}_t(\theta). \qquad (5)
\end{aligned}$$

This is the process of one-step-ahead beliefs that enters the specification of recursive multiple priors preferences (3).

The Bayesian model of learning from conditionally i.i.d. signals obtains as the special case of (5) when both the prior and likelihood sets have only a single element. For that model, the de Finetti theorem implies that one-step-ahead beliefs can be written equivalently as the conditionals of a single *exchangeable* probability P^* on the set of sequences S^∞. Similarly, when there is a single likelihood, that is, signals are *unambiguous,* then there is a set $\mathcal{P}^*$ of exchangeable measures on S^∞, such that $\mathcal{P}_t(s^t)$ equals the set of all one-step-ahead conditionals induced by measures in $\mathcal{P}^*$.

3 TREE PRICING

In this section, we derive two key properties of asset pricing with ambiguous news: market participants respond more strongly to bad news than to good news, and returns must compensate market participants for enduring periods of ambiguous news. We derive these properties first in a simple three period setting. In this context, we also compare the properties of information quality in our model to those of Bayesian models. We then move to an infinite horizon setting, where we derive a number of implications for observed moments.

3.1 An Asset Market with Ambiguous News

There are three dates, labelled 0, 1 and 2. We focus on news about one particular asset (asset A). There are $\frac{1}{n}$ shares of this asset outstanding, where each share is a claim to a dividend

$$d = m + \varepsilon^a + \varepsilon^i.$$

Here m is the mean dividend, ε^a is an aggregate shock and ε^i is an idiosyncratic shock that affects only asset A. In what follows, all shocks are mutually independent and normally distributed with mean zero. We summarize the payoff on all other assets by a dividend $\tilde{d} = \tilde{m} + \varepsilon^a + \tilde{\varepsilon}^i$, where $\tilde{m}$ is the mean dividend and $\tilde{\varepsilon}^i$ is a shock. There are $\frac{n-1}{n}$ shares outstanding of other assets and each pays $\tilde{d}$. The market portfolio is therefore a claim to $\frac{1}{n}d + \frac{n-1}{n}\tilde{d}$. In the special case $n = 1$, asset A is itself the market. For n large, it can be interpreted as stock in a single small company.

News

Dividends are revealed at date 2. The arrival of news about asset A at date 1 is represented by the signal

$$s = \alpha\varepsilon^a + \varepsilon^i + \varepsilon^s. \tag{6}$$

Here the number $\alpha \geq 0$ measures how specific the signal is to the particular asset on which we focus. For example, suppose n is large, and hence that d represents future dividends of a small company. If $\alpha = 1$, then the signal s is simply a noisy estimate of future cash flow d. As such, it partly reflects future aggregate economic conditions ε^a. In contrast, if $\alpha = 0$, then the news is 100% company-specific: while it helps to forecast company cash flow d, the signal is not useful for forecasting the payoff on other assets (that is, $\tilde{d}$). Examples of company-specific news include changes in management or merger announcements.

We assume that the signal is ambiguous: the variance of the shock ε^s is known only to lie in some range, $\sigma_s^2 \in [\underline{\sigma}_s^2, \overline{\sigma}_s^2]$. This captures the agent's lack of confidence in the signal's precision. This setup is very similar to the simple example (**??**) in the introduction. The one difference is that the parameter $\theta = (\varepsilon^a + \varepsilon^i, \varepsilon^a)'$ that agents try to infer from the signal s is now two-dimensional. Apart from that, there is again a single normal prior for θ and a set of normal likelihoods for s parametrized by σ_s^2. The set of one-step-ahead beliefs about s at date 0 consists of normals with mean zero and variance $\alpha^2\sigma_a^2 + \sigma_i^2 + \sigma_s^2$, for $\sigma_s^2 \in [\underline{\sigma}_s^2, \overline{\sigma}_s^2]$. The set of posteriors about θ at date 1 is calculated using standard rules for updating of normal random variables. For fixed σ_s^2, let γ denote the regression coefficient

$$\gamma\left(\sigma_s^2\right) = \frac{cov\left(s, \varepsilon^a + \varepsilon^i\right)}{var\left(s\right)} = \frac{\alpha\sigma_a^2 + \sigma_s^2}{\alpha^2\sigma_a^2 + \sigma_i^2 + \sigma_s^2} \in [0, 1]\,.$$

Given s, the posterior density of $\theta = (\varepsilon^a + \varepsilon^i, \varepsilon^a)'$ is normal with mean vector $\gamma(\sigma_s^2)\begin{pmatrix} 1 \\ \alpha \end{pmatrix} s$. The covariance matrix consists of $var(\varepsilon^a + \varepsilon^i|s) = (1 - \gamma(\sigma_s^2))(\sigma_a^2 + \sigma_i^2)$ and $var(\varepsilon^a|s) = cov(\varepsilon^a, \varepsilon^a + \varepsilon^i|s) = (1 - \alpha\gamma(\sigma_s^2))\sigma_a^2$. As σ_s^2 ranges over $[\underline{\sigma}_s^2, \overline{\sigma}_s^2]$, the coefficient $\gamma(\sigma_s^2)$ also varies, tracing out a family of posteriors. In other words, the ambiguous news s introduces ambiguity into beliefs about fundamentals.

Measuring Information Quality

To compare information quality across situations, it is common to measure the information content of a signal *relative* to the volatility of the parameter. For fixed σ_s^2, the coefficient $\gamma(\sigma_s^2)$ provides such a measure since it determines the *fraction* of prior variance in θ that is resolved by the signal. Under ambiguity, $\underline{\gamma} = \gamma(\bar{\sigma}_s^2)$ and $\bar{\gamma} = \gamma(\underline{\sigma}_s^2)$ provide lower and upper bounds on (relative) information content, respectively. In the Bayesian case, $\bar{\gamma} = \underline{\gamma}$, and agents know precisely how much information the signal contains. More generally, the greater is $\bar{\gamma} - \underline{\gamma}$, the less confident they feel about the true information content. This is the new dimension of information quality introduced by ambiguous signals. At the same time, $\underline{\gamma}$ continues to measure known information content - if $\underline{\gamma}$ increases, everybody knows that the signal has become more reliable.

In the present asset market example, the signal s captures the sum of all intangible information that market participants obtain during a particular trading period, such as a day. The range $\bar{\gamma} - \underline{\gamma}$ describes their confidence in that information. It may differ across markets or time due to differences in information production. For example, consider the case of a stock which suddenly becomes "hot", that is, popular news coverage increases. This often happens when a stock has done well in the past, for example. Increased popular coverage will typically not increase the potential for truly valuable news: $\bar{\gamma}$ remains nearly constant. However, given the new flood of information, the "typical day's news" s will be affected more by trumped up, irrelevant news items that cannot be easily distinguished from relevant ones: $\underline{\gamma}$ falls. As a second example, suppose a foreign stock is newly listed on the New York Stock Exchange. This will entice more U.S. analysts to research this particular stock, because trading costs for their American clients have now fallen. Again, the competence of the information providers is uncertain, especially since the stock is foreign. It again becomes harder to know how reliable is the typical day's news. However, since most of the new coverage is by experts, one would now expect $\bar{\gamma}$ to increase, while $\underline{\gamma}$ remains nearly constant.

3.2 Asymmetric Reponse and Price Discount

We assume that there is a representative agent who does not discount the future and cares only about consumption at date 2. He has recursive multiple-priors utility with beliefs as described above. We begin with a Bayesian benchmark, where the agent maximizes expected utility and beliefs are as above with $\underline{\gamma} = \bar{\gamma}$. We also allow for risk aversion: let period utility be given by $u(c) = -e^{-\rho c}$, where ρ is the coefficient of absolute risk aversion.

Bayesian Benchmark

It is straightforward to calculate the price of asset A at dates 0 and 1:

$$\begin{aligned} q_0 &= m - \rho cov\left(d, \frac{1}{n}d + \frac{n-1}{n}\tilde{d}\right) \\ &= m - \rho\left(\sigma_a^2 + \frac{1}{n}\sigma_i^2\right); \\ q_1(s) &= m + \underline{\gamma}s - \rho\left(\left(1 - \alpha\underline{\gamma}\right)\sigma_a^2 + \frac{1}{n}\left(1 - \underline{\gamma}\right)\sigma_i^2\right). \end{aligned} \tag{7}$$

Price equals the expected present value minus a risk premium that depends on risk aversion and covariance with the market. The latter consists of two parts, the variance of the common shock ε^a, and the variance of the idiosyncratic shock multiplied by $\frac{1}{n}$, the market share of the asset. As n becomes large, idiosyncratic risk is diversified away and does not matter for prices. If the signal is of any value $\left(\underline{\gamma} > 0\right)$, the price at date 1 will react to it. In addition, the risk premium will be reduced as the signal resolves some uncertainty.

