Learning and Asset Prices under Ambiguous Information*

Markus Leippold[a,b] Fabio Trojani[c,*] Paolo Vanini[b,d]

[a] *Capital Markets Research, Federal Reserve Bank of New York*

[b] *Swiss Banking Institute, University of Zurich, Switzerland*

[c] *Swiss Institute of Banking and Finance and Department of Economics, University of St. Gallen, Switzerland*

[d] *Zurich Cantonal Bank, Switzerland*

(First version: September 2003; This version: August 2005)

---

* We are grateful to Andrew Abel for many valuable suggestions on an earlier draft. We also thank Freddy Delbaen, David Feldman, Günter Franke, Rajna Gibson, Jens Jackwerth, Yvan Lengwiler, Marco LiCalzi, Abraham Lioui, Alessandro Sbuelz, Sandra Sizer, 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, 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.

*Correspondence address: Fabio Trojani, Swiss Institute of Banking and Finance, University of St. Gallen, Rosenbergstr. 52, CH-9000 St. Gallen, e-mail: Fabio.Trojani@unisg.ch.

Learning and Asset Prices under Ambiguous Information

Abstract

We propose a new continuous-time framework for studying asset prices under learning and ambiguity aversion. In a Lucas economy with time-additive power utility, a discount for ambiguity arises for a relative risk aversion below one or, equivalently, an intertemporal elasticity of substitution above one. The joint presence of learning and ambiguity enforces large equity premia and model predictions consistent with well-known asset pricing puzzles. For realistic amounts of ambiguity, absence of either learning or ambiguity aversion implies low volatilities or low equity premia.

Keywords: Financial Equilibria, Learning, Knightian Uncertainty, Ambiguity Aversion, Model Misspecification.

JEL Classification: C60, C61, G11.

In this paper we study the equilibrium asset-pricing implications of learning when the distinction between risk and ambiguity (Knightian uncertainty) aversion matters. We define ambiguity as those situations in which investors do not rely on a single probability law to describe the relevant random variables. Ambiguity aversion means that investors dislike the ambiguity of the probability law of asset returns.

Using a continuous-time economy, we study the joint impact of learning and ambiguity aversion on asset prices and learning dynamics. More specifically, we tackle the problem of asset pricing under learning and ambiguity aversion in a continuous-time Lucas (1978) exchange economy in which economic agents have partial information on the ambiguous dynamics of some aggregate endowment process. We develop a new continuous-time setting of learning under ambiguity aversion that allows us to study analytically both the conditional and unconditional implications for equilibrium asset prices.

Whether ambiguity aversion gives a plausible explanation for salient features of asset prices when learning is accounted for is an open issue. For instance, researchers ask if the equity premium puzzle can be still addressed in a model of ambiguity aversion as new data are observed and more data-driven knowledge about some unobservable variable becomes available. The answer to this question depends on the ability of investors to learn completely the underlying probability laws under a misspecified belief. Rational models of Bayesian learning[1] cannot address such issues, because they are based on a single-prior/single-likelihood correct specification assumption about the beliefs that define the learning dynamics. Therefore, to study asset prices under learning and ambiguity aversion, we must consider settings that explicitly address a possible misspecification of beliefs and the corresponding learning dynamics.

In our model, agents learn only some of the global ambiguous characteristics of the underlying endowment process, parameterized by a finite set of relevant ambiguous states of the economy. Moreover, we account for a set of multiple likelihoods in the description of the local ambiguous properties of the underlying endowment process, conditional on any relevant state

[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) and (2000), and Xia (2001). For a recent review of the literature, see Feldman (2005).

of the economy. Since we allow for multiple likelihoods, our model does not resolve ambiguity in the long run, even when the underlying endowment process is not subject to changes in regime.

Using the exchange economy framework, we are able to compute analytically the equilibrium equity premia, equity expected returns and volatilities, interest rates, and price dividend ratios. Since we also allow for external exogenous signals on the unobservable expected growth rate of the aggregate endowment, we can study the relation between asset prices, information noisiness and ambiguity.

Our main focus is on studying how learning under ambiguity aversion affects the functional form of the equilibrium variables and if it worsens existing asset pricing puzzles. For instance, although there is plenty of evidence that ambiguity aversion settings do help to explain the equity premium and the low interest-rate puzzles, we also know that in a pure setting of learning the equity premium can be even more of a puzzle (see, e.g., Veronesi (2000)). So we ask if the combination of learning and ambiguity aversion helps to give a reasonable explanation for the equity premium puzzle. We also know that pure settings of learning can explain excess volatility and volatility clustering of asset returns. At the same time, simple constant opportunity set models of ambiguity aversion do not affect substantially expected equity returns and equity volatility (see, e.g., Maenhout (2004) and Sbuelz and Trojani (2002)). So we ask if the combination of learning and ambiguity aversion still generates excess volatility and volatility clustering.

We can directly address these questions in our model. First, we find that in a Lucas economy, learning under ambiguity aversion implies an equilibrium discount for ambiguity, if and only if relative risk aversion is smaller than one, or equivalently, if the elasticity of intertemporal substitution (EIS) is above one. Under low risk aversion, learning and ambiguity aversion increase conditional equity premia and volatilities. In addition, the part of equity premium due to the interaction of learning and ambiguity aversion is the largely dominating one. 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 a promising feature of our setting, in that it explains simultaneously the equity premium and the risk-free rate puzzles without an ad hoc use of preference parameters. That is, already for moderate amounts of ambiguity. Third, under learning and ambiguity aversion, the true theoretical equilibrium relation between excess returns and conditional variances is highly time-varying. This feature can generate (i) estimated relations between excess returns and conditional variances that have undetermined signs over time and (ii) a huge time-varying bias in the naively estimated risk-return trade-off using, e.g., regression methods. In addition, estimates of the EIS based on standard Euler equations for equity returns are strongly downward biased. Therefore, in our model an EIS above one can be consistent with observed estimated EIS clearly below one. Finally, since in our setting ambiguity does not resolve asymptotically, we show explicitly that asset pricing relations under ambiguity aversion, but no learning, can be interpreted as the limit of an equilibrium learning process under ambiguity aversion. The paper is organized as follows. The next section reviews the relevant literature on learning and ambiguity. Section 2 introduces our setting of learning under ambiguity aversion. The properties of the optimal learning dynamics are studied in Section 3. Section 4 characterizes and discusses conditional asset pricing relations under learning and ambiguity aversion. Section 5 concludes.

1. Background

Distinguishing between ambiguity aversion and risk aversion is economically and behaviorally important. As the Ellsberg (1961) paradox illustrates, investors behave differently under ambiguity and risk aversion. Moreover, ambiguity itself is pervasive in financial markets. 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.

Recently, authors have attempted to incorporate ambiguity aversion also in an intertemporal context. These approaches have been largely inspired by the Gilboa and Schmeidler (1989) Max-Min expected utility setting. 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) later provide a discrete-time axiomatic foundation for that model in which they show that a dynamically consistent conditional version of the Gilboa and Schmeidler (1989) preferences can be represented by using a recursive Max-Min expected utility criterion over a set of multiple distributions. Chen and Epstein (2002) extend that setting to continuous time. Hansen, Sargent, and Tallarini (1999, in discrete time) and Anderson, Hansen, and Sargent (2003, in continuous time) propose a second setting of intertemporal ambiguity aversion based on an alternative form of Max-Min expected utility preferences. Their settings apply robust control theory to economic problems.

Various authors have proposed continuous-time models of full information economies with ambiguity aversion that give plausible explanations for several important characteristics of asset prices. 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). By construction, these models exclude any form of learning. Investors observe perfectly the state variables that determine the opportunity set, but they are not fully aware of the probability distribution of the state variables. Consequently, some form of conservative worst-case optimization determines their optimal decision rules.

Only more recently have a few authors addressed the issue of learning under ambiguity aversion. Using a production economy that is subject to exogenous regime shifts and driven by a two-state Markov chain, Cagetti, Hansen, Sargent, and Williams (2002) apply robust filtering theory to show the impact of learning and ambiguity aversion on the aggregate capital stock, equity premia, and price dividend ratios. Using numerical methods, they provide evidence that ambiguity aversion increases precautionary saving in a way that is similar to the effect of an increased subjective time preference rate. Moreover, the equity premium increases substantially due to ambiguity aversion and price dividend ratios turn out to be lower.

Our model differs from the Cagetti et al. (2002) setting in several aspects. We work with an exchange Lucas economy without regime shifts and employ a more tractable homothetic setting of preferences under ambiguity aversion. These two features allow us to solve the model in closed form and to discuss the convergence of equilibria under learning and ambiguity aversion to equilibria with no learning. Due to the tractability of the model, we are able to study theoretically and in great detail all relevant asset-pricing relations and their dependence on model parameters. For instance, we show that in the partial information exchange Lucas economy, ambiguity aversion can fail to increase equity premia if standard risk aversion is too high or, equivalently, the EIS is too low. Moreover, we allow for external public signals, in excess of dividends, and for heterogeneous ambiguity sizes across the relevant states of the economy. These extensions have nontrivial implications for the resulting asset pricing relations. For instance, the effect of ambiguity aversion on price dividend ratios cannot be mapped into an adjustment of the subjective time preference rate. Moreover, ambiguity premia caused by extraneous signals have an important role in the determination of equity premia.

Epstein and Schneider (2002) use a discrete-time setting to highlight that learning about an unknown parameter under multiple likelihoods can fail to resolve ambiguity asymptotically, even when the underlying state process is not subject to regime shifts. Epstein and Schneider (2004) present a related learning model under ambiguity and illustrate the impact of an ambiguous signal precision on asset prices. Using numerical methods applied to a setting with risk neutral investors, they show that an ambiguous quality of information, defined in terms of a set of possible values of the signal precision parameter, can generate skewed asset returns and excess volatility. The focus of our paper is different from Epstein and Schneider (2002). While they focus on ambiguous signal precision and its impact on returns volatility and skewness, we focus on ambiguous signals about expected dividend growth and the resulting implications, primarily for equity premia. In this context, we show that for realistic amounts of ambiguity the joint interaction of learning and ambiguity aversion is responsible for very large equity premia. At the same time, we show that the additional model predictions are consistent with the interest rate and excess volatility puzzles. In addition, the tractability of our setting allows us to compute all equilibrium quantities analytically under standard assumptions on the utility function. For instance, since we assume general power utility investors, we can disentangle the impact of ambiguity aversion and risk aversion on equity premia, and we show that a positive ambiguity premium arises only for a moderate risk aversion or, equivalently, a sufficiently large

EIS. Finally, our model allows for ambiguity on fundamental and extraneous ambiguous signals. This feature emphasizes the distinction between ambiguous fundamental signals, which affect both the underlying opportunity set and expectations of the economy's growth rate, and ambiguous extraneous signals, which affect only the expected growth rate of the economy.

Finally, Knox (2005) proposes an axiomatic setting of learning about a model parameter under ambiguity aversion. However, he does so without studying the general equilibrium asset pricing implications.

2. The Model

We start with a continuous-time Lucas economy. The drift rate in the diffusion process for the dividend dynamics is unobservable. Investors learn about the "true" drift by observing dividends and a second distinct signal. In contrast to most other models of rational learning, we explicitly allow for a distinction between noisy and ambiguous signals. For a purely noisy signal, the distribution conditional on a given parameter value is known. For ambiguous signals, the distribution conditional on a given parameter value is either unknown or at least not uniquely identified. This distinction broadens the notion of information quality. In many situations, it is plausible that agents are aware of a host of poorly understood or unknown factors that obscure the interpretation of a given signal. Such obscuring factors can depend on economic conditions or on some specific aspects of a given state of the economy.

