Learning and Asset Prices Under Ambiguous Information

Markus Leippold
Tanaka Business School, Imperial College London

Fabio Trojani
Swiss Institute of Banking and Finance, University of St. Gallen

Paolo Vanini
Zurich Cantonal Bank

In a Lucas exchange economy with standard power utility, we study asset prices under learning and ambiguous information. In contrast with models featuring only learning or ambiguity, our model is successful in matching the equity premium, the interest rate, and the volatility of stock returns under empirically reasonable parameters. Our closed-form formulas also show that a severe downward bias arises in the empirical relation between stock returns and return volatility. We quantify this bias in simulations and show that our model can explain why such a relation is difficult to detect in the data. (*JEL* G1, G11, G12)

The equity premium puzzle, the interest rate puzzle, and the excess volatility are well-known empirical failures of the consumption capital asset pricing model pioneered by Lucas (1978) and Breeden (1979). In this article, we present an extension of the Lucas exchange economy that helps explain these puzzles in a setting of learning and ambiguity (Knightian uncertainty) aversion. Moreover, we show that learning and ambiguity aversion imply a nonlinear equilibrium trade-off between risk and return, which generates a severe downward bias in the empirical relation between stock returns and returns volatility.

Models of pure learning can explain excess volatility, but they can make the equity premium even more of a puzzle [see, e.g. Veronesi (2000)].

We are grateful to Andrew Abel, Tobias Moskowitz (the editor), and an anonymous referee for many valuable suggestions. We also thank Freddy Delbaen, David Feldman, Günter Franke, Rajna Gibson, Jens Jackwerth, Yvan Lengwiler, Marco LiCalzi, Abraham Lioui, Alessandro Sbuelz, Sandra Sizer, Pietro Veronesi, Yihong Xia, the participants of the 2004 European Summer Symposium in Financial Markets in Gerzensee, the 2005 International Finance Conference at the University of Copenhagen, the EFMA 2006 (Madrid), and of the finance seminars at the University of Basel, the University of Konstanz, the University of Frankfurt, the University of Venice, the University of Zurich, and the ETH Zurich. The authors gratefully acknowledge the financial support of the Swiss National Science Foundation (NCCR FINRISK and grants 101312-103781/1 and 100012-105745/1) and the University Research Priority Program "Finance and Financial Markets" of the University of Zurich. Address correspondence to Markus Leippold, Tanaka Business School, Imperial College London, South Kensington Campus, London SW7 2AZ, UK, or e-mail: m.leippold@imperial.ac.uk.

doi:10.1093/rfs/hhm035 Advance Access publication September 12, 2007

On the other hand, in an economy with a constant opportunity set and time-additive preferences, ambiguity aversion decreases the interest rate, but fails to generate excess volatility and a sizeable equity premium under realistic model parameters [see, e.g. Maenhout (2004)]. Therefore, we ask if the interaction of learning and ambiguity aversion can simultaneously explain the above puzzles.

We extend rational settings of Bayesian learning[1] in a parsimonious and tractable way. As in Veronesi (2000), the representative investor has a standard coefficient of relative risk aversion (CRRA) utility and does not observe the expected growth rate of dividends. She tries to infer the "true" expected growth rate through the observation of dividends and another noisy signal. However, the investor does not know the form of the data-generating process of dividends. This assumption introduces ambiguity into the economy. We model investor's aversion to ambiguity by a max–min expected utility optimization problem.

First, we find that learning under ambiguity aversion implies an equilibrium discount for ambiguity if and only if the relative risk aversion is less than 1, or, equivalently, the elasticity of intertemporal substitution (EIS) is above 1. For low risk aversion, learning and ambiguity aversion increase conditional equity premia and volatilities. We also show that the part of the equity premium that is due to the interaction of learning and ambiguity aversion is the dominant part.

Second, learning and ambiguity aversion imply lower equilibrium interest rates, regardless of risk aversion. Thus, with low risk aversion we get both a higher equity premium and a lower interest rate. This finding is an important feature of our setting, because it explains both the equity premium and the interest rate puzzles for moderate amounts of ambiguity.

Third, the true theoretical equilibrium relation between excess returns and conditional variances is highly time-varying. Estimated relations between excess returns and conditional variances have undetermined signs over time, and generate a time-varying bias in the naively estimated risk-return trade-off using, for examble., regression methods. This feature of our model is consistent with the weak empirical evidence of a positive relationship between expected returns and conditional return variance.[2]

The article is organized as follows. The next section reviews the literature on ambiguity aversion. Section 2 introduces our model. In Section 3, we study the properties of the optimal learning dynamics. Section 4

[1] See the early work of Detemple (1986), Dothan and Feldman (1986), and Gennotte (1986). More recent contributions include, e.g. Barberis (2000), Brennan (1998), Brennan and Xia (2001), Kandel and Stambaugh (1995), Pástor (2000), Pástor and Stambaugh (2000), Veronesi (1999), Veronesi (2000), and Xia (2001). For a recent review of the literature, see Feldman (2005).

[2] For example, French, Schwert, and Stambaugh (1987) and Campbell and Hentschel (1992) find a positive, mostly nonsignificant, risk-return trade-off, and Campbell (1987) and Brandt and Kang (2004) find a negative relation between the conditional variance and the conditional expected return.

characterizes and discusses conditional asset pricing relations. Section 5 concludes.

1. Background

From the Ellsberg (1961) paradox, we know that investors behave differently under ambiguity and risk aversion. Therefore, distinguishing between ambiguity aversion and risk aversion is economically and behaviorally important. Gilboa and Schmeidler (1989) suggest an atemporal axiomatic framework of ambiguity aversion in which preferences are represented by max–min expected utility over a set of multiple prior distributions. Inspired by their approach, Epstein and Wang (1994) study some asset pricing implications of max–min expected utility in a discrete-time infinite horizon economy. Epstein and Schneider (2003) provide a discrete-time axiomatic foundation for that model, and show that a dynamically consistent version of the Gilboa and Schmeidler (1989) preferences can be represented by a recursive max–min expected utility over a set of multiple distributions. Chen and Epstein (2002) extend the model of Epstein and Schneider (2003) to continuous time. Hansen, Sargent, and Tallarini (1999, in discrete time) and Anderson, Hansen, and Sargent (2003, in continuous time) suggest another model of intertemporal ambiguity aversion based on an alternative form of max–min expected utility.

Various authors propose models of full information economies with ambiguity aversion to explain several important characteristics of asset prices. In these models, the investor observes perfectly the state variables that determine the opportunity set, but she is not fully aware of the probability distribution of the state variables. Consequently, some form of conservative worst-case optimization determines her optimal decision rules.[3] However, none of these models addresses the equity premium, excess volatility, and low interest rate puzzles simultaneously. Moreover, Maenhout (2004) and Sbuelz and Trojani (2002) show that it is difficult to fit the equity premium puzzle under standard power utility without imposing an unrealistically high degree of ambiguity.

Only more recently have a few authors addressed the issue of learning under ambiguity aversion. In a production economy driven by a two-state Markov chain, Cagetti, Hansen, Sargent, and Williams (2002) use robust filtering theory to study the impact of learning and ambiguity aversion on the aggregate capital stock, equity premia, and price–dividend

[3] Examples of such models include, among others, Gagliardini, Porchia, and Trojani (2004, term structure of interest rates), Epstein and Miao (2003, home bias), Liu, Pan, and Wang (2004, option pricing with rare events), Maenhout (2004, equity premium puzzle), Routledge and Zin (2001, liquidity), Sbuelz and Trojani (2002, equity premium puzzle), Trojani and Vanini (2002, 2004, equity premium puzzle and stock market participation), and Uppal and Wang (2003, home bias).

ratios. Using numerical methods, they show that the precautionary saving increases in a way similar to the effect of a higher subjective time preference rate. Equity premia increase and price–dividend ratios decrease.

Our model differs from that in Cagetti, Hansen, Sargent, and Williams (2002) in several aspects. We work with a Lucas exchange economy and use a tractable homothetic setting of preferences under ambiguity aversion. Because our model is tractable, we can study and analyze in closed form all the relevant asset-pricing relations and their dependence on model parameters. For instance, we show that in the partial-information Lucas economy, ambiguity aversion can fail to increase equity premia if standard risk aversion is too high or, equivalently, the EIS is too low. We also allow for external public signals and for heterogeneous ambiguity across the possible states of the economy. These extensions have implications for asset prices. For example, we find that ambiguity premia caused by public signals can be quite large and that there is an ambiguity premium for information quality.

Epstein and Schneider (2002) use a discrete-time model to highlight that learning about an unknown parameter under multiple likelihoods can fail to resolve ambiguity asymptotically. In a similar model, Epstein and Schneider (2004) illustrate the impact of an ambiguous signal precision on asset prices. In a risk-neutral economy, they find that ambiguous information quality (defined by a set of possible signal precision parameters) can increase the equity premium, the volatility of returns, and can generate skewness. Using numerical solutions for a model with risk aversion, they demonstrate that their model can help understand stock prices in the aftermath of 9/11.

Our article is different from that of Epstein and Schneider (2004). Their focus is on ambiguous signal precision and its impact on return volatility and skewness. We focus on ambiguous signals about the expected dividend growth and the resulting implications primarily for equity premia. In this context, we show that for low risk aversion and a realistic amount of ambiguity, the joint interaction of learning and ambiguity aversion is responsible for large equity premia. Additionally, the model predictions are consistent with the interest rate and excess volatility puzzles. Since we compute equilibrium quantities under standard assumptions on the utility function, we can also disentangle the impact of learning, ambiguity aversion, and risk aversion on equity premia.

2. The Model

We develop an equilibrium setting of learning under ambiguity aversion. Our model consists of three main components: a parametric reference model for the unobservable dividend drift, a set of multiple likelihoods for the dynamics of dividends, and a max–min expected utility optimization problem for the representative investor.[4]

2.1 The reference model

We consider a Lucas (1978) economy populated by a CRRA investor with utility function

$$u(C,t) = e^{-\delta t}\frac{C^{1-\gamma}}{1-\gamma}, \tag{1}$$

where $\gamma > 0$ is the CRRA. We give the representative investor a parametric reference model, which is an approximate description of the dynamics of dividends D.

To focus on the genuine implications of learning and ambiguity aversion, we assume a reference model in which dividends follow a geometric Brownian motion, as in Veronesi (2000).

Definition 1. *Under the reference model probability $\mathbb{P}$, dividends have the dynamics*

$$\frac{dD}{D} = \theta dt + \sigma_D dB_D, \tag{2}$$

where $\theta \in \Theta := \{\theta_1, \theta_2, ..., \theta_n\}$, with $\theta_1 < \theta_2 < ... < \theta_n$, $\sigma_D > 0$ and B_D is a $\mathbb{P}-$Brownian motion.

The reference model is a rough approximation of reality. Therefore, the representative investor has some motivated doubts about the specification of the reference model. The representative investor observes only dividends and a noisy unbiased signal e for the dividend drift, which is unobservable. Under the reference model probability, the signal dynamics are

$$de = \theta dt + \sigma_e dB_e, \tag{3}$$

where $\sigma_e > 0$ and B_e is a $\mathbb{P}$–Brownian motion independent of B_D.