Ambiguous Signals

We now calculate prices when the signal is ambiguous. For simplicity, we assume that the agent is risk neutral.[3] Of course, he is still averse to uncertainty, since he is averse to ambiguity. As discussed in Section 2, with recursive multiple-priors utility, actions are evaluated under the worst-case conditional probability. We also know that the representative agent must hold all assets in equilibrium. It follows that the worst-case conditional probability minimizes conditional mean dividends. Therefore, the price of asset A at date 1 is

$$q_1(s) = \min_{\sigma_s^2 \in \left[\underline{\sigma}_s^2, \bar{\sigma}_s^2\right]} E\left[d|s\right] = \begin{cases} m + \underline{\gamma}s & \text{if } s \geq 0 \\ m + \bar{\gamma}s & \text{if } s < 0. \end{cases} \tag{8}$$

A crucial property of ambiguous news is that the worst-case likelihood used to interpret a signal depends on the value of the signal itself. Here the agent interprets bad news ($s < 0$) as very informative, whereas good news are viewed as imprecise. The price function $q_1(s)$ is thus a straight line with a kink at zero, the cutoff point that determines what "bad news" means. If the agent is not ambiguity averse $\left(\bar{\gamma} = \underline{\gamma}\right)$, the price function is the same as that for a Bayesian agent who is not risk averse $(\rho = 0)$.

At date 0, the agent knows that an ambiguous signal will arrive at date 1. His one-step-ahead conditional beliefs about the signal s are normal with mean zero and variance $\alpha^2\sigma_a^2 + \sigma_i^2 + \sigma_s^2$, where σ_s^2 is unknown. Again, the worst-case probability is used to evaluate portfolios. Since the date 1 price is concave in the signal s, the date zero conditional

[3]This approach allows us to derive transparent closed form solutions for key moments of prices and returns. In the numerical example considered below, risk aversion is again introduced.

mean return is minimized by selecting the highest possible variance $\bar{\sigma}_{\varepsilon}^{2}$. We thus have

$$
\begin{aligned}
q_0 &= \min_{\sigma_s^2 \in \left[\underline{\sigma}_s^2, \overline{\sigma}_s^2\right]} E\left[q_1\right] \\
&= \min_{\sigma_s^2 \in \left[\underline{\sigma}_s^2, \overline{\sigma}_s^2\right]} E\left[m_0 + \underline{\gamma} s + \left(\bar{\gamma} - \underline{\gamma}\right) \min\left\{s, 0\right\}\right] \\
&= m - \left(\bar{\gamma} - \underline{\gamma}\right) \frac{1}{\sqrt{2\pi\underline{\gamma}}} \sqrt{\alpha\sigma_a^2 + \sigma_i^2} \qquad (9)
\end{aligned}
$$

The date zero price thus exhibits a discount, or ambiguity premium. This premium is directly related to the extent of ambiguity, as measured by $\bar{\gamma} - \underline{\gamma}$. It is also increasing in the volatility of fundamentals, including the volatility σ_i^2 of idiosyncratic risk. Again without ambiguity aversion, we obtain risk neutral pricing ($q_0 = m$), exactly as in the case of no risk aversion ($\rho = 0$ in 7)

Comparison of (9) and (7) reveals two key differences between risk premia and premia induced by ambiguous news. The first is the role of *idiosyncratic shocks* for the price of small assets. Ambiguous company-specific news not only induces a premium, but the size of this premium depends on total (including idiosyncratic) risk. In the Bayesian case, whether company-specific news is of low quality barely matters even ex post. Indeed, for $\sigma_a^2 = 0$ and n large, the effect of (in . Second, under ambiguity, prices depend on the *prospect* of low information quality. It is intuitive that if it becomes known today that information about asset A will be more difficult to intepret *in the future*, this makes asset A less attractive, and hence cheaper, already today. This is exactly what happens when the signal is ambiguous. In contrast, a change of information quality in the Bayesian model does not have this effect. While the prospect of lower information quality in the future produces a larger discount ex post after the news has arrived (q_1 is increasing in $\underline{\gamma}$), the ex ante price q_0 is independent of γ.

Both properties can be traced to one behavioral feature: for ambiguity averse investors, uncertainty about the distribution of future payoffs is a first-order concern. We have discussed above that the Bayesian model fails to predict behavior in the Ellsberg experiment, because it assumes that agents treat all model uncertainty as risk. The multiple-priors model accomodates Ellsberg-type behavior because agents act as if they adjust the mean of the uncertain assets (or bets). The same effect is at work here. To elaborate, consider first the impact of idisoyncratic shocks. If uncertainty about mean earnings changes because of company-specific news, then Bayesians treat this as a change in risk. There will be only a second order effect on Bayesian valuation of a company as long as the covariance with the market remains the same. In contrast, ambiguity averse investors act as if mean earnings themselves have changed. This is a first-order effect, even if the company is small.

Second, suppose that Bayesian market participants are told at date 0 that hard-to-interpret news will arrive at date 1. They believe that, at date 1, everybody will simply form subjective probabilities about the meaning of the signal at date 1 and average different scenarios to arrive at a forecast for dividends. As long as the volatility of fundamentals does not change, total risk is the same and there is no need for prices to change. In contrast, ambiguity averse market participants know that they will not be confident enough to assign subjective probabilities to different interpretations of the signal at date 1. Instead they will demand a discount once they have seen the signal. As a result, prices reflect this discount even at date 0. The prospect of ambiguous news is thus enough to cause a drop in prices.

3.3 Asset Price Properties

To compare the predictions of the model to data, we embed the above three-period model of news release into an infinite-horizon asset pricing model. Specifically, we chain together a sequence of short learning episodes of the sort modeled above. Agents observe just *one* intangible signal about the next innovation in dividends before that innovation is revealed and the next learning episode starts.

Assume that there is an exogenous riskless interest rate r and that the agent's discount factor is $\beta = \frac{1}{1+r}$. In addition, we omit the distinction between systematic and idiosyncratic shocks, since agents' reaction to ambiguous signals is similar in the two cases. The level of dividends on some asset is given by a deterministic trend plus a mean-reverting process,

$$d_t = \kappa \bar{d} + (1-\kappa)\, d_{t-1} + u_t, \tag{10}$$

where u_t is a shock and $\kappa \in (0,1)$. The parameter κ measures the speed with which dividends adjust back to their mean $\bar{d}$.[4]

Every period, agents observe an ambiguous signal about next period's shock:

$$s_t = u_{t+1} + \varepsilon_t^s,$$

where the variance of ε_t^s is $\sigma_{s,t}^2 \in [\underline{\sigma}_s^2, \overline{\sigma}_s^2]$. The relevant state of the world for the agent is (s_t, d_t). The components s_t and d_t are conditionally independent, because s_t provides information only about u_{t+1}, which in turn is independent of d_t. Beliefs about s_{t+1} are normal with mean zero and (unknown) variance $\sigma_u^2 + \sigma_{s,t}^2$. Beliefs about d_{t+1} are given by (10) and the set of posteriors about u_{t+1} given s_t described in the previous section.

Our goal is to derive asset pricing properties that would be observed by an econometrician who studies the above asset market. We thus assume that there is a *true* variance of noise $\sigma_s^{*2} \in [\underline{\sigma}_s^2, \overline{\sigma}_s^2]$. It is also useful to define $\gamma^* = \gamma(\sigma_s^{*2})$, a measure of the true information content of the news that arrives in a typical trading period. In addition, we assume that the true distribution of the fundamentals u coincides with the subjective

[4] Under these assumptions, dividends are stationary in levels, which is not realistic. However, it is straightforward to extend the model to allow for growth. Let observed dividends be given by

$$\begin{aligned} \hat{d}_t &= g^t(\bar{d} + d_t) \\ d_t &= (1-\kappa)\, d_{t-1} + u_t, \end{aligned} \tag{11}$$

where $g-1$ is the average growth rate, $g-1 < r$. The observed stock price in the growing economy is then $\widehat{q}_t = g^t q_t$. The analysis below applies to the detrended stock price q_t if β is replaced by βg.

beliefs of agents. The latter assumption distinguishes the present model from existing approaches to asset pricing under ambiguity. Indeed, existing models are driven by ambiguity about fundamentals. The degree of ambiguity is then often motivated by how hard it is to measure fundamentals. The present setup illustrates that ambiguity can matter even if the true process of dividends is *known* by both the econometrician and market participants. The point is that market participants typically have access to ambiguous information other than past dividends that is not observed by the econometrician.