In our model, signals on the state of the economy are ambiguous and can be interpreted differently, depending on whether agents condition on good or bad economic information. We model this feature by a set of multiple likelihoods on the underlying dividend dynamics. The size of this set of multiple likelihoods can depend on the state of the economy. Disentangling the properties of noisy and ambiguous signals across the possible relevant states of the economy gives the model builder a more realistic way to specify a learning behavior with multiple beliefs.

Our objective is to characterize equilibrium asset returns in the presence of noisy signals in ambiguous states of the economy. Therefore, we develop an equilibrium model of learning under ambiguity aversion that consists of the following key elements:

a) A parametric reference model for the underlying dividend process and the unobservable dividend drift. We explicitly treat the reference model as an approximation to reality, rather than an exact description of it. Therefore, economic agents possess some motivated specification doubts. Specification doubts arise, e.g., when agents are aware that, based on an empirical specification analysis, they choose the reference model from a set of statistically close models.

b) A set of multiple likelihoods on the dynamics of the unobservable dividend drift. We use these multiple likelihoods to compute a set of beliefs about the unknown dividend drift dynamics. This set of beliefs represents the investor's ambiguity on the dynamic structure of the unobservable expected dividend growth rate. The set of multiple likelihoods can also serve as a description of a class of alternative specifications to the reference model. Since these specifications are statistically close, they are difficult to distinguish from the reference belief.

c) An intertemporal Max-Min expected utility optimization problem.[2] The Max-Min problem models the agents' optimal behavior given their attitudes to risk and ambiguity and under the relevant set of multiple ahead beliefs.

Given the three key elements above, a set of standard market clearing conditions on good and financial markets closes the model. After solving the model, we provide equilibrium asset prices under learning and ambiguity aversion.

[2]See also Gilboa and Schmeidler (1989), Chen and Epstein (2002), Epstein and Schneider (2003), and Knox (2005).

2.1. The Reference Model Dynamics

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

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

where $\gamma > 0$. The representative investor has a parametric reference model that describes in an approximate way the dynamics of dividends D

$$\frac{dD}{D} = E_t\left(\frac{dD}{D}\right) + \sigma_D dB_D \quad , \tag{2}$$

where $\sigma_D > 0$ and $E_t\left(\frac{dD}{D}\right)$ is the unobservable drift of dividends at time t. Investors further observe a noisy unbiased signal e on $E_t(dD/D)$ with dynamics

$$de = E_t\left(\frac{dD}{D}\right) + \sigma_e dB_e \quad , \tag{3}$$

where $\sigma_e > 0$. The standard Brownian motions B_D and B_e are independent.

The parametric reference model to describe the dividend drift dynamics is a rough approximation to reality. It implies a simple geometric Brownian motion dynamics for dividends with a constant drift that takes one of a finite number of candidate values, as in (2000).

Definition 1 *The reference model dividend drift specification is given by*

$$\frac{1}{dt}E_t(dD/D) = \theta \quad , \tag{4}$$

for all $t \geq 0$*, where* $\theta \in \Theta := \{\theta_1, \theta_2, ..., \theta_n\}$ *and* $\theta_1 < \theta_2 < ... < \theta_n$*. The representative investor has some prior beliefs* $(\widehat{\pi}_1, .., \widehat{\pi}_n)$ *at time* $t = 0$ *on the validity of the candidate drift values* $\theta_1, ..., \theta_n$*.*

In a single-likelihood Bayesian framework, Definition 1 implies a parametric single-likelihood model for the dividend dynamics, where the specific value of the parameter θ is unknown. The only relevant statistical uncertainty about the dynamics in equation (2) is parametric. Therefore, in a single-likelihood Bayesian setting, a standard filtering process leads to asymptotic learning of the unknown constant dividend drift θ in the class Θ of candidate drift values. Moreover, the equilibrium asset returns dynamics can be determined and the pricing impact of learning can be studied analytically.

In the sequel, we strongly depart from such a Bayesian asset pricing setting by allowing for the possibility of a misspecification in the reference model of Definition 1. Relevant misspecifications take a general nonparametric form, so that they cannot be consistently detected even by means of parametric Bayesian model selection approaches.

2.2. Multiple Likelihoods

In reality, a Bayesian (single-likelihood) specification hypothesis of the type given in Definition 1 is very restrictive. Such a specification assumes that even when the dividend drift is unobservable, the investor can identify a parametric model that is able to describe exactly, in a probabilistic sense, the relevant dividend drift dynamics. More realistically, we propose a model of learning in which economic agents have some specification doubts about the given parametric reference model. Such a viewpoint is motivated by considering that any empirical specification analysis provides a statistically preferred model only after having implicitly rejected several alternative specifications that are statistically close to it. Even if such alternative specifications to the reference model are statistically close, it is quite possible that they can quantitatively and qualitatively affect the optimal portfolio policies derived under the reference model's assumptions.[3] To avoid the negative effects of a misspecification on the optimal policies derived from the reference model, we prefer to work with consumption/investment optimal policies that account explicitly for the possibility of model misspecifications. This approach should ensure some degree of robustness of the optimal policies against misspecifications of the reference model dynamics.

[3]The importance of this issue has been recognized, e.g., by Huber (1981) in his influential introduction to the theory of robust statistics and has been further developed, e.g., in econometrics to motivate several robust procedures for time series models. See Krishnakumar and Ronchetti (1997), Sakata and White (1999), Ronchetti and Trojani (2001), Gagliardini, Trojani, and Urga (2005), Mancini, Ronchetti, and Trojani (2005), and Ortelli and Trojani (2005), for some recent work in the field.

We explicitly incorporate specification doubts by modeling agents' beliefs, conditional on any possible reference model drift θ, by using a set of multiple likelihoods. To define these sets, we restrict ourselves to absolutely continuous misspecifications of the geometric Brownian motion process in equations (2) and (4). By Girsanov's theorem, the likelihoods implied by absolutely continuous probability measures can be equivalently described by a corresponding set of drift changes in the model dynamics in equations (2) and (4).

Let $h(\theta)\,\sigma_D$ be a process describing the dividend drift change implied by such a likelihood function. We assume that $h(\theta) \in \Xi(\theta)$, where $\Xi(\theta)$ is a suitable set of standardized change of drift processes (see Subsection 2.3, Assumption 2 below). Under such a likelihood, the prevailing dividend dynamics are

$$\frac{dD}{D} = E_t^{h(\theta)}\left(\frac{dD}{D}\right) + \sigma_D dB_D \quad , \tag{5}$$

with signal dynamics

$$de = E_t^{h(\theta)}\left(\frac{dD}{D}\right) + \sigma_e dB_e \quad . \tag{6}$$

Ambiguity on D's dynamic arises as soon as for some $\theta \in \Theta$, the set $\Xi(\theta)$ contains a drift distortion process $h(\theta)$ different from the zero process. In this case, we consider several possible functional forms of the drift in equation (5) together with the reference model dynamics in equations (2) and (4). The set of possible drifts implied by the multiple likelihoods in $\Xi(\theta)$ represents the relevant beliefs of an agent who does not completely trust the reference model dynamics.

2.3. A Specific Set of Multiple Likelihoods

Compared with the Bayesian single-likelihood specification hypothesis in Definition 1, an agent with multiple likelihood beliefs is less ambitious. Thus, we have the following assumption.

Assumption 1 *The "true" dividend drift specification is given by*

$$\frac{1}{dt}E_t^{h(\theta)}\left(\frac{dD}{D}\right) = \theta + h(\theta, t)\,\sigma_D \quad , \tag{7}$$

for all $t \geq 0$, some $\theta \in \Theta$ and some $h(\theta) \in \Xi(\theta)$. The representative investor has some beliefs $(\widehat{\pi}_1, .., \widehat{\pi}_n)$ at time $t = 0$ on the a priori plausibility of the different sets $\Xi(\theta_1), .., \Xi(\theta_n)$ of candidate drift processes.

Under Assumption 1, the representative agent recognizes that a whole class $\Xi(\theta)$ of standardized drift changes is statistically hardly distinguishable from a zero drift change, i.e., from the reference model dynamics with drift θ given in Definition 1. If $\Xi(\theta) = \{0\}$ for all $\theta \in \Theta$, the Bayesian setup in (2000). follows. Then, agents would be concerned only with the pure noisiness of a signal on the parameter value θ. Therefore, the distinction between ambiguity and noisiness is absent in a pure Bayesian setting.

The size of the set $\Xi(\theta)$ describes the degree of ambiguity associated with any possible reference model dividend drift θ. The broader the set $\Xi(\theta)$, the more ambiguous are the signals about a specific dividend drift $\theta + h(\theta)\sigma_D \in \Xi(\theta)$. Such ambiguity reflects the fact that there are aspects of the unobservable dividend drift dynamics that agents think are hardly possible, or even impossible, to ever know. For example, the representative agent is aware of the problem that identifying the exact functional form for a possible mean reversion in the dividend drift dynamics is empirically a virtually infeasible task.[4] Accordingly, the agent tries to understand only a limited number of features on the dividend dynamics.

In our setting, we represent this limitation with a learning model about the relevant neighborhood $\Xi(\theta)$, rather than with a learning process on the specific form of $h(\theta)$. Therefore, the learning problem under multiple beliefs becomes one of learning the approximate features of the underlying dividend dynamics across a class of model neighborhoods $\Xi(\theta)$, $\theta \in \Theta$. Hence, the representative agent has ambiguity about some local dynamic properties of equity returns, conditional on some ambiguous local macroeconomic conditions. The agent tries to infer some more global characteristics of asset returns in dependence of such ambiguous macroeconomic states. Finally, since the size of the set $\Xi(\theta)$ can depend on the specific value of θ, our setting allows also for a degree of ambiguity that depends on economic conditions.

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

Next, we specify the set $\Xi(\theta)$ of multiple likelihoods relevant for our setting. The set contains all likelihood specifications that are statistically close (in some appropriate statistical measure of model discrepancy) to the one implied by the reference model dynamics. This feature makes more precise the general principle that $\Xi(\theta)$ should contain only models for which agents have some well-motivated specification doubt, relative to the given reference model dynamics. The relevant reference model misspecifications are constrained to be small and are thus hardly statistically detectable. Moreover, the set $\Xi(\theta)$ contains any misspecification which is statistically close to the reference model. This property defines a whole neighborhood of slight, but otherwise arbitrary, misspecifications of the reference model distributions.

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

$$\Xi(\theta) := \left\{ \theta + h(\theta)\,\sigma_D : \frac{1}{2} h^2(\theta, t) \leq \eta(\theta) \text{ for all } t \geq 0 \right\} \quad , \tag{8}$$

where $\eta(\theta_1), .., \eta(\theta_n) \geq 0$. Moreover, for any $i \neq j$ it follows:

$$\Xi(\theta_i) \cap \Xi(\theta_j) = \emptyset \ . \tag{9}$$

Under Assumption 2, we can constrain the discrepancy between the reference model distributions under a drift θ and those under any model implied by a drift distortion process $h(\theta) \in \Xi(\theta)$ to be statistically small. A moderate bound $\eta(\theta)$ implies that for any likelihood in the set $\Xi(\theta)$ there will be a small statistical discrepancy relative to a reference model dynamics with drift θ. In all our model calibrations below in the paper, we will impose moderate sizes for parameter $\eta(\theta)$, in order to avoid unrealistically pessimistic beliefs in our model. Since equation (8) does not make any specific assumption on a parametric structure for $h(\theta)$, the neighborhood $\Xi(\theta)$ is nonparametric and contains all likelihood models that are compatible with the bound in set (8).