In the standard Bayesian framework, Definition 1 implies a single likelihood for the dividend dynamics. The specific value of the parameter θ is unknown and the only relevant statistical uncertainty about the dividend dynamic (2) is parametric. Therefore, the standard filtering process leads to asymptotic learning of the unknown dividend drift θ in the class of candidate drift values Θ.

We depart from the above setting by allowing for a misspecification in the reference model. Misspecifications can have a general nonparametric form, so that they cannot be consistently detected even if we use approaches based on parametric Bayesian model selection.

[4] See also Gilboa and Schmeidler (1989), Chen and Epstein (2002), Epstein and Schneider (2003).

2.2 Multiple likelihoods

Compared with the Bayesian single-likelihood hypothesis in Definition 1, an investor with multiple likelihood computes a set of relevant beliefs about the unknown dividend drift. This set of beliefs represents the investor's ambiguity about the unobservable expected dividend growth rate. It can also be interpreted as a class of alternative specifications of the reference model.

Assumption 1. *(i) The set of multiple likelihoods consists of absolutely continuous probabilities $\mathbb{P}^h(\theta)$, such that the dividend dynamics are*

$$\frac{dD(t)}{D(t)} = (\theta + h(\theta)\sigma_D)dt + \sigma_D dB_D(t), \quad t \geq 0, \tag{4}$$

for some $\theta \in \Theta$ and some drift distortion $\theta + h(\theta)\sigma_D \in \Xi(\theta)$. (ii) Under any probability $\mathbb{P}^h(\theta)$ signals are unbiased:

$$de(t) = (\theta + h(\theta)\sigma_D)dt + \sigma_e dB_e(t), \quad t \geq 0. \tag{5}$$

(iii) The representative investor has some beliefs $(\widehat{\pi}_1, .., \widehat{\pi}_n)$ at time $t = 0$ about the a priori plausibility of the different sets $\Xi(\theta_1), \ldots, \Xi(\theta_n)$ of drift distortions h.

Under Assumption **1**, the representative investor recognizes that statistically, the whole neighborhood $\Xi(\theta)$ of dividend drift changes is hardly distinguishable from a zero drift change, that is, from the reference model dynamics with drift θ in Definition 1. If $\Xi(\theta) = \{0\}$ for all $\theta \in \Theta$, the Bayesian setup of Veronesi (2000) follows and the investor is concerned exclusively with the noisiness of the signal on the parameter value θ.

The size of the neighborhood $\Xi(\theta)$ describes the degree of ambiguity associated with any possible reference model drift θ. The broader $\Xi(\theta)$, the more ambiguous are the signals about a specific dividend drift $\theta + h(\theta)\sigma_D \in \Xi(\theta)$. This ambiguity reflects the fact that there are aspects of the unobservable dividend dynamics that are hardly possible, or even impossible, to ever know. Therefore, our representative investor is aware that, empirically, identifying the exact functional form of a possible mean reversion in the dividend drift dynamics is a virtually infeasible task.[5] Accordingly, the investor tries to understand only a limited number of features of the dividend dynamics.

We specify $\Xi(\theta)$ by ensuring that it contains likelihood specifications that are statistically close, in some appropriate statistical measure of model discrepancy, to the reference model. Since we constrain the relevant misspecifications to be small, they are hardly statistically detectable. This

[5] Shepard and Harvey (1990) show that in finite samples, it is very difficult to distinguish between a purely iid process and one which that incorporates a small persistent component.

property defines a whole neighborhood of small, but otherwise arbitrary, misspecifications of the reference model.

Assumption 2. *For any $\theta \in \Theta$ we define $\Xi(\theta)$ by*

$$\Xi(\theta) := \left\{ \theta + h(\theta)\sigma_D : \frac{1}{2}h^2(\theta) \leq \eta(\theta) \text{ for all } t \geq 0 \right\}, \tag{6}$$

where $\eta(\theta) \geq 0$. Moreover, for any $i \neq j$,

$$\Xi(\theta_i) \cap \Xi(\theta_j) = \emptyset. \tag{7}$$

Assumption **2** states that for any probability $\mathbb{P}^h(\theta)$ implied by an admissible likelihood, the maximal drift distortion relative to the reference model is bounded by:

$$-\sqrt{2\eta(\theta)} \leq h(\theta) \leq \sqrt{2\eta(\theta)}. \tag{8}$$

Therefore, by using the parameter η we can control the size of the set of multiple likelihoods of the representative investor and parameterize in a parsimonious way her perception of the ambiguity of dividends. For $\eta = 0$, ambiguity vanishes and we obtain the standard Bayesian setting. For increasing η, we obtain settings in which ambiguity is more pronounced. Since Equation (6) does not make specific assumptions on the parametric functional form of $h(\theta)$, the set of likelihoods implied by the neighborhood $\Xi(\theta)$ is nonparametric and there is no simple way we can put a prior on it.

Equation (6) specifies the size of ambiguity within each neighborhood $\Xi(\theta_i)$, $i = 1, \ldots, n$. Condition (7) requires these neighborhoods to be pairwise disjoint. This condition allows us to map uniquely any admissible dividend drift $\theta_i + h(\theta_i)\sigma_D$ into a single neighborhood $\Xi(\theta_i)$. Any such drift has then the interpretation of a drift process difficult to distinguish statistically from the reference drift θ_i. We can think of the reference model drifts as describing different approximate growth rates of the economy. Conditional on the reference model growth rate, there is a whole neighborhood of similar models for dividends that are very difficult to distinguish statistically. We can interpret these neighborhoods as a family of disjoint sets of approximate scenarios for the growth rate of dividends. To model a different degree of statistical uncertainty associated with these different sets of scenarios, we allow for different sizes for the neighborhoods. The representative investor faces ambiguity within but not between neighborhoods. Therefore, she tries to learn the neighborhood containing the true model for dividends, but not the specific model that generates the dividend data.

2.3 The size of ambiguity

Given a set of multiple likelihoods, we have to ask how we can impose realistic maximal drift distortions $\theta + h(\theta)\sigma_D \in \Xi(\theta)$ relative to the reference model.

Given the reference belief, condition (7) already imposes a constraint on the size of $\eta(\theta)$. Therefore, we can only select a moderate size for the ambiguity parameter $\eta(\theta)$, relative to the distances between reference model drifts $\theta_1, \ldots, \theta_n$. In addition, we note from condition (7) that the size of $\eta(\theta)$ determines the degree of pessimism implied by a worst-case dividend drift $\theta - \sqrt{2\eta(\theta)}\sigma_D$ in the neighborhood $\Xi(\theta)$. Given the model parameters, we can compute the quantity $\sqrt{2\eta(\theta)}\sigma_D$. In this way, we can study the difference between the expected consumption growth rate under the reference belief and under the worst-case belief. In all our subsequent model calibrations, we use this approach to constrain the size of $\eta(\theta)$ in a way that avoids worst-case beliefs that are too pessimistic. Table 1 reports the differences between the reference belief and the worst-case belief about the expected consumption growth rate. For all calculations, we use $\sigma_D = 0.0325$, which corresponds to the volatility of consumption growth as estimated in Campbell (1999) for an annual sample from 1891 to 1994. For the given choices of ambiguity $\eta(\theta)$ in Table 1, the differences between the reference belief and the worst-case belief of the expected consumption growth rate are always less than 0.46%.

Finally, from a statistical viewpoint, we can make use of an argument proposed by Anderson, Hansen, and Sargent (2003) to constrain the degree of pessimism in our model. They quantify the statistical closeness of two likelihood beliefs by calculating the average error probability in a Bayesian likelihood ratio test of two competing models, given a continuous record of data of length N. Intuitively, likelihoods that are statistically close must

Table 1
Degree of pessimism and model detection error probabilities

		Error probabilities $\epsilon_N(\theta)$			
$\eta(\theta)$	Degree of pessimism	$N = 100$	$N = 200$	$N = 300$	$N = 400$
0.0001	0.05 %	0.47	0.46	0.45	0.44
0.0010	0.15 %	0.41	0.38	0.35	0.33
0.0025	0.23 %	0.36	0.31	0.27	0.24
0.0050	0.33 %	0.31	0.24	0.19	0.16
0.0075	0.40 %	0.27	0.19	0.14	0.11
0.0100	0.46 %	0.24	0.16	0.11	0.08

The entries report the degree of pessimism and the average model detection error probability $\epsilon_N(\theta)$ implied by different sizes for the ambiguity parameter $\eta(\theta)$ and for sample sizes $N = 100, 200, 300, 400$. We calculate the degree of pessimism as the difference between the expected consumption growth rate under the reference belief and under the worst-case belief. For the calculations, we assume $\sigma_D = 0.0325$.

be associated with large error probabilities, but likelihoods that are easy to distinguish have to imply low error probabilities. Therefore, we can use the size of these error probabilities as a metric for the closeness of the worst-case and reference beliefs.

Let $l_N(\theta)$ be the log-likelihood function of the worst-case likelihood relative to the likelihood of the reference model associated with neighborhood $\Xi(\theta)$. Treating these likelihoods symmetrically, the average probability of a model detection error in the corresponding likelihood ratio test is

$$\epsilon_N(\theta) = 0.5 \cdot \mathbb{P}(l_N(\theta) > 0) + 0.5 \cdot \mathbb{P}^{h^*}(l_N(\theta) < 0), \tag{9}$$

where $h^* = -\sqrt{2\eta(\theta)}$ is the standardized worst-case drift distortion associated with neighborhood $\Xi(\theta)$. The quantity $\epsilon_N(\theta)$ is a simple equally weighted average of the probability of rejecting the reference model when it is true ($\mathbb{P}(l_N(\theta) > 0)$) and the probability of accepting the reference model when the worst-case model is true ($\mathbb{P}^{h^*}(l_N(\theta) < 0)$. Using the geometric Brownian motion structure of dividends under the reference model and the worst-case likelihood, it follows that

$$\epsilon_N(\theta) = \Phi\left(-\sqrt{\frac{\eta(\theta)N}{2}}\right), \tag{10}$$

where Φ is the cumulative distribution function of the standard normal distribution. The error probability $\epsilon_N(\theta)$ is decreasing in N and $\eta(\theta)$. Therefore, if $\eta(\theta)$ is large, the investor could use the dividend data to statistically separate the reference belief from the worst-case belief.

In Table 1, we see that for a century of annual observations ($N = 100$) the smallest error probability $\epsilon_N(\theta)$ is 24% ($\eta = 0.01$). In this case, the investor would not be able to identify the correct model 24% of the times. For a century of quarterly observations ($N = 400$), the smallest error probability is 8%. In both cases, the error probability is above the significance level typically used in many empirical studies to reject the null hypothesis.

2.4 Ambiguity aversion and intertemporal max–min expected utility

We use $\mathcal{F}(t)$ to denote the information available at time t. This information contains all possible realizations of dividends and signals. P is the price of the risky asset in the economy, r the instantaneous interest rate, and $\eta(\theta)$ the function that describes ambiguity. The representative investor determines consumption and investment plans $C(t)$ and $w(t)$ by solving the intertemporal max–min expected utility optimization problem

$$\max_{C,w} \inf_{h(\theta)} E\left[\int_0^\infty u(C,s)\,ds \middle| \mathcal{F}(0)\right], \tag{11}$$

subject to the dividend and wealth dynamics

$$dD = (\theta + h(\theta)\,\sigma_D)\,Ddt + \sigma_D DdB_D,$$
$$dW = W\left[w\left(\frac{dP + Ddt}{P}\right) + (1-w)\,rdt\right] - Cdt,$$

where for any $\theta \in \Theta$ the standardized drift distortion is such that $\theta + h(\theta)\,\sigma_D \in \Xi(\theta)$ and Assumption **2** holds.