Let q_t denote the stock price. In equilibrium, the price at t must be the worst-case conditional expectation of the price plus dividend in period $t+1$:

$$q_t = \min_{\left(\sigma_{s\,t}^2, \sigma_{s,t+1}^2\right) \in \left[\underline{\sigma}_s^2, \overline{\sigma}_s^2\right]^2} \beta E_t \left[q_{t+1} + d_{t+1}\right]. \tag{12}$$

We focus on stationary equilibria. The price is given by

$$q_t = \frac{\bar{d}}{r} + \frac{1-\kappa}{r+\kappa} d_t + \frac{1}{r+\kappa} \gamma_t s_t - \left(\bar{\gamma} - \underline{\gamma}\right) \frac{\sigma_u}{r\sqrt{2\pi\underline{\gamma}}}, \tag{13}$$

where γ_t is a random variable that is equal to $\bar{\gamma}$ if $s_t < 0$ and equal to $\underline{\gamma}$ otherwise.[5]

The first two terms reflect the present discounted value of dividends – without news, the model reduces to a version of the Gordon growth model, where prices are determined by the interest rate, the growth rate and the current dividend level. The third term captures the response to the current ambiguous signal. As in (8), this response is asymmetric: the distribution of γ_t implies that bad news are incorporated into prices more strongly. In addition, the strength of the reaction now depends on the persistence of dividends: if κ is smaller, then the effect of news on prices is stronger since the information matters more for payoffs beyond just the next period. The fourth term captures anticipation fear of *future* ambiguous news – it is the present discounted value of the discount in (9). As before, it may contain compensation for asset-specific shocks.

We are interested also in the behavior of excess returns. In the present setup, it is convenient to focus on *per share* excess returns, defined as

$$\begin{aligned} R_{t+1} &= q_{t+1} + d_{t+1} - (1+r)\, q_t \\ &= \frac{1+r}{r+\kappa} \left(\beta \gamma_{t+1} s_{t+1} + u_{t+1} - \gamma_t s_t\right) + \left(\bar{\gamma} - \underline{\gamma}\right) \frac{\sigma_u}{\sqrt{2\pi\underline{\gamma}}} \end{aligned} \tag{14}$$

If there is no intangible information, we have risk neutral pricing. In this case, all γs are zero and excess returns are $\frac{1+r}{r+\kappa} u_{t+1}$ – they depend only on the shock to fundamentals and are always expected to be zero. The gradual diffusion of information through news implies that returns will also depend on the signals, and hence on noise.

Equity Premium and Idiosyncratic Risk

[5] Conjecture a time invariant price function of the type $q_t = \bar{Q} + Q_{\hat{d}} \hat{d}_t + Q_s \gamma_t s_t$. Inserting the guess into (12) and matching undetermined coefficients delivers (13).

The mean excess return under the true probability is

$$E^*[R_{t+1}] = (\bar{\gamma} - \underline{\gamma}) \frac{1}{\sqrt{2\pi\underline{\gamma}}} \left(1 + \frac{r}{r+\kappa}\sqrt{\underline{\gamma}/\gamma^*}\right)\sigma_u.$$

The presence of ambiguous news induces an ambiguity premium. It is well-known that such a premium can arise as a result of ambiguity in fundamentals. What is new here is that it is driven by ambiguity about the quality of news. Since news is in turn driven by fundamentals, this introduces a direct link between the *volatility* of fundamentals and the ambiguity premium: $E^*[R_{t+1}]$ is increasing in σ_u. Ambiguous signals thus provide a reason why idiosyncratic risk is priced. Indeed, as discussed above, it does not matter for the order of magnitude of the effect whether the news is company-specific or not.

Recent literature has documented a link between idiosyncratic risk and excess returns in the cross section (for example, Lehmann [28] and Malkiel and Xu [29]). The explanation typically put forward is that agents do not fully diversify. There is indeed evidence that some agents, such as individual households, hold undiversified portfolios.[6] However, a Bayesian model will produce equilibrium premia for idiosyncratic risk only if there is *no* agent who holds a well-diversified portfolio. Any well-diversified mutual fund, for example, will bid up prices until the discount on idiosyncratic risk is zero. In contrast, the present model features one, well-diversified, investor. As long as this investor views company-specific news as ambiguous, he will want to be compensated for it. Since institutional investors with low transaction costs still have to process news, this delivers a more robust story for why idiosyncratic risk can matter.

Additional implications of the ambiguous news model could be used to explore further the tradeoff between idiosyncratic shocks and expected returns. In particular, the premium changes to different degrees depending on the way in which information quality is increased. Other things equal, the premium is increasing in both $\bar{\gamma}$ and $\underline{\gamma}$, but the derivative with respect to $\bar{\gamma}$ is always larger. This implies that, at the margin, an increase in coverage by potential experts (higher $\bar{\gamma}$) induces a larger increase in the premium than an increase in popular news coverage $(\text{lower } \underline{\gamma})$. The intuition is that potential high quality news moves prices more, and hence induces more uncertainty per unit of volatility of fundamentals σ_u.

Excess Volatility

A classic question in finance is why stock prices are so much more volatile than measures of the expected present value of dividends. We now reconsider the link between "excess volatility" and information quality. The variance of the stock price is

$$var(q_t) = \sigma_u^2 \left(\frac{1-\kappa}{r+\kappa}\right)^2 \left(\frac{1}{\kappa(2-\kappa)} + \frac{1}{2\gamma^*}\left(\bar{\gamma}^2 + \underline{\gamma}^2 - \frac{1}{\pi}(\bar{\gamma} - \underline{\gamma})^2\right)\right).$$

Price volatility is proportional to the volatility of the shock u. When there is no intangible information $(\underline{\gamma} = \bar{\gamma} = 0)$, the second term in the big bracket is zero, and the volatility of prices is equal to that of the present value of dividends .

[6] In fact, ambiguity aversion may be responsible for such underdiversification. See Epstein and Schneider [19] for an analysis of portfolio choice with recursive multiple-priors.

The big bracket reflects the propagation of shocks through news. In the Bayesian case ($\underline{\gamma} = \bar{\gamma} = \gamma^*$), it is equal to $\frac{1}{\kappa(2-\kappa)} + \gamma^*$: given persistence κ, price volatility is increasing in information quality and bounded above. In particular, with persistent dividends (κ small), changes in information quality will typically have only a small effect on price volatility – it is dominated by the volatility of the present value of dividends. With ambiguous signals, at the benchmark $\bar{\gamma} = \underline{\gamma} = \gamma^*$, $var\,(q_t)$ is increasing in both $\bar{\gamma}$ and $\underline{\gamma}$. In other words, more news coverage by potential experts (higher $\bar{\gamma}$) increases volatility, while more popular coverage (lower $\underline{\gamma}$) tends to reduce it. Also, the possibility of ambiguous news removes the upper bound on price volatility imposed by the Bayesian model.[7] Ambiguous signals can thus contribute to excess volatility of prices.

Recent empirical work on firm level volatility has mostly focused on volatility of *returns*. In our model,

$$var\,(R_t) = \sigma_u^2 \left(\frac{1+r}{r+\kappa}\right)^2 \left\{1 - \bar{\gamma} - \underline{\gamma} + \frac{1+\beta^2}{2\gamma^*}\left(\bar{\gamma}^2 + \underline{\gamma}^2 - \frac{1}{\pi}\left(\bar{\gamma} - \underline{\gamma}\right)^2\right)\right\}$$

At the point $\bar{\gamma} = \underline{\gamma} = \gamma^*$, the derivatives of $var(R_t)$ with respect to both $\bar{\gamma}$ and $\underline{\gamma}$ are again positive. Changes in information quality due to changes in the ambiguity of signals thus affect the volatility of prices and returns in the same way. This is in sharp contrast to the Bayesian case, where price and return volatilities move in opposite directions.[8] As explained by West [40], the latter result obtains because higher information quality simply means that more information about future cash flows is released earlier, when the cash flows are still being discounted at a higher rate.

Campbell et al. [12] have documented an upward trend in individual stock return volatility. One question is whether this development can be connected to improvements in information technology. Campbell et al. interpret such improvements as an increase in γ^*, and dismiss the explanation, because return volatility should decrease, rather than increase. In our framework, this argument only applies if the effect of the new technology is immediately fully known. If agents have also become less certain about how much improvement there is, the outcome is no longer obvious. In particular, while higher γ^* lowers return volatility, higher $\bar{\gamma}$ increases it. Increased uncertainty about the potential of information technology is thus consistent with higher volatility of returns.