Condition (9) means that economic agents have ambiguity only about candidate drifts within neighborhoods, but not between neighborhoods. In other words, different macroeconomic conditions can be mapped into disjoint sets of likely drift dynamics. Such a situation arises when the degree of ambiguity $\eta(\theta)$ in the economy is not too high, relative to the distance between reference model drifts θ. Therefore, we focus on situations in which ambiguity in the economy is moderate.

2.4. Ambiguity Aversion and Intertemporal Max-Min Expected Utility

We denote by $\mathcal{F}(t)$ the information available at time t that 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 the amount of ambiguity relevant to investors. The representative investor determines consumption and investment plans $C(t)$ and $w(t)$ by solving the intertemporal Max-Min expected utility optimization problem

$$(P) \quad : \quad \max_{C,w} \inf_{h(\theta)} E\left[\int_0^\infty u(C,s)\,ds \right| \left. \mathcal{F}(0) \right] \quad , \tag{10}$$

subject to the dividend and wealth dynamics

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

where for any $\theta \in \Theta$ the standardized drift distortion is such that $h(\theta) \in \Xi(\theta)$ and Assumption 2 holds. In problem (10), the representative agent must select, in excess of an optimal consumption/investment policy, an optimal worst-case belief h out of the admissible class $\Xi(\theta)$. The fact that such an optimal belief is determined endogenously as a function of investors' preferences differs sharply from the standard Bayesian setting in which beliefs are fixed by a parametric assumption on the unobservable dynamics of the underlying dividend drift process.

3. Multiple Filtering Dynamics under Ambiguity

3.1. Bayesian Learning and Likelihood Misspecification

Learning under ambiguity requires constructing a set of standard Bayesian ahead beliefs for $E_t^{h(\theta)}\left(\frac{dD}{D}\right)$ that are functions of likelihoods $h(\theta) \in \Xi(\theta)$. For a given likelihood model $h(\theta) \in \Xi(\theta)$, let $\pi_i(t)$ be the investor's belief that the drift rate is $\theta_i + h(\theta_i)\sigma_D$, conditional on past dividend and signal realizations, i.e.,

$$\pi_i(t) = \Pr\left(\frac{1}{dt}E_t^{h(\theta)}\left(\frac{dD}{D}\right) = \theta_i + h(\theta_i)\sigma_D \middle| \mathcal{F}(t)\right) \quad . \tag{11}$$

The distribution $\Pi(t) := (\pi_1(t), .., \pi_n(t))$ summarizes investors beliefs at time t, under a given likelihood $h(\theta) \in \Xi(\theta)$. Given such beliefs, investors can compute the expected dividend drift at time t as

$$\frac{1}{dt}E^{h(\theta)}\left(\frac{dD}{D}\middle| \mathcal{F}(t)\right) = \sum_{i=1}^{n}(\theta_i + h(\theta_i)\sigma_D)\pi_i(t) = m_{\theta,h} \quad , \tag{12}$$

where

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

The filtering equations implied by any given likelihood $h(\theta) \in \Xi(\theta)$ are standard (see, e.g., Liptser and Shiryaev (2001)).

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

$$d\pi_i = \pi_i(\theta_i + h(\theta_i)\sigma_D - m_{\theta,h})\left(k_D d\widetilde{B}_D^h + k_e d\widetilde{B}_e^h\right) \quad ; \quad i = 1, .., n \quad , \tag{14}$$

where

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

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

To study how a likelihood misspecification affects the dynamic properties of the perceived beliefs, we express the dynamics in equation (14) in terms of the original Brownian motions B_D and B_e. This description helps to highlight how a likelihood misspecification can fail to imply consistency of a Bayesian learning process.

Corollary 1 *Let* $h(\theta) \in \Xi(\theta)$ *be an admissible likelihood and* $\theta_l + h_D\sigma_D$, $l \in \{1,..,n\}$, *be the true dividend drift process. It then follows that*

$$d\pi_i = \pi_i\left(\theta_i + h(\theta_i)\sigma_D - m_{\theta,h}\right)\left[k\left(\theta_l + h_D\sigma_D - m_{\theta,h}\right)dt + k_D dB_D + k_e dB_e\right] \ ; \ i = 1,..,n \quad , \tag{15}$$

where $k = k_D^2 + k_e^2$.

Expression (15) gives the dynamics of the posterior probability π_i for the general case in which the likelihood $h(\theta)$ might be different from the true underlying drift distortion h_D, i.e., the case in which the likelihood $h(\theta)$ might be misspecified. The case of a correctly specified likelihood arises when $h(\theta_l) = h_D$. In this case, an inspection of the dynamics in equation (15) shows that the underlying dividend drift $\theta_l + h_D\sigma_D$ will eventually be learned.

Corollary 2 *If the likelihood* $h(\theta)$ *is correctly specified, i.e., if* $h_D = h(\theta_l)$ *for some* $\theta_l \in \Theta$, *then* $\pi_l \underset{t\to\infty}{\to} 1$, *almost surely.*

Corollary 2 shows that consistency of a Bayesian learning process is inherently linked to the correct specification of the given likelihood. Intuitively, consistency cannot be generally expected under a misspecified likelihood $h(\theta)$. To illustrate the basic point, we can study the resulting learning dynamics in a setting with only two possible dividend-drift states.

Example 1 Consider the simplified model structure:

$$\Theta = \{\theta_1, \theta_2\} \quad , \quad h(\theta_1) = h(\theta_2) = 0 \quad . \tag{16}$$

Let $\theta_1 + h_D\sigma_D$ be the true underlying dividend drift process. Then, equation (15) implies the learning dynamics:

$$d\pi_1 \quad = \quad \pi_1 (1-\pi_1)(\theta_1-\theta_2)\left[k\left(\theta_1 + h_D\sigma_D - m_{h,\theta}\right)dt + k_D dB_D + k_e dB_e\right]. \tag{17}$$

From Example 1, we see immediately that if

$$\theta_1 + h_D\sigma_D < m_{\theta,h} \quad (\theta_1 + h_D\sigma_D > m_{\theta,h}) \quad , \tag{18}$$

then $\pi_1 \to 1$ ($\pi_1 \to 0$) as $T \to \infty$, almost surely. Under these conditions, investors will "learn" asymptotically a constant dividend-drift process θ_1 (θ_2), even if the true one, $\theta_1 + h_D\sigma_D$, might be time-varying in a nontrivial and unpredictable way.

This observation implies that we will always have $\pi_1 \to 1$ ($\pi_1 \to 0$) as $T \to \infty$ for all settings in which the true drift $\theta_1 + h_D\sigma_D$ is uniformly lower than θ_1 (higher than θ_2). In the more general case with $\theta_1 + h_D\sigma_D$ between θ_1 and θ_2, both outcomes are possible (i.e., either $\pi_1 \to 1$ or $\pi_1 \to 0$). Figure 1 illustrates this point.

Insert Figure 1 about here

In Figure 1, we plot two different trajectories of π_1 under a dividend-drift process such that

$$\theta_1 + h_D(t)\,\sigma_D = \begin{cases} (\theta_1+\theta_2)/2 + a & t \in (k, k+1], \\ (\theta_1+\theta_2)/2 - a & t \in (k+1, k+2] \end{cases} \quad , \tag{19}$$

where $k \in \mathbb{N}$ is even and $|a| < (\theta_2 - \theta_1)/2$. Process (19) describes a deterministic and piecewise constant dividend-drift misspecification. Although we could consider more complex (possibly nonparametric) misspecifications, the main message of Figure 1 would not change.

Figure 1 shows that under a dividend-drift process (19), a Bayesian investor could converge to infer asymptotically both θ_1 and θ_2 as the dividend-drift process that generated asset prices, even if the true drift process is always strictly between θ_1 and θ_2. In Panel A, we plot two possible posterior probabilities trajectories when no shift arises ($a = 0$). In Panel B, we add two alternative trajectories implied by $a = 0.015$, when a yearly deterministic shift in the underlying parameters is present. The only attainable stationary points in the dynamics (17) are the points $\pi_1 = 1$ and $\pi_1 = 0$. Any value $\pi_1 \in (0, 1)$ such that

$$\theta_1 + h_D \sigma_D = m_{h,\theta} \tag{20}$$

makes the drift, but not the diffusion, equal to zero in the dynamics (17). Consequently, π_1 will never stabilize asymptotically in regions such that $m_{h,\theta} \approx \theta_1 + h_D \sigma_D$. If the goal is to approximate adequately $\theta_1 + h_D \sigma_D$ by means of $m_{h,\theta}$, even under a misspecified likelihood, an asymptotic behavior such that $m_{h,\theta} \approx \theta_1 + h_D \sigma_D$ would be ideally more natural. However, this behavior will never arise under the given misspecified likelihood. Richer, but qualitatively similar, patterns emerge when we enlarge the set of possible states of the economy or when the form of introduced misspecification in the likelihood is more complex.

This discussion highlights that a Bayesian investor will not be able to evaluate exactly the utility of a consumption/investment strategy, because she will not identify exactly the underlying dividend-drift process, even asymptotically. Therefore, we work with a learning setting in which investors explicitly exhibit some well-founded specification doubts about the given reference model.

3.2. Learning Under Ambiguity Aversion

Which learning behavior should agents adopt in an ambiguous environment? Since agents are not particularly comfortable with a specific element of $\Xi(\theta)$, they base their beliefs on the whole set of likelihoods $\Xi(\theta)$. By Corollary 1, this approach generates a whole class $\mathcal{P}$ of indistinguishable dynamic dividend-drift prediction processes given by

$$\mathcal{P} = \{ m_{\theta,h} : h(\theta) \in \Xi(\theta) \} \quad , \tag{21}$$

where the dynamics of any of the corresponding posterior probabilities $\pi_1,..,\pi_n$ under the likelihood $h(\theta)$ is given by

$$d\pi_i = \pi_i \left(\theta_i + h\left(\theta_i\right)\sigma_D - m_{\theta,h}\right)\left(k_D d\widetilde{B}_D^h + k_e d\widetilde{B}_e^h\right) \quad , \quad i = 1,..,n \quad . \tag{22}$$

The set $\mathcal{P}$ of dynamic dividend-drift predictions represents investor's ambiguity on the true dividend-drift process, conditional on the available information generated by dividends and signals. As expected, the larger the size of the set of likelihoods $\Xi(\theta)$ (i.e., the ambiguity about the dividend dynamics), the larger the size of the set $\mathcal{P}$ of dynamic dividend-drift prediction processes.

Using the set $\mathcal{P}$ of dynamic dividend-drift predictions, we can write the continuous-time optimization problem (10) as a full information problem in which we define the relevant dynamics in terms of the filtration $\{\mathcal{F}(t)\}$. Indeed, since all beliefs implied by likelihoods $h(\theta) \in \Xi(\theta)$ are absolutely continuous, all relevant processes $(\widetilde{B}_D^h, \widetilde{B}_e^h)'$ generate the same filtration $\{\mathcal{F}(t)\}$ and the dynamic budget constraint associated with problem (10) can be equivalently formulated in terms of $m_{\theta,h}$ and $(\widetilde{B}_D^h, \widetilde{B}_e^h)'$. See also Miao (2001) for a related discussion.