The max–min problem (11) models the representative investor's optimal behavior, given her attitudes to risk and ambiguity. In addition to an optimal consumption/investment policy, the representative investor must select an optimal worst-case belief $\theta + h(\theta)\sigma_D$ out of the admissible class $\Xi(\theta)$. Therefore, an optimal belief is determined endogenously as a function of preferences. This differs sharply from the standard Bayesian setting, in which beliefs are fixed by a parametric assumption on the unobservable dynamics of the dividend drift process.

3. Asset Prices

We first study the representative investor's optimal learning behavior under ambiguity. With the endogenous set of optimal dividend-drift prediction processes at hand, we solve the equilibrium optimization problem and compute asset prices.

3.1 Learning under ambiguity aversion

We start with determining the endogenous set of optimal dividend-drift prediction processes. Learning under ambiguity requires a set of Bayesian ahead beliefs as functions of drift distortions $\theta + h(\theta)\,\sigma_D \in \Xi(\theta)$. For a given likelihood $h(\theta)$, let $\pi_i(t)$ be the investor's belief that the drift rate is $\theta_i + h(\theta_i)\,\sigma_D$, conditional on past dividends and signals, that is,

$$\pi_i(t) = \mathbb{P}(\theta + h(\theta)\sigma_D = \theta_i + h(\theta_i)\,\sigma_D | \mathcal{F}(t))\,. \tag{12}$$

The distribution $\Pi(t) := (\pi_1(t), .., \pi_n(t))$ summarizes the investor's beliefs at time t, under the likelihood $h(\theta)$. Given such beliefs, the investor can compute the expected dividend drift at time t as

$$m_{\theta,h} := m_\theta + m_{h(\theta)} := \sum_{i=1}^{n} (\theta_i + h(\theta_i)\,\sigma_D)\,\pi_i(t)\,, \tag{13}$$

where

$$m_{\theta} = \sum_{i=1}^{n} \theta_i \pi_i (t), \quad m_{h(\theta)} = \sum_{i=1}^{n} h(\theta_i) \pi_i (t) \sigma_D. \tag{14}$$

The filtering equations implied by any likelihood $h(\theta)$ are standard [see Liptser and Shiryaev (2001)].

Lemma 1. *Suppose that at time zero, the investor's beliefs are represented by the prior probabilities* $\widehat{\pi}_1, \ldots, \widehat{\pi}_n$*. Under a likelihood* $h(\theta)$*, the dynamics of the optimal filtering probabilities vector* $\pi_1, \ldots, \pi_n$ *are given by*

$$d\pi_i = \pi_i \left(\theta_i + h(\theta_i) \sigma_D - m_{\theta,h}\right) \left(k_D d\widetilde{B}_D^h + k_e d\widetilde{B}_e^h\right); \quad i = 1, .., n, \tag{15}$$

where

$$d\widetilde{B}_D^h = k_D \left(dD/D - m_{\theta,h} dt\right), \quad d\widetilde{B}_e^h = k_e \left(de - m_{\theta,h} dt\right),$$

$k_D = 1/\sigma_D$, $k_e = 1/\sigma_e$. *In this equation,* $\left(\widetilde{B}_D^h, \widetilde{B}_e^h\right)$ *is a* $\mathbb{P}^h(\theta)-$*standard Brownian motion in* $\mathbb{R}^2$ *with respect to the filtration* $\{\mathcal{F}(t)\}$.

Which learning behavior should the investor adopt in an ambiguous environment? Since the investor is not particularly comfortable with a specific element of $\Xi(\theta)$, she will base her beliefs on the whole set of likelihoods implied by $\Xi(\theta)$. This approach generates a whole class $\mathcal{P}$ of dividend-drift prediction processes, which are given by

$$\mathcal{P} = \left\{m_{\theta,h} : \theta + h(\theta) \sigma_D \in \Xi(\theta)\right\}, \tag{16}$$

where the dynamics of the posterior probabilities $\pi_1, .., \pi_n$ under the likelihood $h(\theta)$ are given in equation (15). The set $\mathcal{P}$ represents investor's ambiguity about the true dividend-drift process, conditional on the available information $\mathcal{F}(t)$ generated by dividends and signals. The larger the size of $\Xi(\theta)$, that is, the ambiguity about the dividend dynamics, the larger the set $\mathcal{P}$ of dividend-drift prediction processes.

Using the set $\mathcal{P}$, we write the continuous-time optimization problem (11) as a full information problem in which we define the relevant dynamics in terms of the filtration $\{\mathcal{F}(t)\}$. Since all beliefs implied by likelihoods $\mathbb{P}^h(\theta)$ are absolutely continuous, all relevant processes $(\tilde{B}_D^h, \tilde{B}_e^h)'$ generate the same filtration $\{\mathcal{F}(t)\}$, and the dynamic budget constraint associated with the optimization problem (11) can be equivalently formulated in terms of $m_{\theta,h}$ and $(\tilde{B}_D^h, \tilde{B}_e^h)'$; see Miao (2001) for a related discussion.

3.2 Equilibrium with learning and ambiguity aversion

In our economy, an equilibrium is a vector of processes $(C(t), w(t), P(t), r(t), h(\theta))$ such that the optimization problem (11) is solved and markets clear, that is, $w(t) = 1$ and $C(t) = D(t)$. In equilibrium, the relevant problem reads

$$J(\Pi, D) = \inf_{h(\theta)} E\left[\int_0^\infty e^{-\delta t}\frac{D(t)^{1-\gamma}}{1-\gamma}dt \,\middle|\, \mathcal{F}(0)\right], \tag{17}$$

subject to the dynamics

$$dD = m_{\theta,h}Ddt + \sigma_D Dd\widetilde{B}_D^h, \tag{18}$$

$$d\pi_i = \pi_i\left(\theta_i + h(\theta_i)\sigma_D - m_{\theta,h}\right)\left(k_D d\widetilde{B}_D^h + k_e d\widetilde{B}_e^h\right), \tag{19}$$

where for any $\theta \in \Theta$ we have $\theta + h(\theta)\sigma_D \in \Xi(\theta)$ and Assumption **2** holds.

The key difference to the standard Bayesian setting is that in Equation (17) the investor must select the optimal worst-case forecast procedure for the unknown dividend drift. This worst-case belief selection generates an endogenous systematic discrepancy between the reference model belief and the belief applied by the investor to value financial assets.

In Proposition 1, we solve Equation (17) and compute the impact of ambiguity aversion on the price of equity and the equilibrium interest rate.

Proposition 1. *Let* $\widehat{\theta}_i := \delta + (\gamma - 1)\theta_i + \gamma(1-\gamma)\frac{\sigma_D^2}{2}$ *and assume that*

$$\widehat{\theta}_i + (1-\gamma)\sqrt{2\eta(\theta_i)}\sigma_D > 0, \quad i = 1, \ldots, n. \tag{20}$$

It then follows:

a) *The normalized misspecification* $h^*(\theta)$ *that solves problem (17) is given by:*

$$h^*(\theta_i) = -\sqrt{2\eta(\theta_i)}, \quad i = 1, .., n. \tag{21}$$

b) *The equilibrium price function* $P(\Pi, D)$ *for the risky asset is given by*

$$P(\Pi, D) = D\sum_{i=1}^{n}\pi_i C_i, \tag{22}$$

where

$$C_i = 1/(\widehat{\theta}_i + (1-\gamma)\sqrt{2\eta(\theta_i)}\sigma_D), \quad i = 1, .., n. \tag{23}$$

c) *The equilibrium interest rate r is*

$$r = \delta + \gamma m_{\theta,h^*} - \frac{1}{2}\gamma(\gamma+1)\sigma_D^2, \tag{24}$$

where $m_{\theta,h^*} = m_\theta + m_{h^*(\theta)}$ *is such that*

$$m_{h^*(\theta)} = \sum_{i=1}^{n} h^*(\theta_i)\,\pi_i\sigma_D = -\sum_{i=1}^{n}\sqrt{2\eta(\theta_i)}\pi_i\sigma_D. \quad (25)$$

The constant C_i in Equation (23) is proportional to the representative investor's expectation of discounted lifetime dividends, conditional on the worst-case drift $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$ selected from the neighborhood $\Xi(\theta_i)$:

$$\begin{aligned} C_i &= E^{h^*(\theta_i)}\left[\int_s^\infty e^{-\delta(t-s)}\left(\frac{D(t)}{D(s)}\right)^{1-\gamma} dt\right] \\ &= \frac{1}{D(s)}E^{h^*(\theta_i)}\left[\int_s^\infty \frac{u_c(D(t),t)}{u_c(D(s),s)}D(t)\,dt\right], \end{aligned} \quad (26)$$

where $E^{h^*(\theta_i)}[\cdot]$ denotes the expectation under a geometric Brownian motion process for D having drift $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$.

A high C_i implies that the representative investor is willing to pay a high price for the ambiguous state $\Xi(\theta_i)$. Since she cannot exactly identify a particular state, the price of the risky asset is a weighted sum, in which C_i is weighted by the posterior probability of the neighborhood $\Xi(\theta_i)$. C_i is a function of both ambiguity aversion, via the parameter $\eta(\theta_i)$, and the relative risk aversion γ.

Remark 1. *We can also describe Equation* (26) *as*

$$C_i = E\left[\int_s^\infty e^{-\left[\delta+(1-\gamma)\sqrt{2\eta(\theta_i)}\sigma_D\right](t-s)}\left(\frac{D(t)}{D(s)}\right)^{1-\gamma} dt|\theta=\theta_i\right], \quad (27)$$

where $E[\cdot|\theta=\theta_i]$ *denotes the reference model expectation conditional on a constant drift* $\theta=\theta_i$*. Therefore, the impact of ambiguity aversion on the price of equity in the ambiguous state* $\Xi(\theta_i)$ *is equivalent to the impact of an adjusted time preference rate*

$$\delta \longrightarrow \delta + (1-\gamma)\sqrt{2\eta(\theta_i)}\sigma_D. \quad (28)$$

The adjustment depends on the amount of ambiguity of neighborhood $\Xi(\theta_i)$*, risk aversion* γ*, and dividend-growth volatility* σ_D*. Cagetti, Hansen, Sargent, and Williams (2002) give numerical evidence that in their production economy ambiguity aversion decreases the aggregate capital stock in a way that is similar, but not identical under general power utility, to the effect of an increased subjective discount rate. In our model, when ambiguity is homogeneous across neighborhoods, that is,* $\eta(\theta_i)=\eta(\theta_j)$ *for all* i *and* j*, its effect on the price of equity is identical to the impact of an increased time preference rate. In the general case of a heterogeneous* *degree of ambiguity, the final effect on asset prices cannot be mapped into an adjustment of the time preference rate.*

In Corollary 1, we summarize the dependence of the price C_i of an ambiguous neighborhood $\Xi(\theta_i)$ on the ambiguity parameter $\eta(\theta_i)$.