Negative Skewness

Since ambiguity averse market participants respond asymmetrically to news, the model tends to produce skewed distributions for prices and returns, even though both dividends

[7] This does not mean that that price volatility in the ambiguous news model is arbitrary. It just says that the model cannot be rejected using relative volatility of prices and dividends. In a quantitative application, $\bar{\gamma}$, $\underline{\gamma}$ and γ^* can be identified from the full distribution of prices. See Section 4 for an example.

[8] Formally, in the Bayesian case,

$$var\,(R_t) = \frac{\sigma_u^2}{(1-\beta\kappa)^2}\left(1 - \gamma^*(1-\beta^2)\right),$$

so that an increase in γ^* makes price more volatile but returns less so.

and noise have symmetric (normal) distributions. Skewness of a random variable x is usually defined by $r = \mu_3(x) / (\sigma(x))^3$, where μ_3 is the centered third moment and σ is the standard deviation. For returns, the third moment is

$$\mu_3(R_{t+1}) = \left(\frac{1+r}{r+\kappa}\right)^3 \left\{ -\left(1-\beta^3\right)\mu_3(\gamma_t s_t) - \frac{3\sigma_u^3}{\sqrt{2\pi\gamma^{*3}}}\left(1-\frac{1}{\pi}\right)\left(\bar{\gamma}-\underline{\gamma}\right)\left(\bar{\gamma}+\underline{\gamma}-\gamma^*\right) \right\}.$$

The appendix shows that $\mu_3(\gamma_t s_t)$ is negative and proportional to σ_u^3, where the proportionality factor depends on $\bar{\gamma}$, $\underline{\gamma}$ and γ^*. Since $\gamma^* \leq \bar{\gamma}$, the second term is negative, so that returns are negatively skewed, provided that the discount factor β is high enough. Moreover, since the standard deviation of returns is proportional to σ_u, skewness is independent of the volatility of fundamentals and depends only on relative information quality. In the Bayesian case, skewness is zero.

The intuition follows from the definition of returns and agents' asymmetric response, captured by the distribution of γ_t. At date $t+1$, the realized return R_{t+1} is affected by two pieces of news. First, the signal s_{t+1} will be weighted more heavily the less favorable it is, which tends to make returns negatively skewed. Second, the dividend shock u_{t+1} will often offset strong negative reponses to signals in the previous period. As a result, the "dividend surprise" $u_{t+1} - \gamma_t s_t$ will be positively skewed. Since the signal s_{t+1} is about future cash flows (beyond $t+1$), the first effect becomes more important as the discount factor increases. Indeed, for $\beta \to 1$ we will always obtain negative skewness. The dividend surprise is less positively skewed than the signal itself is negatively skewed, because the signal does provide some information. We conclude that, at least at high frequencies, the model predicts negative skewness in returns.

Negative skewness of stock returns at high frequencies has been widely documented in the empirical literature. It is closely connected to the finding that volatility increases in times of low returns, which has been found in both aggregate (index) and individual stock returns.[9] In the cross section, conditional negative skewness tends to be more pronounced in stocks that have (i) witnessed a runup in prices in the last six months and (ii) low book-to-market value (Harvey and Siddique [26], Chen et al. [15]). This evidence is consistent with our comparative statics analysis of information quality. Stocks that have recently risen in price and glamour stocks with low book-to-market value all tend to attract a lot of popular media attention. As traders grapple with intangible information – news of uncertain quality – they induce negative skewness in returns.

The above cited papers also show that negative skewness tends to be more pronounced for stocks with larger market capitalization. A potential explanation is again that large cap stocks are "in the news" more, so that traders in those stocks have to digest more ambiguous, intangible information. It is important for this point that our model does *not* predict that skewness and volatility are related in the cross section: large stocks can be more negatively skewed, but still have lower volatility and hence ambiguity premia, since their fundamentals are more stable (lower σ_u^2). It is also important that skewness is driven by the ambiguity of signals, and not by ambiguous *prior* information. One might

[9] See, for example, Bekaert and Wu [5] for references.

expect that investors perceive small stocks to be more ambiguous ex ante. Incorporating prior ambiguity about the mean of u_t, say, would increase the ambiguity premium and could thus help explain the size effect. However, it would not change the centered third moment and skewness. The latter would still only depend on the perception of news.

4 SHOCKS TO INFORMATION QUALITY

The previous section compared markets with different information quality. We now focus on changes to information quality in a given market. In particular, asset markets often witness shocks that not only increase uncertainty about fundamentals, but also force markets to deal with unfamiliar news sources. Our leading example for such a shock is September 11, 2001. On the one hand, the terrorist attack increased uncertainty about economic growth. On the other hand, news about terrorism and foreign policy – that were arguably less important for U.S. growth and stock returns earlier – suddenly became important for market participants to follow. It thus became more difficult for market participants to assess how much weight to place on any given piece of news. For familiar news, such as Fed announcements and macro statistics releases, market participants had – through years of experience – developed a feel for the relevance of any given piece of news. Such experience was lacking for the news that became relevant after 9/11.

We view 9/11 as the beginning of a learning process where market participants were trying to infer the possibility of structural change to the U.S. economy from unfamiliar signals. Figure 2 plots the price-dividend ratio for the S&P 500 index for 19 trading days, starting 9/17, including the pre-attack value, 9/10, as day 0. The stock market was closed in the week after the attack; trading resumed on Monday, 9/ 17. The large drop on that day was followed by another week of losses, before the market began to rebound. At the end of our window – Friday, October 5 – the market had climbed, for the first time, to the pre-attack level. It subsequently remained between 68 and 73 for another three weeks (not shown). With hindsight, we know that structural change did not occur. The goal is thus to explain what moved the market during the learning process, conditional on that knowledge. We argue that the key implications of ambiguous news, asymmetric response and price dicsount, help to explain the observed price pattern. The exercise also illustrates how belief parameters can be identified from the distribution of prices.

Setup

There is an infinitely-lived representative agent. A single Lucas tree yields dividends $Y_t = \exp\left(\sum_{j=1}^{t} \Delta y_j\right) Y_0$, with Y_0 given. According to the true data generating process, the growth rate of dividends is $\Delta y_t \sim i.i.\mathcal{N}\left(\theta^{hi}, \sigma^2\right)$ for all t. The agent knows that the mean growth rate is θ^{hi} from time 0 up to some given time $T+1$. However, he believes that with probability $1-\mu$, the mean growth rate drops permanently to θ^{lo} after $T+1$. Information about growth beyond $T+1$ is provided, at each date $t \leq T$, by a signal s_t that takes the values 1 or 0. Signals are serially independent and also independent of dividends before $T+1$; they satisfy $\Pr(s_t = 1) = \pi$. At time $T+1$, the long run mean growth rate is revealed.

The information structure captures the following scenario. First, there was no actual permanent structural change caused by the attack.[10] Second, agents were initially unsure if there would be such a change. Third, news reports were initially much more informative about the possibility of structural change than were dividend or consumption data. Of course, to the extent that dividend data were available, they may have provided some information. But initially, they are likely to have largely reflected decisions taken before the attack occurred, becoming more informative only with time. Our model captures this shift in relative informativeness in a stark way. We divide the time after the attack into two phases, a learning phase ($t \leq T$) where dividends are entirely uninformative about structural change, and a "new steady state" phase ($t > T$) where structural change actually materializes in dividends. In our calibration below, T corresponds to 26 days.[11] Finally, imposing a fixed T at which θ is revealed is not a strong restriction if beliefs are already close to the true θ at time T. We show a plot of our posterior means below.

The agent believes that signals are informative about growth, but views them as ambiguous. This feature is modeled via a set of likelihoods ℓ, where

$$\ell\left(s_t = 1|\theta^{hi}\right) = \ell\left(s_t = 0|\theta^{lo}\right) = \lambda \in [\underline{\lambda}, \overline{\lambda}], \tag{15}$$

with $\underline{\lambda} > \frac{1}{2}$. Beliefs about signals up to time T are represented by the parameter space $\Theta = \left\{\theta^{hi}, \theta^{lo}\right\}$, the single prior given by μ and the set of likelihoods $\mathcal{L}$ defined by (15). The special case $\underline{\lambda} = \overline{\lambda}$ is a Bayesian model. To ease notation, assume that signals continue to arrive after T, but that for $t > T$, $\ell\left(s_t = 1|\theta^{hi}\right) = 1 = \ell\left(s_t = 0|\theta^{lo}\right)$.