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

$$(P) \quad : \quad J(\Pi, D) = \inf_{h(\theta)} E\left[\int_0^\infty e^{-\delta t}\frac{D(t)^{1-\gamma}}{1-\gamma}dt \middle| \mathcal{F}(0)\right] \quad , \tag{23}$$

subject to the dynamics

$$dD = m_{\theta,h} D dt + \sigma_D D d\widetilde{B}_D^h \quad , \tag{24}$$

$$d\pi_i = \pi_i \left(\theta_i + h\left(\theta_i\right)\sigma_D - m_{\theta,h}\right)\left(k_D d\widetilde{B}_D^h + k_e d\widetilde{B}_e^h\right) \quad , \tag{25}$$

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

The key difference with a standard (single-likelihood) equilibrium Bayesian setting of learning is that in equation (23) investors must select optimally a worst-case forecast procedure for the unknown dividend drift. Such a worst-case belief selection generates an endogenous systematic discrepancy between the reference model belief and the one applied by investors to value risky assets under ambiguity. The worst-case belief selection affects investors' relevant belief for pricing future asset pay-offs by enforcing a conservative max-min utility behavior. Therefore, in equilibrium the worst-case belief selection has a direct impact on the level of asset prices. However, a more indirect effect arises for the equilibrium equity-return dynamics under the reference model belief. The systematic bias between the reference model belief and investors' worst-case belief affects reference model equilibrium quantities in a nontrivial way.

In Proposition 1, we study the direct impact of ambiguity aversion on the price of equity and equilibrium interest rates. We do so by presenting the solution to Problem (23).

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

$$\widehat{\theta}_i + (1-\gamma)\sqrt{2\eta\,(\theta_i)}\sigma_D > 0 \quad , \quad i = 1,..,n \quad . \tag{26}$$

Then, we have:

a) The normalized misspecification $h^*(\theta)$ *solving Problem (23) is given by*

$$h^*(\theta_i) = -\sqrt{2\eta\,(\theta_i)} \quad , \quad i = 1,..,n \quad . \tag{27}$$

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

$$P\,(\Pi, D) = D\sum_{i=1}^{n}\pi_i C_i \ , \tag{28}$$

where

$$C_i = 1/(\widehat{\theta}_i + (1-\gamma)\sqrt{2\eta\,(\theta_i)}\sigma_D) \quad , \quad i = 1,..,n \quad . \tag{29}$$

c) The equilibrium interest rate r *is*

$$r = \delta + \gamma m_{\theta,h^*} - \frac{1}{2}\gamma\,(\gamma+1)\,\sigma_D^2 \quad , \tag{30}$$

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

$$m_{h^*(\theta)} = \sum_{i=1}^{n} h^*(\theta_i)\, \pi_i \sigma_D = -\sum_{i=1}^{n} \sqrt{2\eta(\theta_i)}\pi_i\sigma_D \quad . \tag{31}$$

Each constant of the form in equation (29) is proportional to investors' expectation of discounted lifetime dividends, conditional on a constant dividend-drift process $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$. The drift process $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$ is the worst-case drift misspecification $\theta_i + h^*(\theta_i)\,\sigma_D$ selected from the neighborhood $\Xi(\theta_i)$. More specifically, we have

$$C_i = E^{h^*(\theta_i)}\left[\int_s^\infty e^{-\delta(t-s)}\left(\frac{D(t)}{D(s)}\right)^{1-\gamma} dt\right] = \frac{1}{D(s)} E^{h^*(\theta_i)}\left[\int_s^\infty \frac{u_c(D(t), t)}{u_c(D(s), s)} D(t)\, dt\right] \quad , \tag{32}$$

where $E^{h^*(\theta_i)}[\cdot]$ denotes expectations 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 investors are willing to pay a high price for the ambiguous state $\Xi(\theta_i)$. Since the state is not observable, they weigh each C_i by the posterior probability π_i to get the price in equation (28) of the risky asset under learning and ambiguity aversion. We note that C_i is a function of both investors' ambiguity aversion, via the parameter $\eta(\theta_i)$, and investor's relative risk aversion γ.

We can also write equation (32) as

$$C_i \quad = \quad E\left[\int_s^\infty e^{-\left[\delta + (1-\gamma)\sqrt{2\eta(\theta_i)}\sigma_D\right](t-s)}\left(\frac{D(t)}{D(s)}\right)^{1-\gamma} dt | \theta = \theta_i\right] \quad , \tag{33}$$

where $E[\cdot|\theta = \theta_i]$ denotes reference model expectations conditional on a constant drift $\theta = \theta_i$. Therefore, the impact of ambiguity aversion on the price of the ambiguous state $\Xi(\theta_i)$ is equivalent to that implied by a corrected time preference rate

$$\delta \longrightarrow \delta + (1-\gamma)\sqrt{2\eta(\theta_i)}\sigma_D \tag{34}$$

under the reference model dynamics. The adjustment in equation (34) depends on the amount of ambiguity of the ambiguous state $\Xi(\theta_i)$, relative risk-aversion γ and dividend-growth volatility σ_D.

Cagetti et al. (2002) give numerical evidence that 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 setting, equation (34) implies that, given a homogeneous degree of ambiguity, the effect of ambiguity aversion is exactly offset by an increase in the subjective time preference rate. However, in the general case of an heterogeneous degree of ambiguity $\eta(\theta)$, we cannot map the final effect of ambiguity on equity prices into an adjustment of one single time preference rate. As we show in Section 4, the additional flexibility implied by heterogeneous ambiguity structures can generate model predictions that are consistent with the well-know puzzles even more easily than can homogeneous ambiguity structures.

Corollary 3 summarizes the dependence of the price C_i on an arbitrary ambiguity parameter $\eta(\theta_i)$.

Corollary 3 *The price of any ambiguous state $\Xi(\theta)$ is a decreasing function in the degree of ambiguity $\eta(\theta)$ if and only if $\gamma < 1$. In such a case, C_i is a convex function of $\eta(\theta_i)$, which is uniformly more convex for smaller risk aversion γ.*

From Corollary 3, the marginal relative price of ambiguity is negative if and only if relative risk aversion γ is less than 1. In the opposite case, if $\gamma > 1$, we obtain the somewhat counterintuitive implication, relative, e.g., to the basic intuition provided by the standard (static) Ellsberg (1961) paradox, that the price of an ambiguous state is higher than the one of an unambiguous one.

To understand this only apparently paradoxical finding, recall that in the determination of C_i the representative investor discounts worst-case future dividends through their marginal utility. In equilibrium, a lower dividend growth rate implies a lower expected future consumption growth and a lower discount rate. Since for high risk aversion the last effect dominates, a lower expected dividend growth deriving from a conservative belief under ambiguity implies a lower discount rate and a higher price for ambiguous states.

For $\gamma > 1$, settings of learning and ambiguity aversion with high risk aversion deliver low (negative) equity premia and low volatilities, together with high and highly variable interest rates. That is, imposing high risk aversion worsens the asset pricing puzzles when learning under ambiguity aversion is considered. Therefore, we focus in the sequel on settings with moderate risk aversion. 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 such experiments, utility functions estimated by procedures that are not robust to the presence of ambiguity are clearly concave.

Assumption 3 *The representative agent in the model has a relative risk-aversion parameter* $\gamma < 1$.

Since we adopt a setting with power utility of consumption, Assumption 3 is equivalent to assuming an elasticity of intertemporal substitution (EIS) $1/\gamma > 1$. This is perfectly consistent with the idea that in our model excess returns are going to reflect mainly some premium for ambiguity, rather than a premium for risk.

The empirical evidence about the size of the EIS is mixed. Hansen and Singleton (1982) and Attanasio and Weber (1989) estimated the EIS to be well above one. Hall (1988) considered aggregation effects and estimated an EIS well below one using aggregate consumption data. Similar low estimates using aggregate consumption variables are obtained in Campbell (1999). Recent empirical work focusing on the consumption of households participating in the stock or the bond market has suggested that such investors have a much larger EIS than individuals that do not hold stocks or bonds. For instance, Vissing-Jorgensen (2002) estimates an EIS well above one for individuals holding portfolios of stocks and bonds in Euler equations for treasury bills. Attanasio and Vissing-Jorgensen (2003) also estimate large EIS for stockholders when using Euler equations for treasury bills and after-tax returns. Attanasio, Banks, and Tanner (2002) findings on UK data suggest an EIS larger than one for Euler equations including treasury bills and equity returns in an econometric model where ownership probabilities are also estimated. Finally, Aït-Sahalia, Parker, and Yogo (2004) estimate EIS above one using Euler equations for treasury bills where consumption is measured by consumption of luxury goods. Typically, in these empirical studies the EIS estimated with Euler equations including US equity returns are lower. However, as noted for instance by Vissing-Jorgensen (2002, p. 840), Attanasio and Vissing-Jorgensen (2003, p. 387) and Aït-Sahalia, Parker, and Yogo (2004, p. 2985), this finding is mainly due to the low predictive power of the instruments for equity returns, which leads to poor finite sample properties of the estimators.

The results in the above literature are based on models that do not explicitly account for fluctuating economic uncertainty. Recently, Bansal and Yaron (2004) argued in a setting with Epstein and Schneider (1989) preferences and fluctuating uncertainty that a model with EIS above one can explain better key asset markets phenomena than a model with EIS below one. Moreover, they showed that neglecting fluctuating economic uncertainty leads to a severe downward bias in the estimated EIS using standard Euler equations. In our setting fluctuating economic uncertainty arises endogenously, via the learning process of our representative agent. Therefore, downward biases in EIS estimates similar to those noted in Bansal and Yaron (2004) will arise. Such biases are particularly large for Euler equations using equity returns; see Section 4.4 below.

3.2.1. Price/Dividend Ratios and Interest Rates

Under Assumption 3, we obtain from Proposition 1 a few implications for the behavior of the price/dividend ratio P/D in the model. We summarize them in Corollary 4.

Corollary 4 *Under Assumption 3 we have the following:*

a) *The price/dividend ratio P/D is a decreasing convex function of the amount of ambiguity $(\eta(\theta_1), .., \eta(\theta_n))$ in the economy. Moreover, P/D is a uniformly more convex function for lower risk aversion γ.*

b) A mean-preserving spread $\widehat{\Pi}$ *of* Π *implies*

$$\widehat{P}/D > P/D \quad ,$$

that is, the price/dividend ratio P/D *is increasing in the amount of uncertainty in the economy.*

Finding *a)* in Corollary 4 is a direct implication of (28) and (33). Finding *b)* follows from the convexity of C_i in (28) as a function of $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$.

Under Assumption 3, the impact of a higher ambiguity on price/dividend ratios (Finding *a)*) has a different sign than the one of a higher uncertainty in the economy (Finding *b)*). This is a distinct prediction of ambiguity aversion for the behavior of P/D.

In Proposition 1, the equilibrium interest rate is given by equation (30). The effect of learning and ambiguity aversion on equilibrium interest rates is always negative, since r is a decreasing convex function of $(\eta(\theta_1), .., \eta(\theta_n))$. We obtain the special case of an equilibrium interest rate r_{NA} under no ambiguity, as in Veronesi (2000), by setting $\eta(\theta) = 0$ for all $\theta \in \Theta$ in (30),

$$r_{NA} = \delta + \gamma m_\theta - \frac{1}{2}\gamma(\gamma+1)\sigma_D^2 \quad . \tag{35}$$

Hence,

$$r - r_{NA} = \gamma m_{h^*(\theta)} < 0 \Longleftrightarrow \text{there exists } \theta \in \Theta \text{ such that } \eta(\theta) > 0 \quad . \tag{36}$$

The case with no uncertainty about the true model neighborhood arises under a degenerate distribution Π, implying $m_\theta + m_{h^*(\theta)} = \theta_l - \sqrt{2\eta(\theta_l)}\sigma_D$ for some $\theta_l \in \Theta$ and

$$r = \delta + \gamma\left(\theta_l - \sqrt{2\eta(\theta_l)}\sigma_D\right) - \frac{1}{2}\gamma(\gamma+1)\sigma_D^2 \quad . \tag{37}$$

The interest rate in equation (37) is the equilibrium interest rate of an economy with ambiguity but no learning. Hence, even in the case of an asymptotic learning about $\Xi(\theta_l)$, the asset-pricing impact of ambiguity on interest rates does not disappear. Asymptotically, the representative agent still has ambiguity about the unknown drift $\theta_l + h(\theta_l)\sigma_D \in \Xi(\theta_l)$ that generates the dividend dynamics, even if she learns that the relevant model neighborhood is $\Xi(\theta_l)$.