Corollary 1. *The price of any ambiguous state $\Xi\ (\theta_i)$ is a decreasing function of the degree of ambiguity $\eta\ (\theta_i)$ if and only if $\gamma < 1$. In such a case, C_i is a convex function of $\eta\ (\theta_i)$, which is more convex for smaller risk aversion γ.*

From Corollary 1, the marginal relative price of ambiguity is negative if and only if the relative risk aversion γ is less than 1. Hence, a risk aversion coefficient γ above 1 contradicts the basic intuition provided by the Ellsberg (1961) paradox. To give an intuition for this finding, recall that in our setting with power utility $1/\gamma$ is the EIS. When $1/\gamma > 1$, the negative pressure on prices due to the lower expected growth rate of consumption under ambiguity aversion is compensated by two effects, a positive hedging demand for risky assets and a positive saving demand needed to smooth consumption over time. Since only for $1/\gamma > 1$ we can get both high equity premia and low interest rates, we make the following assumption.[6]

Assumption 3. *The representative investor has a relative risk aversion parameter below 1 ($\gamma < 1$) or, equivalently, an EIS above 1 ($1/\gamma > 1$).*

There is mixed empirical evidence on the size of the EIS. However, our Assumption **3** is supported by some of the recent empirical studies. Vissing-Jorgensen (2002) estimates an EIS above 1 when using Euler equations for Treasury bills in a model with time-additive preferences. For the United Kingdom, Attanasio, Banks, and Tanner (2002) estimate an EIS between 0.81 and 2.54 for "likely" shareholders. Vissing-Jorgensen and Attanasio (2003) estimate a model with Epstein-Zin preferences using after-tax returns and find an EIS parameter between 1.03 and 1.43 for U.S. stockholders. In a model with long-run risk and fluctuating economic uncertainty, Bansal and Yaron (2004) use a calibrated EIS parameter of 1.5 to justify several empirical features of asset returns.

[6] There is some experimental evidence favoring low risk aversion under ambiguity, collected by Wakker and Deneffe (1996), who estimate a virtually linear utility function when using a utility elicitation procedure robust to the presence of ambiguity. In these experiments, utility functions estimated by procedures that are not robust to the presence of ambiguity are clearly concave.

3.3 Price-dividend ratios and interest rates

Under Assumption **3**, we obtain from Proposition 1 a few direct implications for the behavior of the price-dividend ratio.

Corollary 2. *Let Assumption **3** be satisfied. It then follows:*

- *(a) The price-dividend ratio P/D is a decreasing convex function of the amount of ambiguity $(\eta(\theta_1), .., \eta(\theta_n))$. Moreover, P/D is a more convex function for lower risk aversion γ.*
- *(b) A mean-preserving spread $\widehat{\Pi}$ of Π implies $\widehat{P}/D > P/D$. Therefore, the price-dividend ratio is increasing in the amount of uncertainty.*

Finding *(a)* is a direct implication of (22) and (27). Finding *(b)* follows from the convexity of C_i in Equation(22) as a function of $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$. From Corollary 2, we see that the impact of a higher ambiguity on P/D ratios and of a higher uncertainty in the economy differ in signs, implying a distinct prediction of ambiguity aversion for the behavior of P/D.

Table 2 illustrates the impact of ambiguity aversion on P/D ratios under no learning (column NL) and learning (column L), for risk-aversion parameters between 0.6 and 0.9. These levels of risk aversion correspond to EIS parameters between 1.1 and 1.7. These EIS parameters are consistent with the empirical findings in the literature cited above.

In Table 2, we compute P/D ratios from Equation (22) under three possible reference model drifts, $\theta_1 = 0.005$, $\theta_2 = 0.0175$, $\theta_3 = 0.03$, and a discrete prior Π centered at θ_2. To fix the prior dispersion, we set θ_1, and $\theta_3 = \theta_2 \pm \sigma_e$, and use a signal standard deviation $\sigma_e = 0.0125$. Under these conditions, the posterior mean for expected consumption growth is $m_\theta = 0.0175$ and coincides with the U.S. average consumption growth estimated by Campbell (1999) for an annual sample spanning the period

Table 2
Price-dividend ratios

	$\gamma = 0.6$		$\gamma = 0.7$		$\gamma = 0.8$		$\gamma = 0.9$	
η	NL	L	NL	L	NL	L	NL	L
0.0000	35.55	36.27	33.49	33.82	31.66	31.78	30.03	30.06
0.0001	35.32	36.02	33.33	33.66	31.57	31.69	29.99	30.02
0.0010	34.83	35.51	33.01	33.32	31.37	31.49	29.90	29.93
0.0025	34.43	35.08	32.73	33.04	31.21	31.32	29.83	29.85
0.0050	33.98	34.61	32.43	32.73	31.02	31.14	29.74	29.77
0.0075	33.65	34.25	32.20	32.49	30.88	31.00	29.68	29.70
0.0100	33.37	33.96	32.01	32.30	30.77	30.88	29.62	29.65

The entries report P/D ratios for different risk aversion coefficients γ and different homogeneous levels of ambiguity η, in model settings without learning (NL) and with learning (L). Calculations are based on a model with three reference model drift states $\Theta = \{0.005, 0.0175, 0.03\}$ with prior probabilities $\Pi = \{0.266, 0.468, 0.266\}$. The true dividend drift is $\theta = 0.0175$. Further parameters are $\delta = 0.035$, $\sigma_D = 0.0325$, $\sigma_e = 0.0125$.

from 1891 to 1994. Again, we fix the volatility of consumption growth at $\sigma_D = 0.0325$.

With this prior and this parameter structure, we compute the model-implied P/D ratios by using the explicit formula (22) in Proposition 1. The first row ($\eta = 0$) in Table 2 presents the quantities that prevail in the absence of ambiguity. The entries in the following rows for η ranging from 0.0001 to 0.01 report the P/D ratios for an increasing (homogeneous) ambiguity parameter. We see that the P/D ratios decrease with risk aversion. owing to the convexity of C_i as a function of $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$, these ratios are higher in a setting of learning. Ambiguity aversion lowers P/D ratios monotonically, and the size of this decrease depends on the investor's risk aversion. For the model with learning and for $\gamma = 0.6$, the P/D ratios decrease from 36.27 to 33.96 for $\eta = 0$ and $\eta = 0.01$. For $\gamma = 0.9$, the corresponding P/D ratio decreases very little, from 30.06 to 29.65.

Since the risk-free rate r is a decreasing convex function of $(\eta(\theta_1), .., \eta(\theta_n))$, learning and ambiguity aversion reduce the equilibrium interest rate. We obtain the interest rate in the Bayesian setting of Veronesi (2000) for $\eta(\theta) = 0$:

$$r = \delta + \gamma m_\theta - \frac{1}{2}\gamma(\gamma+1)\sigma_D^2. \tag{29}$$

The interest rate under ambiguity but no learning arises for a degenerate distribution Π. In this case, we have $m_\theta + m_{h^*(\theta)} = \theta_l - \sqrt{2\eta(\theta_l)}\sigma_D$ for some $\theta_l \in \Theta$, and the interest rate is equal to

$$r = \delta + \gamma\left(\theta_l - \sqrt{2\eta(\theta_l)}\sigma_D\right) - \frac{1}{2}\gamma(\gamma+1)\sigma_D^2. \tag{30}$$

When Π is nondegenerate, the contribution of ambiguity aversion to the interest rate is summarized by the term $m_{h^*(\theta)}$, which is a weighted sum of the contributions in the single-model neighborhoods $\Xi(\theta_1), \ldots, \Xi(\theta_n)$. The weights in $m_{h^*(\theta)}$ are the posterior probabilities Π.

4. Conditional Asset Returns

Given the worst-case dividend drift $\theta - \sqrt{2\eta(\theta)}\sigma_D$ conditional on the ambiguous state $\Xi(\theta)$, we obtain the dynamics of the equilibrium equity excess return R defined by

$$dR = \frac{dP + Ddt}{P} - rdt. \tag{31}$$

Proposition 2. *(i) Under the investor's subjective optimal worst-case belief $h^*(\theta)$ in Proposition 1, the equilibrium excess return satisfies*

$$dR = \mu_R^{wc}dt + \sigma_D d\widetilde{B}_D^{h^*} + V_{\theta,h^*}\left(k_D d\widetilde{B}_D^{h^*} + k_e d\widetilde{B}_e^{h^*}\right), \tag{32}$$

where

$$\mu_R^{wc} = \gamma\left(\sigma_D^2 + V_{\theta,h^*}\right), \quad V_{\theta,h^*} = \sum_{i=1}^{n} \frac{\pi_i C_i\left(\theta_i - \sqrt{2\eta\,(\theta_i)}\sigma_D\right)}{\sum_{i=1}^{n}\pi_i C_i} - m_{\theta,h^*}, \tag{33}$$

with Brownian motion increments with respect to the filtration $\{\mathcal{F}(t)\}$ *given by*

$$d\widetilde{B}_D^{h^*} = k_D\left(\frac{dD}{D} - m_{\theta,h^*}dt\right), \quad d\widetilde{B}_e^{h^*} = k_e\left(de - m_{\theta,h^*}dt\right).$$

(ii) Under the reference-model belief, the equilibrium excess return satisfies

$$dR = \mu_R dt + \sigma_D d\widetilde{B}_D + V_{\theta,h^*}\left(k_D d\widetilde{B}_D + k_e d\widetilde{B}_e\right), \tag{34}$$

where

$$\mu_R = \mu_R^{wc} - m_{h^*(\theta)}\left(1 + kV_{\theta,h^*}\right), \quad k = k_D^2 + k_e^2, \tag{35}$$

with Brownian-motion increments with respect to the filtration $\{\mathcal{F}(t)\}$ *given by*

$$d\widetilde{B}_D = k_D\left(\frac{dD}{D} - m_\theta dt\right), \quad d\widetilde{B}_e = k_e\left(de - m_\theta dt\right).$$

By using the first part of Proposition 2, we can study the first effect of learning and ambiguity aversion on excess returns. The description of R with respect to the Brownian motions $\tilde{B}_D^{h^*}$ and $\tilde{B}_e^{h^*}$ yields the dynamics under the worst-case scenario $h^*(\theta) \in \Xi(\theta)$. Thus, we can interpret the resulting expected excess return on equity as the worst-case equity premium.

The second part of Proposition 2 allows us to study the second impact of learning and ambiguity aversion on excess returns. This impact arises because of the discrepancy between the likelihood belief implied by the reference model and the optimal worst-case belief that the representative investor adopts to compute asset prices. This discrepancy generates an additional premium for ambiguity. We can analyze this second effect of learning and ambiguity aversion by describing excess returns with respect to the filtered Brownian motions $\tilde{B}_D, \tilde{B}_e$, which are the optimal ones implied by the reference-model belief. This description yields the correct excess return from the perspective of an outside observer such as an econometrician, who believes in the reference model as an approximate description of dividends and who knows that the representative investor is ambiguity averse.