In terms of the notation of Section 2, the state space is $S = \{0,1\} \times \mathbb{R}$. Since Y is independent of s, the one-step-ahead beliefs $\mathcal{P}_t\left(s^t, Y^t\right)$ for $t \leq T$ are given by the appropriate product of one-step-ahead beliefs about s_{t+1} and the conditional probability law for Y_{t+1}. Preferences over consumption streams are then defined recursively by

$$V_t\left(c; s^t, Y^t\right) = \min_{p_t \in \mathcal{P}_t\left(s^t, Y^t\right)} \left(c_t^{1-\gamma} + \beta E^{p_t}\left[\left(V_{t+1}\left(c; s^{t+1}, Y^{t+1}\right)\right)^{1-\gamma}\right]\right)^{\frac{1}{1-\gamma}}, \tag{16}$$

where β and γ are the discount factor and the coefficient of relative risk aversion, respectively. Since only the signals are ambiguous, the minimization in (16) may be viewed as a choice over sequences $\lambda^{t+1} = (\lambda_1, ..., \lambda_{t+1})$ of precisions.

Connection between Truth and Beliefs

Discipline on beliefs is imposed in two ways. First, as above, assume that the true precision λ lies in $[\underline{\lambda}, \overline{\lambda}]$. This condition ensures that an agent's view of the world is not

[10] The model can nevertheless accommodate drops in dividends in September and stock price movements that reflect these drops. All that is required is that such movements come from the same distribution as movements before September 11.

[11] The model could be extended to relax this strict division into phases. One might want to assume that both news reports and dividends are informative about structural change at all times. However, in such a setup, one would still like to let the informativeness of news reports decrease over time relative to that of dividends. It is plausible that the main effects of our setup would carry over to this more general environment.

contradicted by the data. Suppose the agent looks back at the history of signals after he is told the true parameter at time T. If he is Bayesian ($\underline{\lambda} = \overline{\lambda}$), the distribution of the signals at the true parameter value is the same as the true distribution of the signals. In this sense, the agent has interpreted the signals correctly.[12] More generally, an ambiguity averse agent contemplates many 'theories' of how the signal history has been generated, each corresponding to a different sequence of precisions. One might thus be concerned that theories that do not satisfy $\lambda_t = \lambda$ infintely often are contradicted by the data. However, this is not the case if $\lambda \in [\underline{\lambda}, \overline{\lambda}]$: there exists a large family of signal processes with *time varying precision* $\lambda_t \in [\underline{\lambda}, \overline{\lambda}]$ that cannot be distinguished from the true distribution on the basis of *any* finite sample.[13] While some of these processes will appear less likely than others in the short run, any of them is compatible with a sample that looks i.i.d. with precision λ. An agent who believes in the whole range $[\underline{\lambda}, \overline{\lambda}]$ need not, with hindsight, feel that he interpreted the signals incorrectly.

The second way in which discipline is imposed on beliefs is through the restriction that agents would learn the true state θ^{hi} even if it were not revealed at $T+1$. This precludes an excessively pessimistic interpretation of news. A sufficient condition is that the posterior probability of θ^{hi}, $\mu_t\left(s^t, \lambda^t\right)$, converges to 1 if the truth equals the *lower bound* of the precision range:

$$\lim_{t\to\infty} \min_{\lambda_1^t} \ \mu_t\left(s^t, \lambda^t\right) = 1, \ \text{ a.s. for } s_t \text{ i.i.d. with } \Pr\left(s_t = 1\right) = \underline{\lambda}. \tag{17}$$

If $\overline{\lambda}$ were too large for given $\underline{\lambda}$, agents could interpret negative signals as very precise and never be convinced that the true state has occurred if the fraction of good signals is $\underline{\lambda}$. Thus the condition bounds $\overline{\lambda}$ for a given $\underline{\lambda}$.

Supporting Measure and Asset Prices

Following [20], equilibrium asset prices can be read off standard Euler equations once a (one-step-ahead) "supporting measure" that achieves the minimum in (16) has been determined. Suppose that the intertemporal elasticity of substitution is greater than one. It is then easy to show that continuation utility is always higher after good news ($s=1$) than after bad news ($s=0$). Thus the sequence of precisions $\left(\lambda^{*t+1}\right)$ that determines the supporting measure at time t and history s^t is chosen to *minimize* the probability of a high signal in $t+1$. For the past signals s^t, this requires maximizing the precision of bad news ($\lambda_j^* = \overline{\lambda}$ if $s_j = 0$) and minimizing the precision of good news ($\lambda_j^* = \underline{\lambda}$ if $s_j = 1$). For the future signal s_{t+1}, it requires maximizing (minimizing) the precision λ_{t+1}^* whenever news are more likely to be bad (good) next period, that is, whenever the posterior probability of θ^{hi} is smaller (larger) than $\frac{1}{2}$.

Let q_t denote the price of the Lucas tree. Since signals and dividends are independent

[12]For example, he has not been "overconfident", interpreting every signal as more precise than it actually was.

[13]To construct such precision sequences, pick any ω such that $\omega\underline{\lambda} + (1-\omega)\overline{\lambda} = \lambda$. Let $\tilde{\lambda}_t$ be an i.i.d. process valued in $\left\{\underline{\lambda}, \overline{\lambda}\right\}$ with $\Pr\left(\tilde{\lambda}_t = \underline{\lambda}\right) = \omega$. For almost every realization (λ_t) of $\left(\tilde{\lambda}_t\right)$, the empirical distribution of the nonstationary signal process with precision sequence (λ_t) converges to the true distribution of the signals. See Nielsen [33] for a formal proof.

for $t \leq T$, the price-dividend ratio $v_t = q_t / Y_t$ depends only on the sequence of signals. It satisfies the difference equation

$$v_t\left(s^t\right) = \hat{\beta} E_t^* \left[\left(1 + v_{t+1}\left(s^{t+1}\right)\right)\right]$$

where E_t^* denotes expectation with respect to the (one-step-ahead) supporting measure and where the new discount factor $\hat{\beta} = \beta e^{(1-\gamma)\theta^{hi} - \frac{1}{2}(1-\gamma)^2\sigma^2}$ is adjusted for dividend risk, where γ here is the coefficient of relative risk aversion. Once θ has been revealed at date $T+1$, the price dividend ratio settles at a constant value.

Calibration

We set the discount rate to 4% p.a. and the coefficient of relative risk aversion to one half. The average growth rate of dividends is fixed to match the price-dividend ratio, yielding a number of 5.2% p.a. This is clearly larger than the historical average, which reflects the high p/d ratio. The volatility of consumption is set at the historical value of 2% p.a. reported by Campbell [11] for postwar data. Finally, we assume that the potential permanent shock corresponds to a drop in consumption growth of .5 % p.a. In steady state, this would correspond to a price-dividend ratio of 61.

Having fixed these parameters, we infer, for every learning model, the sequence of signals that must have generated our price-dividend ratio sample if the model is correct. If the signals had a continuous distribution, this map would be exact. Here we assume that agents observe 20 signals per day. We then compute the model-implied price path that best matches the data. While the price distribution is still discrete, it is sufficiently fine to produce sensible results. A model is discarded if its 'pricing errors' are larger than .5 at any point in time. Finally, we compute the likelihood conditional on the first observation for each model, using the distribution of the fitted price paths. This is a useful criterion for comparing models, since the first observation is basically explained by the choice of the prior.

Numerical Results

To select a Bayesian model, we search over priors μ and precision parameters λ to maximize the likelihood. This yields an interior solution for both parameters. For example, for precision, the intuition is as follows: the path of posteriors is completely determined by the path of p/d ratios. Thus performance differences across Bayesian models depend on how likely the path of posteriors is under the truth. If precision is very large, then it is highly unlikely that there could have been enough bad news to explain the initial price decrease. In contrast, if precision is very small, then signals are so noisy that posteriors do not move much in response to any given news. Highly unlikely 'clusters' of first bad and then good news would be required to explain the price path. This tradeoff gives rise to an interior solution for precision.