For the case in which the asymptotic distribution of Π is nondegenerate, the contribution $m_{h^*(\theta)}$ of ambiguity aversion to the level of interest rates is a weighted sum of the contributions of ambiguity aversion under the single-model neighborhoods $\Xi(\theta_1)$,.., $\Xi(\theta_n)$. The weights in $m_{h^*(\theta)}$ are given by the posterior probabilities Π. Moreover, the dynamics of Π depend on $\eta(\theta)$, i.e., on the worst-case likelihood $h^*(\theta)$, that has been optimally selected by the ambiguity-averse investor. Therefore, time-varying aggregate ambiguity arises fully endogenously, because its Π–dependent dynamics cannot be determined regardless of the ambiguity parameter η in the economy.

The impact of ambiguity aversion on interest rates and price/dividend ratios is illustrated numerically in Table 1 for a setting of no learning (column NL) and a setting with learning (column L). The first row ($\eta = 0$) in Table 1 presents quantities prevailing in the absence of ambiguity. The following rows ($\eta = 0.001, 0.005, 0.01, 0.02$) give equilibrium interest rates and price/dividend ratios for an increasing (homogenous) ambiguity parameter η.

Insert Table 1 about here

For a moderate risk aversion parameter $\gamma = 0.5$, low interest rates of about 2.8% are obtained, even in the absence of ambiguity ($\eta = 0$), because the elasticity of intertemporal substitution $1/\gamma$ implied by our setting is sufficiently large. Increasing ambiguity aversion lowers interest rates further to a level of about 2.45% for $\eta = 0.02$. Interest rates under learning (column L) and no learning (column NL) are identical because r is a linear function of posterior probabilities Π and we have chosen Π to be symmetric around the true underlying dividend drift θ. Due to the convexity of C_i as a function of $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$, price/dividend ratios are always higher in a setting of learning, compared to a model without prior uncertainty about the underlying dividend drift. However, ambiguity aversion lowers $P/D-$ratios monotonically, from a level of about 96 in the absence of ambiguity (column L for $\eta = 0$) to a level of about 69 (column L for $\eta = 0.02$).

3.2.2. Endogenous Learning Dynamics

The normalized worst-case drift distortion in equation (27) of Proposition 1 determines the description of the endogenous relevant $\Pi-$dynamics under ambiguity aversion. We focus on a description under the reference model dynamics from the perspective of an outside observer who knows that the dividend dynamics indeed satisfies the reference model in (2) and (3), and who also knows the specific value of the parameter θ.

Despite the fact that the true dynamics are those under the reference model, misspecification doubts coupled with ambiguity aversion force investors to follow different learning dynamics than the optimal Bayesian learning dynamics under the reference model's likelihood. We highlight this issue in the next Corollary.

Corollary 5 *Under the reference model in Definition 1, the filtered probabilities dynamics of a representative agent solving the equilibrium optimization problem (23) are*

$$d\pi_i = \pi_i \left(\theta_i - \sqrt{2\eta\left(\theta_i\right)}\sigma_D - m_{\theta,h^*}\right)\left[k\left(\theta - m_{\theta,h^*}\right)dt + k_D dB_D + k_e dB_e\right] \tag{38}$$

Equation (38) gives us a way to study the learning dynamics realized under ambiguity aversion. We observe that ambiguity aversion can imply a tendency to overstate the probability of good states, relatively to the probabilities implied by the learning dynamics of a Bayesian investor. To emphasize this point, we consider the case of a constant ambiguity aversion $\eta\left(\theta_1\right) = ... = \eta\left(\theta_n\right) = \eta$, implying filtered probability dynamics given by

$$d\pi_i = \pi_i\left(\theta_i - m_\theta\right)\left[k\left(\theta - m_\theta + \sqrt{2\eta}\sigma_D\right)dt + k_D dB_D + k_e dB_e\right] \quad . \tag{39}$$

For $\eta = 0$, the dynamics in equation (39) are those of a standard (single-likelihood) Bayesian learner. More precisely, the difference in the drift in (39) with and without ambiguity ($\eta \neq 0$ and $\eta = 0$) is given by

$$k\pi_i\left(\theta_i - m_\theta\right)\sqrt{2\eta}\sigma_D dt. \tag{40}$$

The difference is positive for above-average candidate reference model drifts $\theta_i \in \Theta$ ($\theta_i - m_\theta > 0$) and negative for below-average candidate reference model drifts $\theta_i \in \Theta$ ($\theta_i - m_\theta < 0$). Therefore, investors subject to ambiguity aversion will tend to "learn" a large reference model drift more rapidly than a low reference model drift. Unconditionally, this property implies learning dynamics where the a posteriori expected reference model drift m_θ under ambiguity aversion is higher than that of a Bayesian investor, i.e., the learning dynamics under ambiguity aversion implies an optimistic tendency to overstate the a posteriori reference model drifts relatively to a standard Bayesian prediction. Such a tendency is more apparent for large precision parameters k.

Figure 2 illustrates these features for a setting with three possible neighborhoods $\Xi(\theta_1)$, $\Xi(\theta_2)$, $\Xi(\theta_3)$. We plot the posterior probabilities π_1 implied by Corollary 5 for the "bad" state $\Xi(\theta_1)$ in Panel A and those for the good state $\Xi(\theta_3)$ (the probabilities π_3) in Panel B. The "true" underlying state is $\Xi(\theta_2)$.

Insert Figure 2 about here

In Panel A, the uniformly higher probabilities π_1 arise in the absence of ambiguity (the solid line corresponds to $\eta = 0$), while for the largest ambiguity-aversion parameter $\eta = 0.05$ the uniformly lowest posterior probabilities arise. Hence, the ambiguity-averse investor systematically understates the probability of the "bad" state θ_1. Similar features, but in the opposite direction, arise for the probabilities π_3 of the "good" state θ_3 in Panel B.

4. Conditional Asset Returns

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

$$dR = \frac{dP + Ddt}{P} - rdt \quad . \tag{41}$$

We first study the direct effect of learning and ambiguity aversion on $R-$dynamics by describing this effect with respect to the filtered Brownian motions $\widetilde{B}_D^{h^*}, \widetilde{B}_e^{h^*}$, which are implied by the selected optimal worst-case likelihood belief $h^*(\theta)$ of Proposition 1. This description provides the dynamics of R under the worst-case scenario $h^*(\theta) \in \Xi(\theta)$ in our economy. In this sense, we can interpret the resulting expected excess return on equity as the worst-case equity premium in the economy.

The indirect impact of learning and ambiguity aversion on $R-$dynamics arises because of the differences between the likelihood belief under the reference model dynamics and the optimal worst-case belief adopted by ambiguity-averse investors in computing asset prices. Under the reference model likelihood belief, such a discrepancy determines an additional ambiguity premium component for misspecification in the $R-$dynamics. We can analyze this important effect of learning and ambiguity aversion, by describing $R-$dynamics with respect to the filtered Brownian motions $\widetilde{B}_D, \widetilde{B}_e$, which are implied by the reference model belief for dividends in Definition 1. This description provides the correct $R-$dynamics from the perspective of an outside observer (e.g., an econometrician), who believes in the reference model of Definition 1 as an approximate description of the dividend dynamics and knows that investors in the economy are ambiguity averse. The resulting expected excess return on equity identifies the structure of equity premia under learning and ambiguity aversion. Proposition 2 summarizes our findings.

Proposition 2 *(i) Under the investor's subjective optimal worst-case belief $h^*(\theta)$ in Proposition 1, the equilibrium return process R under ambiguity aversion has dynamics*

$$dR = \mu_R^{wc} dt + \sigma_D d\widetilde{B}_D^{h^*} + V_{\theta,h^*} \left(k_D d\widetilde{B}_D^{h^*} + k_e d\widetilde{B}_e^{h^*} \right) \quad , \tag{42}$$

where

$$\mu_R^{wc} = \gamma \left(\sigma_D^2 + V_{\theta,h^*} \right) \quad , \quad V_{\theta,h^*} = \sum_{i=1}^{n} \frac{\pi_i C_i \left(\theta_i - \sqrt{2\eta\left(\theta_i\right)} \sigma_D \right)}{\sum_{i=1}^{n} \pi_i C_i} - m_{\theta,h^*} \quad , \tag{43}$$

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

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

(ii) Under the reference model belief, the equilibrium excess return process R under ambiguity aversion has dynamics

$$dR = \mu_R dt + \sigma_D d\widetilde{B}_D + V_{\theta,h^*} \left(k_D d\widetilde{B}_D + k_e d\widetilde{B}_e \right) \quad , \tag{44}$$

where

$$\mu_R = \mu_R^{wc} - m_{h^*(\theta)} \left(1 + k V_{\theta,h^*} \right) \quad ,$$

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

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

We can analyze in more detail how learning under ambiguity aversion affects the conditional structure of asset returns. We first study the sign and comparative statics for quantities $m_{h^*(\theta)}$ and V_{θ,h^*} arising in equations (42) and (44). In a second step, we discuss the impact of learning under ambiguity aversion on equity premia and volatilities. The term

$$m_{h^*(\theta)} = m_{\theta,h^*} - m_\theta = -\sum_{i=1}^{n} \sqrt{2\eta(\theta_i)} \pi_i \sigma_D \tag{45}$$

is a conservative correction to the reference model's a posteriori expectations m_θ. This correction accounts for misspecification doubts in the a posteriori expectations for the growth rate of the economy and is always negative. The term V_{θ,h^*} reflects the difference between the worst-case expected growth rate of the economy, m_{θ,h^*}, and its value-adjusted counterpart. V_{θ,h^*} is larger either when agents have more diffuse beliefs about $\Xi(\theta_1), .., \Xi(\theta_n)$, or when they value the asset very differently across the different states. These differences in valuation depend on the heterogeneity of the worst-case growth rate $\theta - \sqrt{2\eta(\theta)}\sigma_D$ across such states. Under Assumption 2, it follows

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

Therefore, we can use similar arguments as in the proof of Lemma 3 in Veronesi (2000) to obtain the following characterization of V_{θ,h^*} in our setting of learning under ambiguity aversion.

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

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*

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

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

In particular, quantity V_{θ,h^*} is positive and increasing with respect to mean-preserving spreads of Π if and only if $\gamma < 1$. The positivity of V_{θ,h^*} is crucial to avoid theoretical asset pricing relations that are clearly inconsistent with the equity premium puzzle predictions.

To study the impact of ambiguity aversion on V_{θ,h^*}, we compute comparative statics for the standardized worst-case drift ambiguity quantities $\sqrt{2\eta(\theta_1)}, .., \sqrt{2\eta(\theta_n)}$ in a neighborhood of $\eta(\theta) = 0$, i.e., the pure Bayesian learning setting.