To analyze excess returns in more detail, we must study the sign and the comparative statics of the quantities $m_{h^*(\theta)}$ and V_{θ,h^*} in Equations (32) and (34). The term

$$m_{h^*(\theta)} = m_{\theta,h^*} - m_\theta = -\sum_{i=1}^{n} \sqrt{2\eta(\theta_i)}\pi_i \sigma_D \tag{36}$$

is a correction to the reference model's posterior mean m_θ. This correction is negative and incorporates misspecification doubts in the posterior mean of the economy's growth rate. V_{θ,h^*} reflects the difference between the worst-case expected growth rate of the economy, m_{θ,h^*}, and its value-adjusted counterpart. This quantity is larger when the representative investor has more diffuse beliefs about $\Xi(\theta_1), .., \Xi(\theta_n)$ or when she values the asset very differently across states. Differences in valuation across states depend on the heterogeneity of the worst-case growth rate $\theta - \sqrt{2\eta(\theta)}\sigma_D$. Under Assumption **2**, it follows that:

$$\theta_1 - \sqrt{2\eta(\theta_1)}\sigma_D < \theta_2 - \sqrt{2\eta(\theta_2)}\sigma_D < \ldots < \theta_n - \sqrt{2\eta(\theta_n)}\sigma_D.$$

Therefore, we can apply the arguments in the proof of Lemma 3 in Veronesi (2000) to characterize the behavior of V_{θ,h^*}.

Lemma 2. *Let Assumption **2** be satisfied. It then follows:*

1. *V_{θ,h^*} is a decreasing function of γ.*
2. *The following statements are equivalent:*
   - *(a) Assumption **3** holds.*
   - *(b) $V_{\theta,h^*} > 0$.*
   - *(c) For any mean-preserving spread $\widetilde{\Pi}$ of Π it follows that*

     $$\widetilde{V}_{\theta,h^*} > V_{\theta,h^*},$$

     where "$\backsim$" denotes quantities under $\widetilde{\Pi}$.

In particular, V_{θ,h^*} is positive and increasing for mean-preserving spreads of Π if and only if $\gamma < 1$. The positivity of V_{θ,h^*} is crucial for avoiding asset prices that are inconsistent with the equity premium puzzle predictions.

4.1 Equity premia

From Equation (33), we see that the equity premium is the sum of three conceptually different components:

$$\mu_R = \underbrace{\gamma(\sigma_D^2 + V_{\theta,h^*})}_{\text{(A)}} \quad \underbrace{-m_{h^*}}_{\text{(B)}} \quad \underbrace{-m_{h^*} k V_{\theta,h^*}}_{\text{(C)}}. \tag{37}$$

Component (A) is both the equity premium for the standard risk exposure, that is, the risk premium, and the worst-case equity premium in our economy. The sum (B) + (C) is the equity premium caused by ambiguity, that is, the ambiguity premium. (B) is the ambiguity premium caused by misspecification doubts about the dynamics of dividends. (C) is the ambiguity premium caused by misspecification doubts about the dynamics of the optimal posterior probabilities Π.

In Figure 1, we compute the risk and equity premia under the same reference model drifts, prior structure, and parameter values as in Table 2. We plot these drifts and priors in Panel A of Figure 1. In Panel B, we also plot the sizes of the ambiguous neighborhoods $\Xi(\theta_1), .., \Xi(\theta_n)$ as ellipses centered at $\theta_1, .., \theta_n$. In Panel D of Figure 1, we present a typical pattern of the equity premium μ_R implied by Equation (35) for a risk-aversion parameter between 0.7 and 1 and for a homogeneous degree of ambiguity $\eta = 0.001$. The equity premium is a monotonically decreasing function of risk aversion. For some risk aversions strictly smaller than 1, this feature turns out to be consistent with the predictions of the equity premium puzzle. For instance, for a risk aversion between 0.7 and 0.8 the equity premium ranges from 12.5% to 4.8%. These substantial premia arise despite the very small size of the ambiguity parameter used. Following our discussion in Section 2.3, the difference between the reference and the worst-case belief is only 0.16%, which implies a worst-case expected consumption growth of

$$m_{\theta,h^*} = m_\theta - \sqrt{2\eta}\sigma_D = 1.75\% - 0.16\% = 1.59\%.$$

Furthermore, for $\eta = 0.001$ and the sample size used by Campbell (1999), the model detection error probability defined in Equation (9) is $\epsilon_N(\theta) = 0.41$, that is, the probability that the investor would not be able to separate statistically the reference and the worst-case model is 41%.

4.1.1 Premia for risk. Component (A) in Equation (37) is the equity premium perceived by an investor under the optimal worst-case likelihood $h^*(\theta)$ in Proposition 1. From Proposition 2, we get

$$\mu_R^{wc} = \gamma\left(\sigma_D^2 + V_{\theta,h^*}\right) = \gamma Cov_t^{h*}(dR, dD/D) = \gamma Cov_t(dR, dD/D), \tag{38}$$

where $Cov_t^{h^*}$ and Cov_t denote conditional covariances under the worst-case likelihood h^* and under the reference model likelihood, respectively. The last equality in Equation (38) arises because worst-case and reference model likelihoods are absolutely continuous. Hence, the covariance $Cov_t(dR, dD/D)$ between returns and aggregate consumption growth is linked only to the fraction (A) of the whole equity premium μ_R.

In our economy, the risk premium (A) is quite small. $\gamma\sigma_D^2$ is the risk premium under ambiguity aversion but no learning.[7] For realistic risk-aversion parameters it is typically small. Moreover, Lemma 2 implies

[7] See Maenhout (2004) and Trojani and Vanini (2002).

that V_{θ,h^*} is decreasing in risk aversion. Therefore, the risk premium (A) is bounded as a function of risk aversion and is negligible for practical purposes, as in Veronesi (2000).

In Panel C of Figure 1, we plot a typical profile of the risk premium (A) (solid line) together with the risk premium implied by the standard Bayesian learning setting (dashed line). The two risk premia almost coincide. In general, the risk premium under learning and ambiguity aversion is different, because the term V_{θ,h^*} in Equation (38) is different from the one in a setting of pure learning. But for realistic structures of the function $\eta(\theta)$, we always find that the two risk premia are numerically similar and small. For $\gamma = 0.6$, the risk premia in the models with and without ambiguity aversion are 0.57% and 0.59%, respectively. As emphasized by Veronesi (2000), the Mehra and Prescott (1985) equity

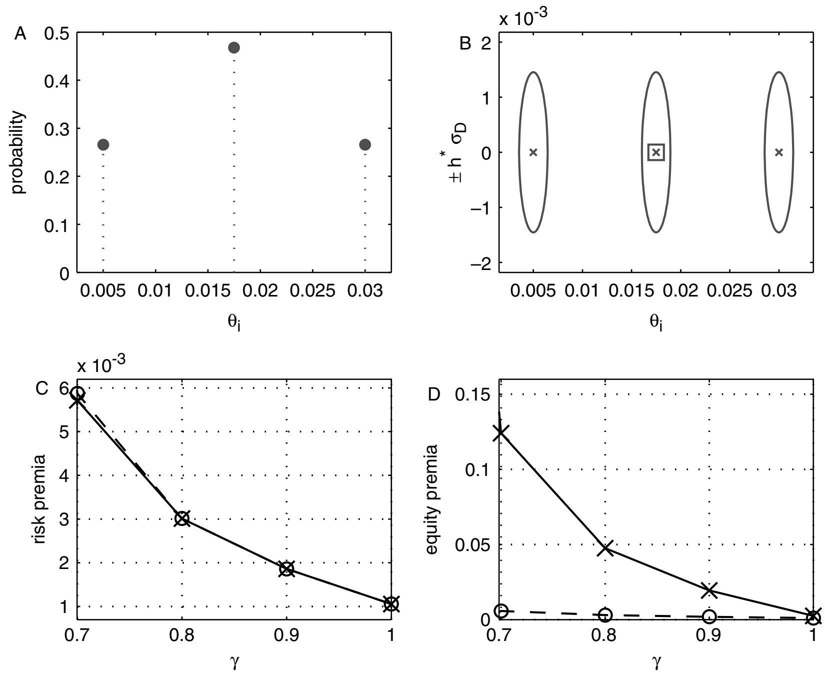

Figure 1
Risk premium and ambiguity premium
Panel A plots the set of probabilities Π relevant for the figure. Panel B plots the different relevant reference model states $\theta_1, .., \theta_n$. The true dividend drift state of the reference model is marked with a square and is equal to the posterior expected value $\sum \pi_i \theta_i$. Calculations are based on a model with three reference model drift states $\Theta = \{0.005,\ 0.0175,\ 0.03\}$ with prior probabilities $\Pi = \{0.266,\ 0.468,\ 0.266\}$. We use a small amount of homogeneous ambiguity $\eta = 0.001$. In Panel B, the size of the ambiguous neighborhoods $\Xi(\theta_1), .., \Xi(\theta_n)$ is highlighted by the ellipses centered at $\theta_1, .., \theta_n$. Further, we set $\delta = 0.01$, $\sigma_D = 0.0325$, and $\sigma_e = 0.0125$. With this prior structure and these parameters, Panels C and D plot the resulting risk premium μ_R^{wc} and the equity premium μ_R as functions of γ for ambiguity aversion (solid line) and no ambiguity aversion (dashed line).

premium puzzle is even more puzzling in a Bayesian setting, because equity premia cannot be matched by risk premia, even for very high risk aversion. This finding on risk premia is also true for our model. However, in our model, equity premia consist of premia for both risk and ambiguity.

4.1.2 Premia for ambiguity. The equity premium in Equation (37) depends on the ambiguity premia (B) and (C), both of which are positive under Assumption **3**. The premium (B) + (C) is due to the discrepancy between the reference-model belief and the worst-case belief used by the representative investor to update her subjective posterior mean for the growth rate of dividends. The resulting worst-case return dynamics in Equation(32) depend on the two filtered random shocks,

$$d\widetilde{B}_D^{h^*} = d\widetilde{B}_D - k_D m_{h^*(\theta)} dt, \quad d\widetilde{B}_e^{h^*} = d\widetilde{B}_e - k_e m_{h^*(\theta)} dt, \tag{39}$$

which define a $\{\mathcal{F}(t)\}$–Brownian motion $(\widetilde{B}_D^{h^*}, \widetilde{B}_e^{h^*})$ under the worst-case belief $h^*(\theta)$, and define a $\{\mathcal{F}(t)\}$–Brownian motion with drift under the reference-model belief. We can interpret the discrepancies

$$\lambda_D^A := (d\widetilde{B}_D^{h^*} - d\widetilde{B}_D)/dt = -k_D m_{h^*(\theta)}, \tag{40}$$

and

$$\lambda_e^A := (d\widetilde{B}_e^{h^*} - d\widetilde{B}_e)/dt = -k_e m_{h^*(\theta)}, \tag{41}$$

as the market prices of ambiguity for $d\widetilde{B}_D$ and $d\widetilde{B}_e$ shocks. The filtered shocks $d\widetilde{B}_D^{h^*}$ and $d\widetilde{B}_e^{h^*}$ influence the worst-case dynamics in Equation (32) in two ways: through the impact of $d\widetilde{B}_D^{h^*}$ on the filtered dynamics for dividends and through the joint impact of $d\widetilde{B}_D^{h^*}$ and $d\widetilde{B}_e^{h^*}$ on the filtered dynamics for Π. Let $h = h^*$ in Equations (18) and (19). It then follows:

$$dD/D = m_{\theta,h^*} dt + \sigma_D d\widetilde{B}_D^{h^*}, \tag{42}$$

$$d\pi_i/\pi_i = \left(\theta_i + h(\theta_i)\sigma_D - m_{\theta,h^*}\right)\left(k_D d\widetilde{B}_D^{h^*} + k_e d\widetilde{B}_e^{h^*}\right), \quad i = 1, .., n. \tag{43}$$

The ambiguity premium deriving from a unit shock in the filtered dynamics in Equation(43) is

$$k_D(d\widetilde{B}_D^{h^*} - d\widetilde{B}_D)/dt + k_e(d\widetilde{B}_e^{h^*} - d\widetilde{B}_e)/dt = k_D\lambda_D^A + k_e\lambda_e^A = -k m_{h^*(\theta)}, \tag{44}$$

where $k = k_D^2 + k_e^2$. According to the worst-case dynamics in Equation(32), V_{θ,h^*} measures the reaction of equilibrium equity returns to shocks in the dynamics in Equation(43). Therefore, component (C) in Equation (37) is the ambiguity premium for the misspecification of these shocks. The ambiguity premium for a shock in the filtered dividend dynamics is given in Equation(40). Therefore, component (B) in Equation (37) is the ambiguity premium for misspecified dividend shocks.