To select a multiple-priors model, we need to specify both the true precision and the range of precisions the agent thinks possible. To sharpen the contrast with Bayesian models, we focus on models where ambiguity is large; we set $\overline{\lambda}$ slightly (.001) below the upper bound associated with the requirement (17). We also assume that the truth

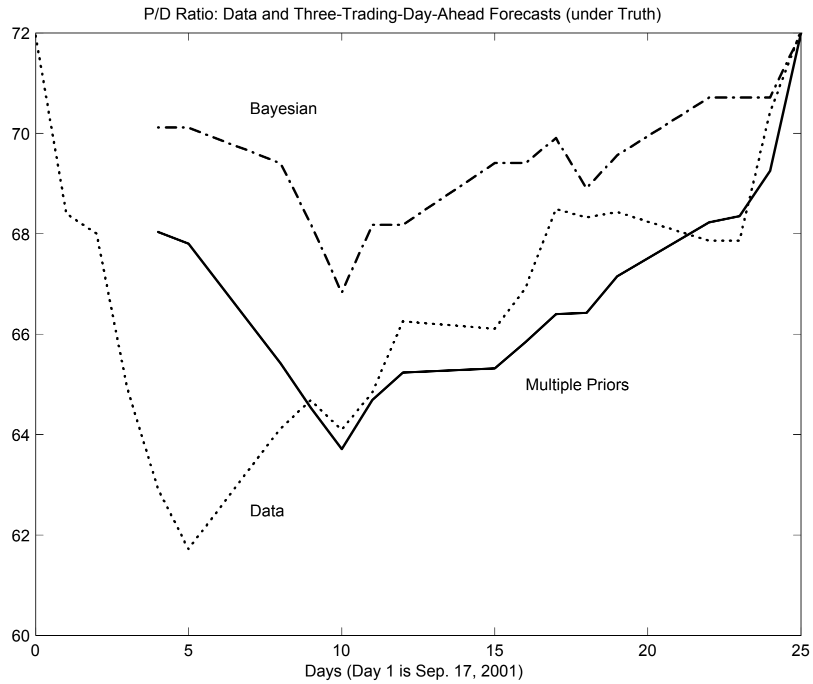

Figure 2: Data and In-Sample Forecasts for 9/11 Calibration.

corresponds to $\underline{\lambda}$.[14] With these two conventions, we search over $\underline{\lambda}$ to find our favorite multiple-priors model. This model is compared to the Bayesian model in Figures 2 and 3. The favorite multiple-priors model begins with a much higher prior probability, and the precision range for $\lambda \in [.58, .608]$ is higher than the precision for the best Bayesian model, $\lambda = .56$. The multiple-priors model (log likelihood $= -33.29$) outperforms the Bayesian model (log likelihood $= -36.82$). Figure 2 plots the one-step-ahead conditional likelihoods to illustrate the source of the difference. The multiple-priors model is better able to explain the downturn in the week of September 17. The models do about the same during the recovery. Figure 2 plots, together with the data, three-trading-day-ahead in-sample forecasts. This shows that the Bayesian model predicts a much faster recovery than the multiple-priors model throughout the sample.

The result shows how the effects discussed in the previous section operate in a setting with many signals. The two models represent two very different accounts of market movements in September 2001. According to the Bayesian story, all price movements reflect changes in beliefs about future growth. In particular, the initial drop in prices arose because market participants expected a permanent drop in consumption of .2% (see Figure 3). During the first week, bad news increased the expected drop to almost .5%. In contrast, the ambiguity story says that agents begin with a prior opinion that basically nothing has changed. However, they know that the next few weeks will be one

[14]Strictly speaking, this polar case is not permitted by the restriction that the truth lie in the interior of the precision range. However, there is always an admissible model arbitrarily close to the model we compute.

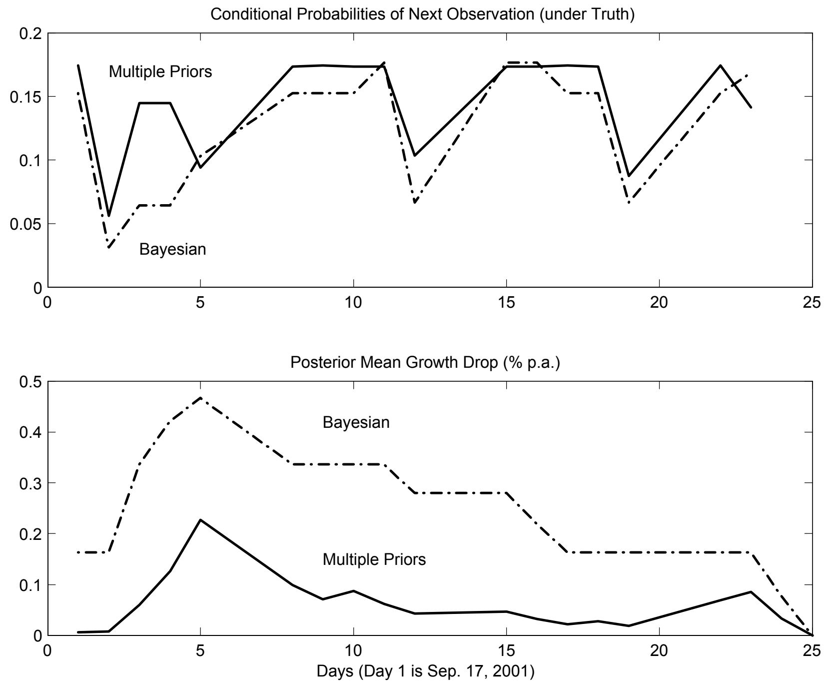

Figure 3: Conditional Probabilities and Posterior Mean for 9/11 Calibration.

of increased confusion and uncertainty. Anticipation of this lowers their willingness to pay for stocks. In particular, they know that future bad news will be interpreted (by future ambiguity averse market participants) as very precise, whereas future good news will be interpreted as noisy. This makes it more likely that the market will drop further in the short run than for the Bayesian model.

For representative agent asset pricing models with multiple-priors utility, there is always an observationally equivalent Bayesian model that yields the same equilibrium price. This begs the questions why one should not consider this Bayesian model directly.[15] Here, the reason is that this Bayesian model cannot be motivated by the same plausible *a priori* view of the environment as our ambiguity aversion model. We want to capture a scenario where signals are generated by a memoryless mechanism, and where precision does not depend on the state of the world: learning in good times is not expected to occur at a different speed than in bad times. An ambiguity aversion model with these features outperforms a Bayesian model with these features. Some other Bayesian model which does not have these features is not of interest. In addition, such a model would yield misleading comparative static predictions. The observationally equivalent model is much like a 'reduced form' which is not invariant to changes in the environment.

[15] This model would be an expected utility model with pessimistic beliefs, similar to the one in Abel [1].

5 RELATED LITERATURE

As mentioned above, our paper [17] provides axiomatic foundations for recursive multiple-priors utility and derives measure-by-measure Bayesian updating as a dynamically consistent updating rule for sets of priors. In Epstein and Schneider [19], we impose further structure on belief sets in order to capture learning from a sequence of conditionally independent signals. That paper introduces the general notion of "ambiguous signals"; formally, they are signals for which connection to the underlying parameter is described by a nonsingleton set of likelihoods. In the present paper, "ambiguous information" refers to ambiguous signals for which likelihoods differ only in precision, or information that is hard to interpret in the sense that its reliability is uncertain. The applications in the present paper also differ from that in [19]. The latter considers partial equilibrium portfolio choice behavior, while here our focus is on equilibrium pricing of securities.

Several papers have derived and discussed the emergence of ambiguity premia in general settings (for example, Epstein and Wang [20], Chen and Epstein [14]). What is new in the present paper is that the premium is due to ambiguity in signals. This additional structure (i) clarifies that the *anticipation* of ambiguous news is sufficient for premia and (ii) creates a direct link between the size of the premium and the *measured volatility* of fundamentals. In addition, we emphasize that the size of the ambiguity premium need not have anything to do with the asset's covariance with the market, as long as the latter is not zero. Uncertainty about asset-specific news cannot be dismissed as negligible, because ambiguity averse agents do not diversify it away. This result is related to the fact that the law of large numbers needs to be modified under ambiguity (see, for example, Marinacci [31]). With ambiguous signals, it naturally leads to compensation for measured idiosyncratic volatility.

The relevance of information not observed by the econometrician has long been recognized. Traditional Euler equation tests (for example, Hansen and Singleton [25]) are attractive precisely because they are robust to agents having more information than the econometrician. However, fully specified equilibrium asset pricing models typically assume that all information is tangible – information sets tend to include only past and present prices, consumption and dividends. An exception is Veronesi [39], who has examined the effect of information quality on the equity premium in a Lucas asset pricing model that also features an intangible (but unambiguous) signal. He shows that with high risk aversion there is no premium for low information quality in a Bayesian model.