Proposition 3 *(i) The comparative statics of* V_{θ,h^*} *for the ambiguity parameter* $\sqrt{\eta(\theta_i)}$ *are*

$$\left.\frac{\partial V_{\theta,h^*}}{\partial\sqrt{2\eta(\theta_i)}}\right|_{\eta(\theta)=0} = -\left[\frac{\pi_i C(\theta_i)}{\sum_{j=1}^n \pi_j C(\theta_j)} - \pi_i + (1-\gamma)\frac{\pi_i C(\theta_i)^2}{\sum_{j=1}^n \pi_j C(\theta_j)}(\theta_i - m_\theta - V_\theta)\right]\sigma_D \quad ,$$

where for any $i = 1, .., n$ coefficient $C(\theta_i)$ is the value of C_i in Proposition 1 for $\eta(\theta_i) = 0$ and V_θ is the value of V_{θ,h^*} for $\eta(\theta_1) = ... = \eta(\theta_n) = 0$. (ii) Let Assumption 3 be satisfied. If both conditions

$$\frac{\pi_i C(\theta_i)}{\sum_{j=1}^n \pi_j C(\theta_j)} \gtrless \pi_i \quad ; \quad \theta_i - m_\theta \gtrless V_\theta \tag{46}$$

are satisfied, then

$$\left. \frac{\partial V_{\theta,h^*}}{\partial \sqrt{2\eta(\theta_i)}} \right|_{\eta(\theta)=0} \lesseqgtr 0 \quad . \tag{47}$$

(iii) Let Assumption 3 be satisfied. If ambiguity is homogeneous ($\eta(\theta_i) = \eta$ for $i = 1, .., n$), it follows

$$\left. \frac{\partial V_{\theta,h^*}}{\partial \sqrt{2\eta}} \right|_{\eta=0} < 0 \quad . \tag{48}$$

To understand the meaning of Proposition 3 and condition (46), define for any $i = 1, .., n$ the value-adjusted probability of the reference model drift θ_i in the absence of ambiguity,

$$\widetilde{\pi}_i = \frac{\pi_i C(\theta_i)}{\sum_{j=1}^n \pi_i C(\theta_j)} \quad . \tag{49}$$

Then, we can rewrite equation (46) as

$$\widetilde{\pi}_i \gtrless \pi \quad ; \quad \theta_i - m_\theta \gtrless \widetilde{m}_\theta - m_\theta \quad , \tag{50}$$

where "$\sim$" denotes quantities under $\widetilde{\Pi}$. Condition (50) is intuitive. It requires that the value-weighted probability $\widetilde{\pi}_i$ of θ_i and the reference model drift θ_i itself are larger (or smaller) than, respectively, the posterior probability π_i and the $\widetilde{\Pi}-$value weighted mean of θ. If equation (50) is valid with ">", then adding ambiguity to state θ_i implies, under Assumption 3, $V_{\theta,h^*} < V_\theta$. The opposite holds if (50) is valid with "<". Under Assumption 3, $C(\theta_i)$ in an increasing convex function of θ_i. Therefore, equation (50) will tend to hold with ">" for large values of θ_i and with "<" for low values of θ_i. (For instance, under Assumption 3 condition (50) is always satisfied with "<" by θ_1 and with ">" by θ_n.) Inequality (47) of Proposition 3 then implies that asymmetric ambiguity structures $\eta(\theta)$ tend to decrease (increase) V_{θ,h^*}, when ambiguity is sufficiently large for high (low) reference model drift states. However, when ambiguity is homogeneous, no increase in V_{θ,h^*} arises when we extend the Bayesian learning setting to incorporate ambiguity aversion.

4.1. Equity Premia

From equation (43), the equilibrium equity premium μ_R is given by

$$\mu_R = \underbrace{\gamma(\sigma_D^2 + V_{\theta,h^*})}_{\text{(A)}} \quad \underbrace{-m_{h^*}}_{\text{(B)}} \quad \underbrace{-m_{h^*}kV_{\theta,h^*}}_{\text{(C)}} \quad . \tag{51}$$

μ_R is the sum of three conceptually different equity premium contributions (A), (B), and (C). (A) is the equity premium part deriving from standard risk exposure, i.e., the standard risk premium. It can also be interpreted as the worst-case equity/risk premium in our economy. The sum (B)+(C) is the equity premium part caused by exposure to ambiguity, i.e., the ambiguity premium. (B) is the part of the ambiguity premium caused by misspecifications in the dividend dynamics. (C) is the part of the ambiguity premium caused by misspecifications in the dynamics for the posterior probabilities Π.

Figure 3, Panel D, presents a typical pattern for the equity premium μ_R for different levels of the risk-aversion parameter and under a homogeneous degree of ambiguity $\eta(\theta) = \eta = 0.01$.

Insert Figure 3 about here

In Figure 3, the equity premium μ_R is a monotonically decreasing function of risk aversion. However, for low risk aversion, such a feature seems to be very compatible with the predictions of the equity premium puzzle. For instance, for moderate risk aversion γ between 0.2 and 0.4, the equity premium ranges from about 8% to about 5%. This effect arises despite the small size of the ambiguity parameters used.

4.1.1. Premia for Risk

The term (A) in equation (51) is the equity premium perceived by an investor under the optimal worst-case likelihood $h^*(\theta)$ selected in Proposition 1. More precisely, from Proposition 2 we have

$$\mu_R^{wc} = \gamma\left(\sigma_D^2 + V_{\theta,h^*}\right) = \gamma Cov_t^{h*}(dR, dD/D) = \gamma Cov_t(dR, dD/D) \quad , \tag{52}$$

where $Cov_t^{h^*}$ (Cov_t) denotes conditional covariances under the worst-case likelihood h^* (under the reference model likelihood) and the last equality arises because worst-case and reference model likelihoods are absolutely continuous. Therefore, the term (A) has the dual interpretation of being the total equity/risk premium arising under the worst-case likelihood belief, and the part of equity premium deriving from pure risk exposure under the reference model likelihood belief.

In particular, equation (52) emphasizes the fact that under learning and ambiguity aversion, the covariance term $Cov_t(dR, dD/D)$ captures only the fraction (A) of the whole equity premium μ_R under the reference model dynamics.

It is natural to expect that the risk premium (A) will be actually quite small in our economy. Indeed, $\gamma\sigma_D^2$ is the risk premium under ambiguity aversion but no learning (see, e.g., Maenhout (2004) and Trojani and Vanini (2002)). It is increasing in γ, but for realistic risk-aversion parameters it is typically a very small number. Moreover, Lemma 2 implies that the quantity V_{θ,h^*} is decreasing in risk aversion. Therefore, the term (A) as a function of risk aversion is bounded and is negligible for practical purposes.

In Panel C of Figure 3, we plot a typical profile of the risk premium (A) as a function of γ (circled curve) together with the equity/risk premium function implied by a setting of pure Bayesian learning (crossed curve). The equity/risk premium function prevailing in Panel C under a pure Bayesian learning setting almost coincides with the risk premium function under ambiguity aversion. In general, the risk premium (A) under learning and ambiguity aversion is different from the risk premium arising under pure learning, because the term V_{θ,h^*} in equation (52) is different from the corresponding term that prevails in a setting of pure learning when $\eta(\theta) = 0$. However, for realistic structures of the ambiguity function $\eta(\theta)$, we always find the two risk premia to be numerically similar. The Mehra and Prescott (1985) equity premium puzzle is even more puzzling in a purely Bayesian setting, because equity premia cannot be matched by risk premia, even for very high risk aversion. However, the equity premium in equation (51) under learning and ambiguity aversion also consists of the ambiguity premium (B)+(C). As we show below, this component is crucial for obtaining model predictions that are consistent with the equity premium puzzle.

4.1.2. Premia for Ambiguity

Under ambiguity aversion, the equity premium in equation (51) depends on the ambiguity premia (B) and (C), which are both positive under Assumption 3. The sum of (B)+(C) represents a premium for ambiguity in the reference model dynamics that derives from the discrepancy between the reference model likelihood belief and the worst-case likelihood belief optimally selected by the ambiguity-averse representative investor. More specifically, recall that the worst-case return dynamics (42) depends on two filtered random shocks

$$d\widetilde{B}_D^{h^*} = d\widetilde{B}_D - k_D m_{h^*(\theta)} dt \quad , \quad d\widetilde{B}_e^{h^*} = d\widetilde{B}_e - k_e m_{h^*(\theta)} dt \quad . \tag{53}$$

By construction, $(\widetilde{B}_D^{h^*}, \widetilde{B}_e^{h^*})$ is a $\{\mathcal{F}(t)\}$–Brownian motion under the worst-case likelihood belief $h^*(\theta)$, but it is a $\{\mathcal{F}(t)\}$–Brownian motion with drift under the reference model likelihood belief. The differences

$$\lambda_D^A \;:=\; (d\widetilde{B}_D^{h^*} - d\widetilde{B}_D)/dt = -k_D m_{h^*(\theta)} \quad , \tag{54}$$

and

$$\lambda_e^A \;:=\; (d\widetilde{B}_e^{h^*} - d\widetilde{B}_e)/dt = -k_e m_{h^*(\theta)} \quad , \tag{55}$$

are the market prices of ambiguity for $d\widetilde{B}_D$ and $d\widetilde{B}_e$ shocks, respectively, that prevail under the reference model belief. Such market prices of ambiguity arise because ambiguity-averse investors apply a worst-case learning approach to price assets. Such an approach systematically understates the expected dividend drift prevailing under the reference model. The filtered shocks $d\widetilde{B}_D^{h^*}$ and $d\widetilde{B}_e^{h^*}$ influence the worst-case dynamics in equation (42) in two distinct ways: through the isolated impact of $d\widetilde{B}_D^{h^*}$ on the worst-case filtered dynamics for dividends and through the joint impact of $d\widetilde{B}_D^{h^*}$ and $d\widetilde{B}_e^{h^*}$ on the worst-case filtered dynamics for the posterior probabilities Π. More precisely, by setting $h = h^*$ in equations (24) and (25) we obtain, for the optimal joint (D, Π) –filtered dynamics:

$$
\begin{aligned}
dD/D &= m_{\theta,h^*}dt + \sigma_D d\widetilde{B}_D^{h^*} \ , && (56)\\
d\pi_i/\pi_i &= (\theta_i + h(\theta_i)\sigma_D - m_{\theta,h^*})\left(k_D d\widetilde{B}_D^{h^*} + k_e d\widetilde{B}_e^{h^*}\right) \quad ; \quad i = 1, .., n \quad . && (57)
\end{aligned}
$$

The ambiguity premium for exposure to shocks in any of the π_i dynamics is

$$
k_D(d\widetilde{B}_D^{h^*} - d\widetilde{B}_D)/dt + k_e(d\widetilde{B}_e^{h^*} - d\widetilde{B}_e)/dt = k_D\lambda_D^A + k_e\lambda_e^A = -km_{h^*(\theta)} \quad , \tag{58}
$$

where $k = k_D^2 + k_e^2$. From the worst-case R–dynamics,

$$
dR \quad = \quad \mu_R^{wc}dt + \sigma_D d\widetilde{B}_D^{h^*} + V_{\theta,h^*}\left(k_D d\widetilde{B}_D^{h^*} + k_e d\widetilde{B}_e^{h^*}\right) \quad , \tag{59}
$$

we see that ambiguity premia for shocks $d\widetilde{B}_D^{h^*}$ in the D–dynamics are multiplied by dividend volatility σ_D. Similarly, ambiguity premia for shocks $k_D d\widetilde{B}_D^{h^*} + k_e d\widetilde{B}_e^{h^*}$ in the Π–dynamics are multiplied by V_{θ,h^*}. These arguments imply

$$
-m_{h^*(\theta)} = \sigma_D\lambda_D^A \quad ; \quad -m_{h^*(\theta)}kV_{\theta,h^*} = V_{\theta,h^*}(k_D\lambda_D^A + k_E\lambda_e^A) \quad , \tag{60}
$$

i.e., the equity premium component (B) in equation (51) is the equilibrium ambiguity premium for misspecification in the D–dynamics, and the equity premium component (C) in equation (51) is the equilibrium ambiguity premium for misspecification in the Π–dynamics .