Component (B) is positive under a degenerate Π−distribution, that is, in the absence of learning. We can interpret it as a pure premium for ambiguity. In contrast, the ambiguity premium (C) is nonzero if and only if Π is nondegenerate, that is, if the representative investor has only partial information about the underlying model neighborhood $\Xi(\theta)$. Therefore, we can interpret (C) as an ambiguity premium component caused by the interaction of learning and ambiguity aversion.

To illustrate the contribution of learning and ambiguity to equity premia, we compute these quantities in Table 3 for a setting of no learning (column NL) and a setting of learning (column L). The first row ($\eta = 0$) in Table 3 presents the quantities that prevail in the absence of ambiguity. The remaining rows show the equilibrium risk and equity premia (column RP and EP) for an increasing (homogeneous) ambiguity parameter η and for risk-aversion parameters $\gamma = 0.8$ (Panel A), $\gamma = 0.85$ (Panel B), and $\gamma = 0.9$ (Panel C). The reference belief dividend drifts, prior structure, and all other model parameters are the same as those used for Figure 1.

In columns RP of Table 3, we obtain small risk premia for all levels of risk aversion and for a setting both with and without learning. The risk premia μ_R^{wc} range between 0.20% and 0.33%.

In column EP, we present the equity premia. In the absence of ambiguity ($\eta = 0$), the risk and equity premia are identical. Introducing ambiguity into the model increases equity premia for both the setting of no learning (column NL) and of learning (column L). In column NL of Panel B with $\gamma = 0.85$, we observe a premium for pure ambiguity that increases the equity premia from 0.09% ($\eta = 0$) to 0.55% ($\eta = 0.01$). We can attribute this increase to the pure ambiguity premium (B), which amounts to 0.44% at $\eta = 0.01$. In relative terms, this increase is substantial. However, the resulting equity premium is still too small for practical purposes.

In contrast, in Column L of Panel B we present large equity premia that increase from 0.26% ($\eta = 0$) to 7.44% ($\eta = 0.01$) due to a large ambiguity premium (B) + (C). Since the ambiguity premium (B) is only 0.44%, the large increase is due to premium component (C) arising from the joint presence of learning and ambiguity aversion. In Panel B for $\eta = 0.01$, this component amounts to 6.89%.

For $\eta = 0.0075$ and $\eta = 0.01$, equilibrium interest rates are close to 2% for all levels of risk aversion. The low interest is due to both ambiguity aversion and the low risk aversion. The latter induces a sufficiently strong desire to substitute consumption intertemporally, which additionally helps to obtain the low interest rate.

Table 3
Equilibrium interest rates, risk premia, equity premia, and volatilities

Panel A: $\gamma = 0.8$

η	IR (%)	RP (%) NL	RP (%) L	EP (%) NL	EP (%) L	Vol (%) NL	Vol (%) L
0.0000	2.32	0.08	0.33	0.08	0.33	3.25	27.90
0.0001	2.29	0.08	0.33	0.13	1.43	3.25	27.92
0.0010	2.21	0.08	0.33	0.23	3.79	3.25	27.94
0.0025	2.14	0.08	0.33	0.31	5.80	3.25	27.94
0.0050	2.06	0.08	0.33	0.41	8.07	3.25	27.93
0.0075	2.01	0.08	0.33	0.48	9.80	3.25	27.91
0.0100	1.96	0.08	0.33	0.54	11.25	3.25	27.88

Panel B: $\gamma = 0.85$

η	IR (%)	RP (%) NL	RP (%) L	EP (%) NL	EP (%) L	Vol (%) NL	Vol (%) L
0.0000	2.40	0.09	0.26	0.09	0.26	3.25	18.49
0.0001	2.37	0.09	0.26	0.14	0.98	3.25	18.49
0.0010	2.28	0.09	0.26	0.24	2.53	3.25	18.49
0.0025	2.21	0.09	0.26	0.32	3.85	3.25	18.49
0.0050	2.13	0.09	0.26	0.41	5.34	3.25	18.49
0.0075	2.07	0.09	0.26	0.49	6.48	3.25	18.48
0.0100	2.01	0.09	0.26	0.55	7.44	3.25	18.47

Panel C: $\gamma = 0.9$

η	IR (%)	RP (%) NL	RP (%) L	EP (%) NL	EP (%) L	Vol (%) NL	Vol (%) L
0.0000	2.48	0.10	0.20	0.10	0.20	3.25	11.62
0.0001	2.44	0.10	0.20	0.14	0.64	3.25	11.62
0.0010	2.35	0.10	0.20	0.24	1.60	3.25	11.62
0.0025	2.28	0.10	0.20	0.32	2.41	3.25	11.62
0.0050	2.19	0.10	0.20	0.42	3.32	3.25	11.62
0.0075	2.13	0.10	0.20	0.49	4.03	3.25	11.61
0.0100	2.07	0.10	0.20	0.55	4.62	3.25	11.61

The entries report the equilibrium interest rates (IR), risk premia (RP), equity premia (EP), and volatilities (Vol) for risk aversions $\gamma = 0.8$ (Panel A), $\gamma = 0.85$ (Panel B), and $\gamma = 0.9$ (Panel C) and for different homogeneous levels of ambiguity η, under models without learning (NL) and with learning (L). Calculations are based on a model with three reference model drift states $\Theta = \{0.005, 0.0175, 0.03\}$ with prior probabilities $\Pi = \{0.266, 0.468, 0.266\}$. The true dividend is $\theta = 0.0175$. Further parameters are $\delta = 0.01$, $\sigma_D = 0.0325$, $\sigma_e = 0.0125$.

Table 4 shows that the equity premium and the interest rates obtained in Table 3 have sizes comparable to those of the interest rate and equity premium estimated in Campbell (1999). The historical average interest rate in Table 4 is about 2% (see column IR), and the equity premium is about 6.26% (see column EP).

As noted above, component (B) is a pure premium for ambiguity. In the absence of learning, we could make this component arbitrarily large

Table 4
Estimated interest rate and returns parameters

IR (%)	EP (%)	Vol (%)
1.95%	6.26%	18.5%

The table reports the interest rate, equity premium, and volatility parameters, estimated in Campbell (1999) for an annual sample spanning the period from 1891 to 1994.

by increasing the parameter η to fit the stylized facts of asset prices. However, to obtain an empirically reasonable equity premium, we would need to impose an excessive degree of ambiguity by making the parameter η unreasonably large. In a model with ambiguity aversion only, we would have to set $\eta = 1.8$ to obtain an equity premium around 6.28%. This ambiguity level implies a rather pessimistic worst-case expected consumption growth $m_{\theta,h^*} = -4.42\%$ and yields a model detection error probability equal to zero. Therefore, only by simultaneously modeling both learning and ambiguity aversion, we can generate a substantial equity premium with a more reasonable amount of ambiguity. The finding that the pure ambiguity premium is small for realistic sizes of ambiguity is consistent with the calibrations in Maenhout (2004), who does not assume any learning and adopts a max-min optimization problem different from ours.

4.1.3 Is there an ambiguity premium for imprecise signals?. Consistent with the findings of Veronesi (2000), there is no quantitatively relevant risk premium for imprecise signals. However, under ambiguity aversion the key equity premium component is the ambiguity premium (C). Therefore, we ask under which conditions an ambiguity premium for imprecise signals exists.

Corollary 3. *Let Assumption 3 be satisfied and suppose that function $\sqrt{\eta(\theta)}$ is a convex function of θ. Then, for any mean-preserving spread $\tilde{\Pi}$ of Π it follows that:*

$$-\tilde{m}_{h^*}(1 + k\tilde{V}_{\theta,h^*}) > -m_{h^*}(1 + kV_{\theta,h^*}), \tag{45}$$

where $\sim$ denotes quantities under $\tilde{\Pi}$.

Corollary 3 states that under a convex function $\sqrt{\eta(\theta)}$, there is always an ambiguity premium for information noisiness. When the realized signal precision is low, the posterior probabilities Π are more diffuse and the quantity V_{θ,h^*} in Equation (45) increases (see again Lemma 2). We can write V_{θ,h^*} as the covariance between equity excess returns R and signals

e. Therefore, the increased ambiguity premium under a less precise signal follows from the higher covariance between equity returns and public signals. This covariance amplifies the impact of the ambiguity premium (C) on the total equity premium. The higher covariance between returns and signals arises because a less precise dividend drift prediction m_{θ,h^*} implies a lower sensitivity of the investor's hedging demand to signals. In that case, positive (negative) signals are more directly linked to a positive (negative) excess demand for equity and imply a higher covariance between equilibrium equity returns and signals.

4.2 Equity volatility and Risk/Return relations

From Proposition 2, the variance of stock returns is

$$\sigma_R^2 = \sigma_D^2 + V_{\theta,h^*}\left(2 + kV_{\theta,h^*}\right). \tag{46}$$

The variance is increasing in the absolute value of V_{θ,h^*} and for $V_{\theta,h^*} = 0$ attains the minimum σ_D^2. As in Veronesi (2000), to obtain empirically relevant volatilities it is crucial to introduce learning into the model. However, we know from Section 4.1 that the only parameter choice consistent with sizable equity premia is $\gamma < 1$.

Column Vol of Table 3 illustrates quantitatively the contribution of learning and ambiguity to equity return volatilities, for both the setting of no learning (column NL) and learning (column L). In the absence of learning, we obtain small equity volatilities of 3.25%. In the presence of learning and for $\gamma = 0.85$ (Panel B), volatility ranges from 18.49%, without ambiguity ($\eta = 0$), to 18.47%, for an ambiguity parameter $\eta = 0.01$. These volatility levels correspond to the 18.5% unconditional volatility of returns estimated in Table 4 (see column Vol).

The ambiguity premium is the main determinant of the level and the time variation of the equity premium in our model. This feature strongly affects the equilibrium relation between expected excess returns and conditional variances. From Proposition 2 and Equation (46), we see that the relations between risk or equity premia and the conditional variance of returns are given by

$$\mu_R^{wc} = \gamma\sigma_R^2 - \gamma V_{\theta,h^*}\left(1 + kV_{\theta,h^*}\right) \tag{47}$$

and

$$\mu_R = \left(\gamma - \frac{m_{h^*(\theta)}}{V_{\theta,h^*}}\right)\sigma_R^2 - \gamma V_{\theta,h^*}\left(1 + kV_{\theta,h^*}\right) + m_{h^*(\theta)}\left(1 + \frac{\sigma_D}{V_{\theta,h^*}}\right). \tag{48}$$

Equation (48) implies a truly positive, but time-varying, theoretical relation between the equity premium μ_R and the conditional variance σ_R^2. This time-varying relation is due to the ambiguity premium and results from the interaction of learning and ambiguity aversion. The true relation between the risk premium μ_R^{wc} and the conditional variance σ_R^2 is linear and constant. Both relations (47) and (48) are biased by a heteroscedastic error term that has a nonzero conditional mean.