Several recent papers have pointed to overconfidence as a source of overreaction to signals and excess volatility (see, for example, Daniel et al. [16], and also the references in Barberis and Thaler [4]). In these models the agent's perceived precision of an intangible (private) signal is higher than "true" precision. This makes reactions to signals more aggressive than under rational expectations. The mechanism that leads to excess volatility is thus similar to that in our model. However, overconfidence by itself does not entail an asymmetric response to news. Moreover, overconfidence is an assumption about the relationship between true and subjective precisions while ambiguity aversion is an assumption about (subjective) preferences only. The two concepts are thus complementary: one can consider a model in which agents are uncertain about precision, but the true precision lies close to (or even below) the lower bound of the range. This would describe overconfident, ambiguity averse agents.

Other recent work in behavioral finance is also motivated by psychological evidence and explores the effect of deviations from expected utility for asset pricing. Caplin and Leahy [13] have studied a model of utility over belief states to formalize anxiety as an anticipatory emotion. As in our model, information flow in the future matters for current utility, which is consistent with experiments on anxiety. In a series of papers, Barberis and Huang (for example, [2], [3]) have looked at the effects of narrow framing. They assume that agents derive utility directly from individual asset positions. This gives rise to premia on individual assets that are unrelated to correlation with the market. Our results show that a concern for both future information flow and asset-specific uncertainty also follow from the first-order concern with uncertainty exhibited by ambiguity averse agents.

A large literature on asset pricing with Bayesian learning argues that learning can explain excess volatility and in-sample predictability of returns. Excess volatility arises because agents' subjective variance of dividends is higher than the true variance. Agents with such diffuse priors react more strongly to news than agents with rational expectations. To keep priors diffuse, the learning process is usually reset periodically, for example because of regime shifts.[16] The model of Section 3.3 is related to this literature since it also chains together a sequence of short learning processes: in fact, agents observe just *one* intangible signal about the next innovation in dividends before that innovation is revealed and the next "learning process" starts. However, a crucial difference is that in our model the sequence of priors about dividends is equal to their *true* distribution. Excess volatility therefore is not due to a high subjective variance of fundamentals. This feature is important – it implies that our model applies when the distribution of (tangible) fundamentals is well understood. The only deviation from rational expectations in our setup is the presence of ambiguous *intangible* signals.

The model of 9/11 in Section 4 focuses on a learning process triggered by an event that increases uncertainty. It is thus closer to a second group of papers that tries to explain post-event abnormal returns ("underreaction") through the gradual incorporation of information into prices.[17] It may be viewed as a model of negative underreaction in periods of ambiguous news. In these periods, underreaction is likely even if there is no change in fundamentals that is gradually revealed. Moreover, the slide in prices is reversed in the long run as agents learn that fundamentals have not changed.

The mechanism that generates negative skewness in our model differs from existing explanations. Veronesi [38] shows, in a Bayesian model with risk averse agents, that prices respond more to bad news in good times and conversely. This obtains because,

[16] See Timmermann [35, 36], Bossaerts [6], and Lewellen and Shanken [30] for models of nonstationary transitions and Brandt, Xeng, and Zhang [7], Veronesi [38] and Brennan and Xia [9] for models with persistent hidden state variables.

[17] See, Brav and Heaton [8] for an overview and discussion of this literature.

in his setup, news that contradict the current belief increase the conditional variance of asset payoffs. Our result differs in two ways. First, since it does not rely on risk aversion, it is relevant also if uncertainty is idiosyncratic and investors are well diversified. Second, ambiguous signals entail an asymmetric response whether or not times are good. They thus induce *unconditional* negative skewness in returns. To explain the latter fact, Hong and Stein [27] and Veldkamp [37] have proposed mechanisms for bad news to be more clustered. Such mechanisms could reinforce skewness in our setting, but they are not necessary for it to obtain. In terms of cross-sectional implications, our model predicts that the relative importance of intangible information generates skewness. As discussed in Section 3, this is consistent with existing evidence on the cross section of stocks in the U.S.

6 Appendix

The key step in calculating moments of prices and returns in Section 3.3 is to find moments of $\gamma_t s_t$, where γ_t denotes the random variable that is equal to $\bar{\gamma}$ when $s_t \leq 0$, and equal to $\underline{\gamma}$ otherwise. We summarize the properties of $\gamma_t s_t$ here. All moments are calculated under the true signal distribution. Since $E\left[s_t|\gamma_t = \bar{\gamma}\right] = E_t\left[s_t|s_t \leq 0\right])$, the calculations make heavy use of formulas for moments of truncated normal distributions, $E\left[s_t|s_t \geq 0\right] = \sigma\sqrt{\frac{2}{\pi}}$, $E\left[s_t^2|s_t \geq 0\right] = \sigma^2$ and $E\left[s_t^3|s_t \geq 0\right] = \frac{\sigma^3}{\sqrt{2\pi}}$. The mean of $\gamma_t s_t$ is

$$\begin{aligned}
E\left[\gamma_t s_t\right] &= E\left[\gamma_t E[s_t|\gamma_t]\right] \\
&= \frac{1}{2}\bar{\gamma}\left(-\sqrt{\frac{2}{\pi}}\sigma\right) + \frac{1}{2}\underline{\gamma}\left(\sqrt{\frac{2}{\pi}}\sigma\right) \\
&= -\left(\bar{\gamma} - \underline{\gamma}\right)\frac{\sigma_s}{\sqrt{2\pi}},
\end{aligned}$$

and the variance is

$$\begin{aligned}
var\left(\gamma_t s_t\right) &= E\left[\gamma_t^2 s_t^2\right] - E\left[\gamma_t s_t\right]^2 \\
&= E\left[\gamma_t^2 E\left[s_t^2|\gamma_t\right]\right] - E\left[\gamma_t E\left[s_t|\gamma_t\right]\right]^2 \\
&= E\left[\gamma_t^2 var\left(s_t|\gamma_t\right)\right] + var\left(\gamma_t E\left[s_t|\gamma_t\right]\right) \\
&= \frac{1}{2}\left(\bar{\gamma}^2 + \underline{\gamma}^2\right)\sigma_s^2\left(1 - \frac{2}{\pi}\right) + \frac{1}{4}\left(\sqrt{\frac{2}{\pi}}\sigma_s\left(\bar{\gamma} + \underline{\gamma}\right)\right)^2 \\
&= \frac{1}{2}\sigma_s^2\left(\bar{\gamma}^2 + \underline{\gamma}^2 - \frac{1}{\pi}\left(\bar{\gamma} - \underline{\gamma}\right)^2\right),
\end{aligned}$$

To determine the variance of returns, we also need the term

$$\begin{aligned}
cov\left(\gamma_t s_t, u_{t+1}\right) &= E\left[\gamma_t s_t u_{t+1}\right] = E\left[\gamma_t s_t E\left[u_{t+1}|s_t, \gamma_t\right]\right] \\
&= \gamma^* E\left[\gamma_t s_t^2\right] = \gamma^* E\left[\gamma_t E\left[s_t^2|\gamma_t\right]\right] \\
&= \gamma^* \frac{\bar{\gamma} + \underline{\gamma}}{2}\sigma_s^2 = \frac{\bar{\gamma} + \underline{\gamma}}{2}\sigma_u^2.
\end{aligned}$$

The third centered moment of $\gamma_t s_t$ is

$$\mu_3(\gamma_t s_t) = E\left[\gamma_t^3 s_t^3\right] - E\left[\gamma_t s_t\right]^3 - 3E\left[\gamma_t s_t\right] var\left(\gamma_t s_t\right).$$

The only as yet unknown term is

$$\begin{aligned} E\left[\gamma_t^3 s_t^3\right] &= E\left[\gamma_t E\left[s_t^3|\gamma_t\right]\right] \\ &= -\frac{1}{2}\sqrt{\frac{2}{\pi}}\left(5-\frac{4}{\pi}\right)\left(\bar{\gamma}-\underline{\gamma}\right)\sigma_s^3, \end{aligned}$$

where the second equality follow from algebra and the third moment of the truncated normal distribution. More algebra then delivers

$$\mu_3(\gamma_t s_t) = -\frac{\sigma_s^3}{\sqrt{2\pi}}\left(\frac{\left(\bar{\gamma}-\underline{\gamma}\right)^3}{\pi} + \bar{\gamma}\underline{\gamma}\left(\bar{\gamma}-\underline{\gamma}\right) + \frac{7\pi-8}{2\pi}\left(\bar{\gamma}^3-\underline{\gamma}^3\right)\right) < 0.$$

Since $\sigma_s^2 = \sigma_u^2/\gamma^*$, this expression is indeed proportional to σ_u^2, where the factor of proportionality depends only on $\bar{\gamma}$, $\underline{\gamma}$ and γ^*.