Component (B) is also nonzero under a degenerate Π–distribution, i.e., in the absence of learning. Under a nondegenerate Π–distribution, it is affected by Π only when ambiguity $\eta(\theta)$

is not homogeneous and to the extent that Π affects the posterior mean of $\sqrt{\eta(\theta)}\sigma_D$. Indeed, under a homogeneous ambiguity $\eta(\theta) = \eta > 0$ it follows

$$-m_{h^*(\theta)} = \sqrt{2\eta}\sigma_D, \tag{61}$$

i.e., the ambiguity premium in a full information economy with ambiguity-averse agents, as obtained, e.g., in Trojani and Vanini (2002). Therefore, we interpret component (B) as a pure premium for ambiguity. As becomes obvious from (61), we note that even in the absence of learning the equity premium can be made large by increasing the parameter η. However, to accommodate for a reasonable equity premium, we would have to make the uncertainty parameter η unreasonably large and thereby impose an excessive degree of pessimism. Only the simultaneous modeling of learning and ambiguity aversion allows us to generate a substantial equity premium with a moderate and reasonable amount of ambiguity.

In contrast to the findings in the risk premia analysis above, the contribution of the ambiguity premium component (B) to the equity premium is given by a first-order effect of ambiguity that is proportional to dividend volatilities σ_D. Moreover, since (B) is nonzero even for degenerate Π–distributions, it will not disappear asymptotically, even in the case of an asymptotic learning about the underlying true model neighborhood $\Xi(\theta)$. In contrast to (B), the ambiguity premium part (C) is zero under a degenerate Π–distribution, implying that the asymptotic level of the ambiguity premium in the case of asymptotic learning is fully determined by component (B). Such an asymptotic ambiguity premium rewards the representative agent for the residual ambiguity about the precise drift that generated the dividend dynamics, out of a relevant neighborhood $\Xi(\theta_l)$. In such a case, we can obtain the ambiguity premium (61) in a full information economy with ambiguity aversion as the limit of a sequence of ambiguity premia in partial information economies with ambiguity averse agents.

Under Assumption 3, component (C) in equation (51) is nonzero if and only if Π is nondegenerate, i.e., if ambiguity-averse investors did not yet fully learn the underlying model neighborhood $\Xi(\theta)$. Therefore, we can interpret (C) as an ambiguity premium component caused by the joint presence of ambiguity aversion and learning.

To illustrate the contribution of learning and ambiguity to risk and equity premia, we compute these quantities in Table 2 for a setting of no learning (column NL) and a setting without learning (column L). The first row ($\eta = 0$) in Table 2 presents quantities prevailing in the absence of ambiguity. The following rows ($\eta = 0.001, 0.005, 0.01, 0.02$) give equilibrium risk and equity premia (column RP and EP) for an increasing (homogenous) ambiguity parameter η.

Insert Table 2 about here

For risk aversion $\gamma = 0.5$, we obtain in column RP of Table 2 tiny risk premia, i.e., component (A) in the total equity premium, both for a setting with and without learning. For a setting of no learning and no ambiguity, very high risk aversions are needed to generate sizable risk premia. As shown in Veronesi (2000), in a setting of learning without ambiguity, risk premia are tiny also for very large risk aversions. Therefore, the equity premium is even more of a puzzle in such economies. In column EP of Table 2 we present equity premia. In the absence of ambiguity ($\eta = 0$), risk and equity premia are identical. Introducing ambiguity in the model increases equity premia both for a setting of no learning (column NL) and a setting of learning (column L). From column NL, we observe a premium for pure ambiguity that increases equity premia from about 0.07% ($\eta = 0$) to about 0.60% ($\eta = 0.01$). In relative terms, the increase in the equity premium is substantial. However, the resulting equity premia are still too small for practical purposes. With $\eta = 0.01$, the pure ambiguity premium (B) is only about 0.53%. Column L, instead, presents large equity premia increasing from about 0.2% ($\eta = 0$) to about 13.4% ($\eta = 0.01$). The total ambiguity premium (B)+(C) is about 13.2%, indicating a large premium (C) for learning and ambiguity of about 12.67% when $\eta = 0.01$.

From Lemma 2, the ambiguity premium (C) is positive and decreasing in γ. Moreover, in contrast to the ambiguity premium (B), it depends on the signal precision parameter k_e. To highlight (C)'s dependence on a general ambiguity parameter vector $\eta(\theta)$, we use the second-order asymptotics provided in Lemma 3.

Lemma 3 *The following second-order asymptotics for $-m_{h^*}V_{\theta,h^*}$ around $\eta(\theta)=0$ holds*

$$\begin{aligned}
-m_{h^*}V_{\theta,h^*} &= -m_{h^*}\left[V_\theta + \sum_{i=1}^{n} \left.\frac{\partial V_{\theta,h^*}}{\partial\sqrt{2\eta(\theta_i)}}\right|_{\eta(\theta)=0}\sqrt{2\eta(\theta_i)}\sigma_D\right] + o(\|\eta(\theta)\|) \\
&= \sum_{j=1}^{n}\pi_j\sqrt{2\eta(\theta_j)}\sigma_D\left[V_\theta + \sum_{i=1}^{n} \left.\frac{\partial V_{\theta,h^*}}{\partial\sqrt{2\eta(\theta_i)}}\right|_{\eta(\theta)=0}\sqrt{2\eta(\theta_i)}\sigma_D\right] + o(\|\eta(\theta)\|)
\end{aligned} \tag{62}$$

From Lemma 3, the first-order impact of ambiguity aversion on the ambiguity premium (C) is given by the term

$$-km_{h^*}V_\theta = kV_\theta\sum_{j=1}^{n}\pi_j\sqrt{2\eta(\theta_j)}\sigma_D \quad . \tag{63}$$

Therefore, for moderate ambiguity sizes, the term m_{h^*} determines the size of (C) as a function of $\eta(\theta)$. Given a posterior distribution Π, we can expect ambiguity structures implying largest ambiguity premia (B) to imply also the largest premium component (C). Given m_{h^*}, the first-order quantitative effect of ambiguity on the ambiguity premium (C) is larger when either the risk premium in a comparable purely Bayesian economy is large (when V_θ is large), or the signal precision parameter k is larger.

From Lemma 3, the second-order effect of ambiguity on the premium (C) is given by the term

$$-km_{h^*}\sum_{i=1}^{n}\left.\frac{\partial V_{\theta,h^*}}{\partial\sqrt{2\eta(\theta_i)}}\right|_{\eta(\theta)=0}\sqrt{2\eta(\theta_i)}\sigma_D \ . \tag{64}$$

Hence, given a first-order impact of ambiguity on (C) (i.e., given m_{h^*}), the second-order effect of ambiguity is determined by its first-order effect on V_{θ,h^*}, which we have characterized by the comparative statics in Proposition 3. From Proposition 3 and the following discussion, we expect the sign of equation (64) to be negative for homogeneous ambiguity structures or for asymmetric ambiguity structures that associate a larger ambiguity with favorable economic states. Asymmetric ambiguity structures that associate a larger ambiguity with unfavorable economic states tend to imply an opposite sign in equation (64). Figure 4 illustrates these features under a parameter choice and a prior structure Π identical to that in Figure 3.

Insert Figure 4 about here

Panels A, C, E, and G present different forms for the ambiguity function $\eta(\theta)$. For comparability, we choose these functions in a way that preserves the same weighted entropy measure $\sum \pi_i \eta(\theta_i)$ as in Figure 3. Panels B, D, F, and H plot the corresponding equity premia μ_R (dashed line) and risk premia μ_R^{wc} (solid line) as a function of risk aversion $\gamma \in [0.1, 1]$.

In Figure 4, the size of $-m_{h^*}$ is 0.27%, 0.19%, 0.23%, and 0.23% in Panels B, D, F, and H, respectively, implying very small pure ambiguity premium components (B) in all panels. The risk premium component (A) is also very small. In all plots it is always below 0.25%.

In Figure 4, we show that the ambiguity premium (C) under learning and ambiguity aversion is quantitatively quite substantial, for moderate risk aversion $\gamma < 1$. For instance, in Panel H, it is above 8% for a risk aversion of about 0.2 and it is above 5% for a risk aversion of about 0.4 (dashed line). For all practical purposes, the key equity premium component is the ambiguity premium (C), i.e., the ambiguity premium part due to the joint presence of ambiguity aversion and learning.

The prior distribution Π underlying Figure 4 is the symmetric distribution plotted in Panel A of Figure 3. Such a prior distribution puts low probabilities on reference model states above or below average and higher probabilities on more central states. Therefore, in Figure 4 the pure ambiguity premium $-m_{h^*}$ is lower for ambiguity structures that put large ambiguity sizes on external reference model states (as for instance, the ambiguity structure in Panel C). According to Lemma 3 and the following discussion, we then expect higher equity premia for situations in which the first-order impact $(-m_{h^*} V_\theta)$ of ambiguity aversion on $-m^* V_{\theta,h^*}$ is larger. This intuition is confirmed by Panels D and F of Figure 4, where larger values of $-m_{h^*}$ are associated with higher equity premium functions. But comparing Panels F and H, we see that the first-order impact on the premium is identical, since $-m_{h^*}$ is the same in both settings. However, from Lemma 3 and Proposition 3, we expect the second-order impact of ambiguity on the premium to be larger for the ambiguity structure in Panel G, a conjecture we confirm by comparing the equity premia plotted in Panels F and H.

4.1.3. Is There an Ambiguity Premium for Imprecise Signals?

In our Lucas economy, the risk premium (A) is quantitatively negligible, even for a large risk aversion. For such a large risk aversion, more precise signals tend to increase, rather than decrease, the equity risk premium. Moreover, when signals are imprecise, the risk premium is bounded from above (Veronesi (2000), Proposition 3). In other words, there is no quantitatively relevant risk premium for imprecise signals.

However, under ambiguity aversion, we show that the key equity premium component is the ambiguity premium (C). For low risk aversions and when signals are imprecise, the resulting equity premium is quantitatively very significant. Therefore, a key question arises: Is there an ambiguity premium for imprecise signals? Corollary 6 presents sufficient conditions for a positive answer to this question.