Figure 2 illustrates the theoretical relation between risk premia, equity premia, and conditional variances of equity returns. Using the same initial prior and the model parameters of Figure 1, we simulate a daily time series of posterior probabilities $\Pi(t)$ and compute the resulting theoretical (time-varying) coefficient

$$b_t = \gamma - m_{h^*}(\theta)/V_{\theta,h^*}, \tag{49}$$

from the expected excess return and variance relation in Equation (48). The time variation in b_t is caused exclusively by ambiguity. Without ambiguity it follows that $b_t = \gamma$ is constant.

The theoretical (time-varying) equity premium "sensitivity" b_t to changes in σ_R^2 is large compared to the sensitivity of the risk premium, which is just the risk aversion coefficient γ. By definition, ambiguity premia derive from model misspecification, not from covariances between asset

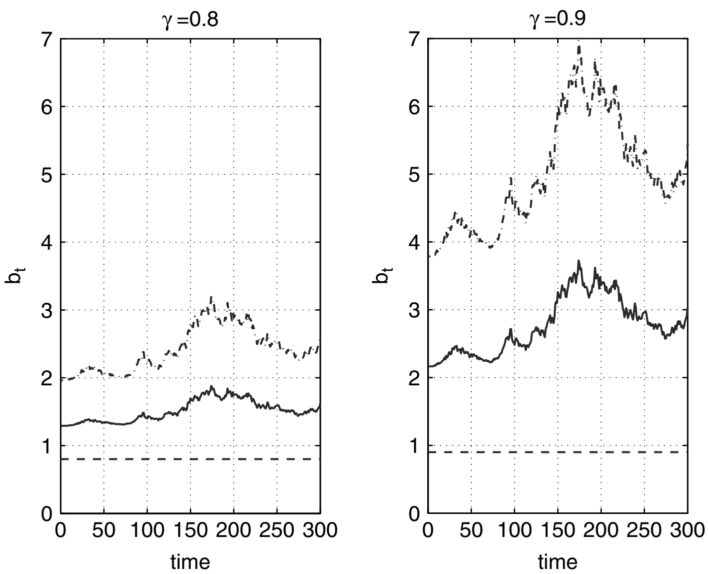

Figure 2
Theoretical time-varying return/variance trade-off
For different parameters $\gamma = 0.8, 0.9$, we simulate the theoretical (time-varying) coefficient b_t in Equation (49). In all panels, we plot b_t as a function of time for homogeneous ambiguity structures $\eta = 0.001$ (solid line) and 0.005 (dash-dotted line). The dashed line represents $b_t = \gamma$. The three reference model drift states are $\Theta = \{0.005,\ 0.0175,\ 0.03\}$ with initial prior probabilities $\Pi = \{0.266,\ 0.468,\ 0.266\}$. The true dividend drift is $\theta = 0.0175$. The remaining parameters are $\delta = 0.01$, $\sigma_D = 0.0325$, $\sigma_e = 0.0125$.

returns and aggregate consumption growth. Therefore, we can expect them to be difficult to identify by, for instance, regression methods. Figure 3 highlights this point by plotting the time series of estimated parameters for a sequence of rolling regressions of R on the theoretical value of σ_R^2 defined by Equation (46). We do these regressions for settings with and without ambiguity. To compute the time series of R and σ_R^2 and the corresponding linear regression estimates, we use the same simulated trajectories of $\Pi(t)$ as in Figure 2. To estimate the rolling regression coefficients, we use samples of 180 daily observations in each regression.

In Figure 3, highly time-varying regression estimates arise. The estimated regression parameters for settings with and without ambiguity are hardly distinguishable. The estimates indicate that the estimated relation between μ_R and σ_R^2 over different time periods can even switch signs. The estimated coefficients clearly fail to identify the theoretical coefficient b_t in Equation (49). This is the case for the settings both with and without ambiguity. However, by comparing with the results in Figure 2, the bias in the presence of ambiguity aversion is much larger. For instance, the estimated

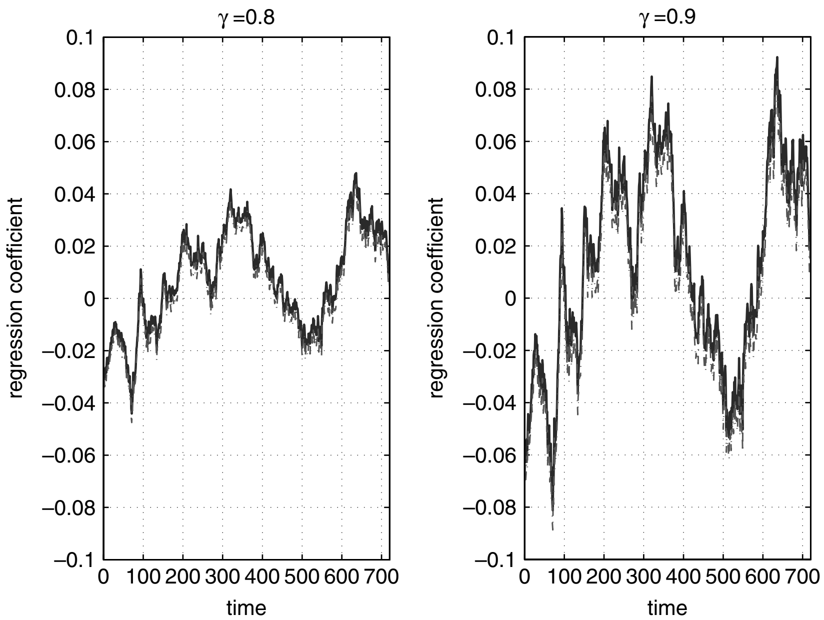

Figure 3
Rolling regression analysis of equity returns on equilibrium variances
For different parameters $\gamma = 0.8, 0.9$, we plot the estimated parameter b in a series of rolling regressions of R on σ_R^2. All panels present the results of one single simulation run. The rolling regressions are based on sample sizes of 180 daily observations, simulated from a model with three reference model drift states $\Theta = \{0.005,\ 0.0175,\ 0.03\}$ with initial probabilities $\Pi = \{0.266,\ 0.468,\ 0.266\}$ and under a homogeneous degree of ambiguity $\eta = 0.001$ (solid line) and $\eta = 0$ (dashed line). The true dividend is $\theta = 0.0175$. The remaining parameters are $\delta = 0.01$, $\sigma_D = 0.0325$, $\sigma_e = 0.0125$.

parameters for $\gamma = 0.9$ are never above 0.1, but the theoretical coefficient b_t in Figure 2 stays above 2 for the ambiguity aversion parameters considered. For the setting without ambiguity aversion, the theoretical coefficient is equal to the risk aversion coefficient $b_t = \gamma = 0.9$.

5. Conclusion

We study asset prices in a continuous-time partial-information Lucas economy with ambiguity aversion. We find that for reasonable parameter values, the joint presence of learning and ambiguity generates a high equity premium, a low interest rate, and excess volatility. Model settings that do not take learning or ambiguity aversion into account require an unrealistically large amount of pessimism to generate sizable equity premia. At the same time, model settings without learning imply small equity volatilities. A further implication of our model is a highly time-varying relation between excess returns and their conditional variances, leading to large time-varying biases in the naively estimated risk-return tradeoff. This result shows that learning and ambiguity aversion can help explaining the weak evidence for a positive relation between expected excess returns and conditional variances in the data.

By using a time-additive power utility function, we can obtain easily interpretable closed-form solutions for the equilibrium, at the cost of constraining the relation between risk aversion and EIS. Disentangling risk aversion and EIS could be useful to introduce an additional degree of freedom in the choice of the risk-aversion parameter, which would make the model more flexible in the fit of, for example higher worst-case equity premia. Such extensions are interesting directions for future research.

Appendix A:

Proof of Proposition 1. We have for any likelihood $h(\theta) \in \Xi(\theta)$,

$$\begin{aligned} V^{h(\theta)}(\Pi, D) &= E^{h(\theta)}\left[\int_s^\infty e^{-\delta(t-s)} \frac{D_t^{1-\gamma}}{1-\gamma} dt \,\middle|\, \mathcal{F}(s)\right] \\ &= E^{h(\theta)}\left[\int_s^\infty e^{-\delta(t-s)} \frac{D(t)^{1-\gamma}}{1-\gamma} dt \,\middle|\, \pi_1(s) = \pi_1, .., \pi_n(s) = \pi_n, D(s) = D\right] \\ &= \frac{D^{1-\gamma}}{1-\gamma} \sum_{i=1}^n \pi_i E^{h(\theta)}\left[\int_s^\infty e^{-\delta(t-s)} \left(\frac{D(t)}{D(s)}\right)^{1-\gamma} dt \,\middle|\, \widetilde{\theta} = \widetilde{\theta}_i\right], \end{aligned} \tag{A1}$$

where $\widetilde{\theta} = \theta + h(\theta)\sigma_D$, $\widetilde{\theta}_i = \theta_i + h(\theta_i)\sigma_D$. Therefore, for any vector Π:

$$
\begin{aligned}
J(\Pi, D) &= \inf_{h(\theta)} V^{h(\theta)}(\Pi, D) \\
&\geq \frac{D^{1-\gamma}}{1-\gamma} \sum_{i=1}^{n} \pi_i \inf_{h(\theta_i)} E^{h(\theta_i)} \left[\int_s^{\infty} e^{-\delta(t-s)} \left(\frac{D(t)}{D(s)} \right)^{1-\gamma} dt \middle| \widetilde{\theta} = \widetilde{\theta}_i \right].
\end{aligned} \tag{A2}
$$

Conditional on $\widetilde{\theta}_i$, the $h(\theta)$ −drift misspecified dynamics are

$$
dD = (\theta_i + h(\theta_i)\sigma_D) D dt + \sigma_D D dB_D. \tag{A3}
$$

Therefore, Assumption **2** implies that we can focus on solving the problem

$$
V^i(D) = \inf_{h(\theta_i)} E\left(\int_s^{\infty} e^{-\delta(t-s)} \frac{D(t)^{1-\gamma}}{1-\gamma} dt \middle| D(s) = D \right),
$$

subject to

$$
\frac{1}{2} h(\theta_i)^2 \leq \eta(\theta_i)
$$

and the dividend dynamics

$$
dD = (\theta_i + h(\theta_i)\sigma_D) D dt + \sigma_D D dB_D. \tag{A4}
$$

The Hamilton Jacobi Bellman equation for this problem reads

$$
0 = \inf_{h(\theta_i)} \left\{ -\delta V^i + \frac{D^{1-\gamma}}{1-\gamma} + (\theta_i + h(\theta_i)\sigma_D) D V_D^i + \frac{1}{2}\sigma_D^2 D^2 V_{DD}^i + \lambda \left(\frac{1}{2} h(\theta_i)^2 - \eta(\theta_i) \right) \right\}, \tag{A5}
$$

where $\lambda \geq 0$ is a Lagrange multiplier for the constraint $\frac{1}{2}h(\theta_i)^2 \leq \eta(\theta_i)$. Equation (A5) implies the optimality condition

$$
h(\theta_i) = -\frac{\sigma_D D}{\lambda} V_D^i. \tag{A6}
$$

Slackness then gives

$$
\frac{\sigma_D^2 D^2}{\lambda^2} \left(V_D^i \right)^2 = 2\eta(\theta_i), \tag{A7}
$$

implying

$$
h(\theta_i) = -\sqrt{2\eta(\theta_i)} \operatorname{sign}[\sigma_D D V_D] = -\sqrt{2\eta(\theta_i)}. \tag{A8}
$$