Finally, consider the third centered moment of returns. Since $u_{t+1} - \gamma_t s_t$ and $\gamma_{t+1}s_{t+1}$ are independent, we just need to compute

$$\begin{aligned} \mu_3(u_{t+1} - \gamma_t s_t) &= \mu_3(u_{t+1}) - \mu_3(\gamma_t s_t) \\ &\quad +3E\left[u_{t+1}\left(\gamma_t s_t - E\left[\gamma_t s_t\right]\right)^2\right] - 3E\left[u_{t+1}^2\left(\gamma_t s_t - E\left[\gamma_t s_t\right]\right)\right]. \end{aligned}$$

The first term is zero since since u_{t+1} is normal. In the third and fourth terms, use conditional normality of u_{t+1} to get

$$\begin{aligned} E\left[E\left[u_{t+1}^2|s_t,\gamma_t\right]\left(\gamma_t s_t - E\left[\gamma_t s_t\right]\right)\right] &= E\left[\left(\gamma^{*2}s_t^2 + (1-\gamma^*)\sigma_u^2\right)\left(\gamma_t s_t - E\left[\gamma_t s_t\right]\right)\right] \\ &= \gamma^{*2}E\left[s_t^2\left(\gamma_t s_t - E\left[\gamma_t s_t\right]\right)\right] \\ E\left[E\left[u_{t+1}|s_t,\gamma_t\right]\left(\gamma_t s_t - E\left[\gamma_t s_t\right]\right)^2\right] &= E\left[\left(\gamma^* s_t\right)\left(\gamma_t s_t - E\left[\gamma_t s_t\right]\right)^2\right] \\ &= \gamma^* E\left[\gamma s_t^2\left(\gamma_t s_t - E\left[\gamma_t s_t\right]\right)\right] \end{aligned}$$

Putting terms together, we have,

$$\begin{aligned} \mu_3\left(\hat{R}_{t+1}\right)(1-\beta g\phi)^3 &= \left((\beta g)^3 - 1\right)\mu_3(\gamma_t s_t) \\ &\quad +3\gamma^*\left(E\left[\gamma_t\left(\gamma_t-\gamma^*\right)s^3\right] - E\left[\gamma_t s_t\right]\left(2E\left[\gamma_t s_t^2\right] - \gamma^* E\left[s^2\right]\right)\right) \\ &= \left((\beta g)^3 - 1\right)\mu_3(\gamma_t s_t) \\ &\quad -6\sqrt{\frac{2}{\pi}}\left(1-\frac{1}{\pi}\right)\sigma_s^3\left(\bar{\gamma}-\underline{\gamma}\right)\left(\bar{\gamma}+\underline{\gamma}-\gamma^*\right). \end{aligned}$$

Since $\sigma_s^2 = \sigma_u^2/\gamma^*$, we obtain the expression in the text.

References

[1] A. Abel, An exploration of the effects of pessimism and doubt on asset returns, *J. Econ. Dynam. Control*, 26 (2002), 7-8, 1075-1092.

[2] N.Barberis and M. Huang, Mental accounting, loss aversion and individual stock returns, *J. Finance* 56 (2001), 1247-1292.

[3] N.Barberis and M. Huang, The loss aversion / narrow framing approach to stock market pricing and participation puzzles, U. Chicago, 2004.

[4] N. Barberis and R. Thaler, A survey of behavioral finance, in G. Constantinides et al., *Handbook of the economics of finance,* 2003.

[5] G. Bekaert and G. Wu, Asymmetric volatility and risk in equity markets, *Rev. Finan. Studies* 13, 1 (2002), 1-42.

[6] P. Bossaerts, Learning-induced securities price volatility, Caltech, 2001.

[7] M. Brandt, Q. Zeng and L. Zhang, Equilibrium stock return dynamics under alternative rules of learning about hidden states, Wharton 2001.

[8] A. Brav and J. Heaton, Competing theories of financial anomalies, *Rev. Finan. Studies* 15 (2002), 575-606.

[9] M. J. Brennan and Y.Xia, Stock price volatility and the equity premium, *J. Monet. Econ.* 47 (2001), 249-283.

[10] M. Cagetti, L.P. Hansen, T. J. Sargent and N. Williams, Robustness and pricing with uncertain growth, *Rev. Financ. Stud*. 15 (2002), 363-404.

[11] J. Campbell, Asset Prices, consumption and the business cycle, in *Handbook of Macroeconomics,* J. Taylor and M. Woodford, eds., Amsterdam: North Holland, 1999.

[12] J. Campbell, M. Lettau, B. Malkiel and Y. Xu, Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *J. Finance,* 56 (2001), 1-43.

[13] A. Caplin and J. Leahy, Psychological expected utility theory and anticipatory feelings, *Quart. J. Econ.* 116 (2001), 55-80.

[14] Z. Chen and L.G. Epstein, Ambiguity, risk and asset returns in continuous time, *Econometrica* 70 (2002), 1403-43.

[15] J. Chen, H. Hong and J. Stein, Forecasting crashes: trading volume, past returns and conditional skewness in stock prices, *J. Finan. Econ.* 61 (2001). 345-381.

[16] K. Daniel, D. Hirshleifer, and A. Subrahmanyam, Investor psychology and security market under- and overreactions, *J. Finance*, 53, 5 (1998), 1839-1886.

[17] L.G. Epstein and M. Schneider, Recursive multiple-priors, *J. Econ. Theory* 113 (2003), 1-31 (with Martin Schneider).

[18] L.G. Epstein and M. Schneider, IID: independently and indistinguishably distributed, *J. Econ. Theory* 113 (2003), 32-50.

[19] L.G. Epstein and M. Schneider, Learning under ambiguity, mimeo, 2004.

[20] L.G. Epstein and T. Wang, Intertemporal asset pricing under Knightian uncertainty, *Econometrica* 62 (1994), 283-322.

[21] M. Feldman, On the generic nonconvergence of Bayesian actions and beliefs, *Econ. Theory* 1 (1991), 301-321.

[22] I. Gilboa and D. Schmeidler, Maxmin expected utility with nonunique prior, *J. Math. Econ.* 18 (1989), 141-153.

[23] L.P. Hansen, T.J. Sargent and T. Tallarini, Robust permanent income and pricing, *Rev. Ec. Stud.* 66 (1999), 873-907.

[24] L.P. Hansen, T.J. Sargent and N. Wang, Robust permanent income and pricing with filtering, *Macroeconomic Dynamics* 6 (2002), 40-84.

[25] L.P. Hansen and K. Singleton, Generalized instrumental variable estiamtion of nonlinear rational expectatioons models, *Econometrica* 50 (1982), 1269-1288.

[26] C. Harvey and A. Siddique, Conditional skewness in asset pricing tests, *J. Finance* 55, 3 (2000), 1263-1295.

[27] H. Hong and J. Stein, Differences of opinion, short-sales constraints and market crashes, Stanford 2002.

[28] B. N. Lehmann, Residual Risk Revisited, *J. Econometrics,* 45 (1990), 71-97.

[29] B.G. Malkiel and Y. Xu, Idiosyncratic risk and security returns, UT Dallas 2002.

[30] J. Lewellen and J. Shanken, Learning, asset pricing tests and market efficiency, *J. Finance,* 58,3 (2002), 1113-1145.

[31] M. Marinacci, Limit laws for non-additive probabilities, and their frequentist interpretation, *J. Econ. Theory* 84 (1999), 145-195.

[32] S. Mukerji and J. Tallon, Ambiguity aversion and incompleteness of financial markets, *Rev. Econ. Stud.* 68 (2001), 883-904.

[33] C.K. Nielsen, Rational belief structures and rational belief equilibria, *Economic Theory* 8 (1996) 3, 399-422.

[34] B. Routledge and S. Zin, Model uncertainty and liquidity, NBER #8683, 2001.

[35] A. Timmermann, How learning in financial markets generates excess volatility and predictability in stock returns, *Quart. J. Economics* 108 (1993), 1135-1145.

[36] A. Timmermann, Excess volatility and predictability of stock returns in autoregressive dividend models with learning, *Rev. Ec. Stud.* (1996), 523-557.

[37] L. Veldkamp, Slow boom, sudden crash, mimeo, NYU 2003.

[38] P. Veronesi, Stock market overreaction to bad news in good times: a rational expectations equilibrium model, *Rev. Finan. Stud.* 12 (1999), 976-1007.

[39] P. Veronesi, How does information quality affect stock returns?, *J. Finance* 55, 2 (2000), 807-837.

[40] W. West, Dividend innovations and stock price volatility, *Econometrica* 56 (1988), 37-61.