This result proves the first statement. To prove the second statement, we note that

$$
V^i(D) = \frac{D^{1-\gamma}}{1-\gamma} E\left[\int_s^{\infty} e^{-\delta(t-s)} \left(\frac{D(t)}{D(s)} \right)^{1-\gamma} dt \middle| \widetilde{\theta} = \theta_i - \sqrt{2\eta(\theta_i)}\sigma_D \right]. \tag{A9}
$$

Conditional on $\widetilde{\theta} = \theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$, the solution of the dividend dynamics gives

$$\left(\frac{D(t)}{D(s)}\right)^{1-\gamma} = \exp\left\{(1-\gamma)\left(\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D - \frac{\sigma_D^2}{2}\right)(t-s) + (1-\gamma)\sigma_D\left(B_D(t) - B_D(s)\right)\right\},$$

implying, under the given assumptions,

$$E\left[\int_s^\infty e^{-\delta(t-s)}\left(\frac{D(t)}{D(s)}\right)^{1-\gamma} dt \,\middle|\, \widetilde{\theta} = \theta_i - \sqrt{2\eta(\theta_i)}\sigma_D\right] = \frac{1}{\widehat{\theta}_i + (1-\gamma)\sqrt{2\eta(\theta_i)}\sigma_D},$$

where

$$\widehat{\theta}_i = \delta - (1-\gamma)\theta_i + \gamma(1-\gamma)\frac{\sigma_D^2}{2} > 0. \tag{A10}$$

We thus obtain for the equilibrium price of the risky asset with dividend process D

$$\frac{P(t)}{D(t)} = \sum_{i=1}^n \pi_i E\left[\int_t^\infty e^{-\delta(s-t)}\left(\frac{D(s)}{D(t)}\right)^{1-\gamma} ds \,\middle|\, \widetilde{\theta} = \theta_i - \sqrt{2\eta(\theta_i)}\sigma_D\right], \tag{A11}$$

or equivalently

$$P(t)\rho(t) = \sum_{i=1}^n \pi_i E\left[\int_t^\infty \rho(s) D(s)\, ds \,\middle|\, \widetilde{\theta} = \theta_i - \sqrt{2\eta(\theta_i)}\sigma_D\right] \tag{A12}$$

where $\rho(t) = u_c(D(t), t) = e^{-\delta t} D(t)^{-\gamma}$. This result proves the second statement of the proposition. Writing Equation (A12) in differential form and applying it to the risky asset paying a "dividend" $D = r$ we obtain

$$\begin{aligned} r dt &= -\sum_{i=1}^n \pi_i E_t\left[\frac{d\rho}{\rho}\,\middle|\, \widetilde{\theta} = \theta_i - \sqrt{2\eta(\theta_i)}\sigma_D\right] \\ &= \left(\delta + \gamma\left(m_\theta + m_{h(\theta)}\sigma_D\right) - \frac{1}{2}\gamma(\gamma+1)\sigma_D^2\right) dt, \end{aligned}$$

that is, the third statement of the proposition, concluding the proof. ■

Proof of Proposition 2. Statement (i) follows by applying the proof of Proposition 2 in Veronesi (2000) to the $D-$dynamics in Equation(18) under the worst-case likelihood $h^*(\theta) = -\sqrt{2\eta(\theta)}$. Statement (ii) follows by expressing the excess return dynamics obtained in (i) with respect to the filtered Brownian motions $\tilde{B}_D$, $\tilde{B}_e$ under the reference model. ■

Proof of Corollary 1. From Lemma 2, V_{θ,h^*} is increasing in mean-preserving spreads $\tilde{\Pi}$, under the given assumptions. Moreover,

$$-m_{h^*} = \sum_{i=1}^n \pi_i \sqrt{\eta(\theta_i)}\sigma_D \tag{A13}$$

is also increasing in mean-preserving spreads, because of the assumed convexity of $\sqrt{2\eta(\theta)}$ as a function of θ. ■

Reference

Anderson, E. W., L. P. Hansen, and T. J. Sargent. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk and Model Detection *Journal of the European Economic Association* 1(1):68–123.

Attanasio, P. O., J. Banks, and S. Tanner. 2002. Asset Holding and Consumption Volatility. *Journal of Political Economy* 110:771–92.

Bansal, R., and A. Yaron. 2004. Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance* 59(4):1481–509.

Barberis, N. 2000. Investing for the Long Run when Returns Are Predictable. *Journal of Finance* 55:225–64.

Brandt, M., and Q. Kang. 2004. On the Relationship between the Conditional Mean and Volatility of Stock Returns: A Latent VAR Approach. *Journal of Financial Economics* 72:217–57.

Breeden, D. 1979. An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities. *Journal of Financial Economics* 7:265–96.

Brennan, M. J. 1998. The Role of Learning in Dynamic Portfolio Decisions. *European Finance Review* 1:295–306.

Brennan, M. J., and Y. Xia. 2001. Stock Return Volatility and Equity Premium. *Journal of Monetary Economics* 47:249–83.

Cagetti, M., L. P. Hansen, T. Sargent, and N. Williams. 2002. Robustness and Pricing with Uncertain Growth. *Review of Financial Studies* 15:363–404.

Campbell, J. Y. 1987. Stock Returns and the Term Structure. *Journal of Financial Economics* 18:373–99.

Campbell, J. Y. 1999. Asset Prices, Consumption, and the Business Cycle in J. Taylor, and M. Woodford (eds), *Handbook of Macroeconomics*, North-Holland, Amsterdam, vol. 1, chap. 19.

Campbell, J. Y., and L. Hentschel. 1992. No News Is Good News: An Asymmetric Model of Changing Volatility in Stock Returns. *Journal of Financial Economics* 31:281–318.

Chen, Z., and L. G. Epstein. 2002. Ambiguity, Risk and Asset Returns in Continuous Time. *Econometrica* 70:1403–43.

Detemple, J. 1986. Asset Pricing in a Production Economy with Incomplete Information. *Journal of Finance* 41:383–91.

Dothan, M., and D. Feldman. 1986. Equilibrium Interest Rates and Multiperiod Bonds in a Partially Observable Economy. *Journal of Finance* 41:369–82.

Ellsberg, D. 1961. Risk, Ambiguity and the Savage Axioms. *Quarterly Journal of Economics* 25:643–69.

Epstein, L. G., and T. Wang. 1994. Intertemporal Asset Pricing Under Knightian Uncertainty. *Econometrica* 62(2):283–322.

Epstein, L. G., and M. Schneider. 2002. Learning under Ambiguity Working paper, University of Rochester.

Epstein, L. G., and M. Schneider. 2003a. Recursive Multiple Priors. *Journal of Economic Theory* 113:32–50.

Epstein, L. G., and J. Miao. 2003b. A Two-Person Dynamic Equilibrium under Ambiguity. *Journal of Economic Dynamics & Control* 27:1253–88.

Epstein, L. G., and M. Schneider. 2004. Ambiguity, Information Quality and Asset Pricing. Working paper, University of Rochester.

Feldman, D. 2005. Incomplete Information Equilibria: Separation Theorems and Other Myths Working paper, University of New South Wales.

French, K., W. Schwert, and R. Stambaugh. 1987. Expected Stock Returns and Volatility. *Journal of Financial Economics* 19:3–29.

Gagliardini, P., P. Porchia, and F. Trojani. 2004. Ambiguity Aversion and the Term Structure of Interest Rates Working paper, University of St.Gallen.

Gennotte, G. 1986. Optimal Portfolio Choice Under Incomplete Information. *Journal of Finance* 41:733–46.

Gilboa, I., and D. Schmeidler. 1989. Maxmin Expected Utility with Non-unique Prior. *Journal of Mathematical Economics* 18:141–53.

Hansen, L. P., T. J. Sargent, and D. Tallarini. 1999. Robust Permanent Income and Pricing. *Review of Economic Studies* 66:873–907.

Kandel, S., and R. F. Stambaugh. 1995. On the Predictability of Stock Returns: An Asset-Allocation Perspective. *Journal of Finance* 51:385–424.

Liptser, R., and A. N. Shiryaev. 2001. *Statistics of Random Processes: General Theory*, vol. 1. Springer, Berlin.

Liu, J., J. Pan, and T. Wang. 2005. An Equilibrium Model of Rare-Event Premia and Its Implication for Option Smirks. *Review of Financial Studies* 18(1):131–64.

Lucas, R. E. 1978. Asset Prices in an Exchange Economy. *Econometrica* 46(6):1429–45.

Maenhout, P. J. 2004. Robust Portfolio Rules and Asset Pricing. *Review of Financial Studies* 17(4):951–83.

Mehra, R., and E. C. Prescott. 1985. The Equity Premium: A Puzzle. *Journal of Monetary Economics* 15:145–61.

Miao, J. 2001. Ambiguity, Risk and Portfolio Choice under Incomplete Information. Working paper, University of Rochester.

Pástor, L. 2000. Portfolio Selection and Asset Pricing Models. *Journal of Finance* 55:179–223.

Pástor, L., and R. F. Stambaugh. 2000. Comparing Asset Pricing Models: An Investment Perspective. *Journal of Financial Economics* 56:335–81.

Routledge, B. R., and S. E. Zin. 2001. Model Uncertainty and Liquidity. Working paper, National Bureau of Economic Research.

Sbuelz, A., and F. Trojani. 2002. Equilibrium Asset Pricing with Time Varying Pessimism. Working paper, University of Lugano, Switzerland.

Shepard, N., and A. C. Harvey. 1990. On the Probability of Estimating a Deterministic Component in the Local Level Model. *Journal of Time Series Analysis* 11:339–47.

Trojani, F., and P. Vanini. 2002. A Note on Robustness in Merton's Model of Intertemporal Consumption and Portfolio Choice. *Journal of Economic Dynamics & Control* 26:423–35.

Trojani, F., and P. Vanini. 2004. Robustness and Ambiguity Aversion in General Equilibrium. *Review of Finance* 8(2):279–324.

Uppal, R., and T. Wang. 2003. Model Misspecification and Under-Diversification. *Journal of Finance* 58:2465–86.

Veronesi, P. 1999. Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model. *Review of Financial Studies* 12:975–1007.

Veronesi, P. 2000. How does Information Quality Affect Stock Returns?. *Journal of Finance* 55:807–37.

Vissing-Jorgensen, A. 2002. Limited Asset Market Participation and the Elasticity of Intertemporal Substitution. *Journal of Political Economy* 110:825–53.

Vissing-Jorgensen, A., and O. P. Attanasio. 2003. Stock-Market Participation, Intertemporal Substitution, and Risk-Aversion. *American Economic Review* 93(2):383–91.

Wakker, P., and D. Deneffe. 1996. Eliciting von Neumann-Morgenstern Utilities when Probabilities Are Distorted or Unknown. *Management Science* 42:1131–50.

Xia, Y. 2001. Learning About Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation. *Journal of Finance* 56:205–46.