Corollary 6 *Let Assumption 3 be satisfied and suppose that function $\sqrt{\eta(\theta)}$ is a convex function of θ. Then, for any mean-preserving spread $\widetilde{\Pi}$ of Π it follows:*

$$-\widetilde{m}_{h^*}(1 + k\widetilde{V}_{\theta,h^*}) > -m_{h^*}(1 + kV_{\theta,h^*}) \quad , \tag{65}$$

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

Corollary 6 states that under a convex ambiguity function $\sqrt{\eta(\theta)}$ there is always an ambiguity premium (B)+(C) for information noisiness.[5] This finding implies that when public signal realizations are less precise, the expected excess return is higher, because there is an ambiguity premium for misspecification in the dividend and posterior probabilities Π dynamics

[5]More generally, in the case in which, e.g., $\sqrt{\eta(\theta)}$ is a concave function, the final result depends on the strength of the effects implied by the ambiguity premium (C) components $-m_{h^*}$ and V_{θ,h^*}. In all our numerical examples, we find that the effect caused by changes in V_{θ,h^*} dominates. This evidence supports the hypothesis that, for practical purposes, ambiguity premia for information noisiness arise also more generally than under the conditions of Corollary 6.

under the reference model. When the realized signal precision is low, the posterior probabilities Π are more diffuse, implying larger market prices of ambiguity λ_D^A and λ_e^A in equations (54) and (55). The quantity V_{θ,h^*} in equation (65) also increases under a less precise signal (see again Lemma 2). V_{θ,h^*} is the covariance between equity returns R and signals e. Indeed, the relevant signal dynamics are

$$de = m_{\theta,h^*}dt + \sigma_e d\widetilde{B}_e^{h^*} = m_{\theta,h^*}dt + \sigma_e(d\widetilde{B}_e - m_{\theta,h^*}dt) \quad , \tag{66}$$

implying, from (44)

$$V_{\theta,h^*} \quad = \quad Cov_t(dR, de) \quad . \tag{67}$$

Therefore, the increased ambiguity premium under less precise signals follows from (i) higher market prices of risk and ambiguity, and (ii) higher equilibrium covariances between equity returns and public signals. The higher covariance (67) under imprecise signals arises because less precise a posteriori dividend drift predictions m_{θ,h^*} imply a lower sensitivity of investors' hedging demand to signals. Therefore, positive (negative) signals tend to generate a positive (negative) excess demand for equity and a positive covariance between equilibrium equity returns and signals. This discussion emphasizes the important role of imprecise signals in determining the level of the ambiguity premium.

4.2. Equity Volatility

From Proposition 2, the volatility of stock returns is given by

$$\sigma_R^2 = \sigma_D^2 + V_{\theta,h^*}\left(2 + kV_{\theta,h^*}\right) \quad . \tag{68}$$

Lemma 2 implies that σ_R^2 is a U-shaped function of risk aversion γ, having a minimum $\sigma_R^2 = \sigma_D$ at $\gamma = 1$. Moreover, under a setting of pure ambiguity aversion (i.e., for a degenerate Π) we also have $V_{\theta,h^*} = 0$, and hence $\sigma_R = \sigma_D$. σ_D is the return volatility in a setting of ambiguity aversion without learning. Maenhout (2004) and Trojani and Vanini (2002) obtain this result under a constant opportunity set dynamics.

Therefore, to obtain nontrivial sizes of equity returns volatility, it is important to introduce learning in the model, in excess of ambiguity, and to ensure that $\gamma \neq 1$. However, from Section 4.1, we see that the only such parameter choice than can be consistent with sizable equity premia is $\gamma < 1$, that is, Assumption 3. Therefore, adding learning to a setting of ambiguity aversion is crucial to obtain qualitative model predictions that can be consistent with the excess volatility puzzle. However, only under Assumption 3 can such predictions also be consistent with the equity premium puzzle.

Column Vol of Table 2 illustrates quantitatively the contribution of learning and ambiguity to equity return volatilities, for a setting of no learning (column NL) and a setting without learning (column L), under homogenous degrees of ambiguity $\eta = 0, 0.001, 0.005, 0.01$. In the absence of learning $\sigma_R = \sigma_D$, irrespective of the degree of ambiguity in the economy, implying tiny equity volatilities of approximately 3.75%. In the presence of learning, volatility ranges from about 28.7%, without ambiguity ($\eta = 0$), and about 24.3%, for an ambiguity size $\eta = 0.01$. As stated in Proposition 3, (iii), for homogenous ambiguity structures $\eta(\theta) = \eta$ equity volatility is always lower under ambiguity aversion. The same conclusion does not hold, in general, for heterogenous ambiguity structures.

Given a nondegenerate learning setting with posterior probabilities Π, we analyze the additional contribution of ambiguity aversion to equity returns volatilities when Assumption 3 is satisfied. To study the behavior of σ_R as a function of a general, possibly heterogenous, parameter vector $\eta(\theta)$, we can again make use of Proposition 3, where we characterize the dependence of V_{θ,h^*} on $\eta(\theta)$, in a neighborhood of the purely Bayesian Lucas economy. From that proposition, we expect higher equilibrium volatilities for asymmetric ambiguity structures $\eta(\theta)$ that associate a higher concern for ambiguity with less favorable states of the economy.

Figure 5 highlights the effect of homogeneous and heterogeneous ambiguity structures on equilibrium equity return volatility.

Insert Figure 5 about here

In Panels A, C, and E of Figure 5, we plot three different heterogeneous ambiguity structures in a setting with five possible reference model drift states. In Panels B, D, and F, we present the equity return volatility σ_R as a function of risk aversion γ. For each such panel, we plot volatility in a setting of pure learning (dotted lines), in a setting of pure ambiguity (dash dotted lines), in a setting of heterogeneous ambiguity given, respectively, by Panels A, C, and E on the left for Panels B, D, and F on the right (solid lines), and in a setting of homogeneous ambiguity (dashed lines). As expected, in the presence of learning, all volatility functions are U–shaped and attain a minimum at $\gamma = 1$, where $\sigma_R = \sigma_D$. The volatility for the pure ambiguity setting is constant at σ_D and is very small. Large differences compared to all other volatility curves arise outside small neighborhoods of the point $\gamma = 1$. For instance, in Panel F the value of σ_D is about 3%, but the volatility for $\gamma = 0.6$ in the setting with learning and homogeneous ambiguity is about 18%. This evidence emphasizes further the dominant role of learning, as opposed to ambiguity aversion, in generating interesting model volatility predictions for the excess volatility puzzle.

Different ambiguity structures $\eta(\theta)$ can imply higher or lower volatilities than in the pure Bayesian setting. Consistently with Proposition 3, the ambiguity structure in Panel E delivers the highest volatility curve, and the structure in Panel C implies the lowest volatilities. Asymmetric ambiguity structures of the type in Panel E tend to link less favorable economic states with a higher ambiguity. Homogeneous ambiguity structures always deliver lower volatilities that do those of a pure Bayesian learning setting. The quantitative differences between equilibrium volatilities under homogeneous and heterogeneous ambiguity structures can be substantial, especially for moderate risk aversion. For instance, in Panel F, equity volatility under heterogeneous ambiguity is 66% for $\gamma = 0.1$ and 38% for $\gamma = 0.3$. For the same risk aversion parameters, equity volatility under homogeneous ambiguity is about 57% and 34%, respectively. In relative terms, equity volatility increases by 13% ($\gamma = 0.1$) and 10% ($\gamma = 0.3$), when we introduce ambiguity in the model. These differences arise already for a very small average ambiguity size $\sum \pi_i \eta(\theta_i) = 0.0012$ and for the given set of possible reference model states Θ assumed in Figure 5. For larger average ambiguity, volatility differences become more significant. For instance, with an average ambiguity $\sum \pi_i \eta(\theta_i) = 0.005$, volatility increases by more than 22% ($\gamma = 0.1$) and 17% ($\gamma = 0.3$) in the presence of ambiguity aversion. Similarly, in a model with only three possible reference model drifts, we found that ambiguity aversion can easily increase equity volatility by about 20% for risk aversion parameters up to 0.3.

4.3. Time-Varying Theoretical Risk/Return Relations

From the results in the previous sections we expect learning and ambiguity aversion to have important implications for the prevailing risk/return relations. Column EP/Vol of Table 2 illustrates quantitatively the issue, for a setting of no learning (column NL) and a setting without learning (column L), under homogenous degrees of ambiguity $\eta = 0, 0.001, 0.005, 0.01$. In particular, in a model of pure learning (column L, $\eta = 0$) the low equity premium and the large equity volatility imply a tiny risk/return trade-off of about 0.71%. For settings including ambiguity aversion, the trade-off is between about 17.9% for $\eta = 0.01$ and 55% for $\eta = 0.01$.

From Proposition 2 and equation (68), the relations between risk or equity premia and the conditional variance of returns are given by

$$\mu_R^{wc} = \gamma\sigma_R^2 - \gamma V_{\theta,h^*}\left(1 + kV_{\theta,h^*}\right) \tag{69}$$

and

$$\mu_R = \left(\gamma - \frac{m_{h^*(\theta)}}{V_{\theta,h^*}}\right)\sigma_R^2 - \gamma V_{\theta,h^*}\left(1 + kV_{\theta,h^*}\right) + m_{h^*(\theta)}\left(1 + \frac{\sigma_D}{V_{\theta,h^*}}\right) \quad , \tag{70}$$

respectively. In particular, when Π is non-degenerate equation (70) implies a truly positive, but time-varying, theoretical relation between the equity premium μ_R and the conditional variance σ_R^2. Such a time-varying relation is due to the ambiguity premium component (C) and derives 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 (69) and (70) are biased by a heteroskedastic error term that has a nonzero conditional mean. More precisely, since the dominating term in such errors is

$$-\gamma V_{\theta,h^*}\left(1 + kV_{\theta,h^*}\right) < 0, \tag{71}$$

both relations are biased downwards.

Figure 6 illustrates the theoretical relation between risk or equity premia and equity return conditional variances.

Insert Figure 6 about here

The theoretical (time-varying) equity premium "sensitivity" to changes in σ_R^2 is huge, compared to the sensitivity of the risk premium, which in turn is given by the risk-aversion coefficient γ. Moreover, ambiguity premia derive by definition from model misspecification, rather than from covariances between asset returns and economic state variables. Therefore, we can expect them to be very difficult to identify by, for instance, regression methods. Figure 7 highlights this point by plotting the time series of estimated parameters in a sequence of rolling regressions of R on σ_R^2.

Insert Figure 7 about here

As expected, highly time-varying regression estimates arise. Such estimates may even indicate a switching sign in the estimated relation between μ_R and σ_R^2 over different time periods. More importantly, the estimated (time-varying) coefficients do not even approximately identify correctly the equity premium "sensitivity" to changing variances under learning and ambiguity aversion. For instance, the estimated parameters for $\gamma = 0.9$ in Figure 7 are never above 0.3, but the theoretical "sensitivities" of equity premia to σ_R^2 in Figure 6 are above eight for all ambiguity aversion parameters.

4.4. Biases in EIS estimates

Our model uses time additive power utility functions to obtain simple closed form solutions for the desired equilibria. Such a choice imposes a specific relation between standard risk aversion and EIS. Risk aversions less than one have to be associated with EIS above one. However, this relation does not imply necessarily large estimated EIS. Indeed, one by-product of learning in our context is to induce a stochastic volatility in the a-posteriori expected dividend growth in the model. Similar to the effects noted by Bansal and Yaron (2004) in a full-information asset pricing setting, stochastic volatility of expected dividend growth can induce a large downward bias in a least-squares regression of consumption growth on asset returns when using Euler equations including equity returns. Such regressions are typically used to estimate the EIS in applied empirical work.[6]

To understand the main reason for a negative bias in the estimation of the EIS, we consider for brevity a pure setting of learning with no ambiguity aversion, that is $\eta(\theta) = 0$. From Proposition 1 and 2 we have

$$r = \delta + \gamma m_\theta - \frac{1}{2}\gamma(\gamma+1)\sigma_D^2 \quad , \quad \mu_R = \gamma\left(\sigma_D^2 + V_\theta\right),$$

where $m_\theta = E\left(dD/D|\,\mathcal{F}_t\right)$. Therefore,

$$E\left(dP/P + D/P|\,\mathcal{F}_t\right) = r + \mu_R = \delta + \gamma E\left(dD/D|\,\mathcal{F}_t\right) - \frac{1}{2}\gamma(\gamma+1)\sigma_D^2 + \gamma\left(\sigma_D^2 + V_\theta\right),$$

and, solving for $E\left(dD/D|\,\mathcal{F}_t\right)$:

$$E\left(dD/D|\,\mathcal{F}_t\right) = a + b \cdot E\left(dP/P + D/P|\,\mathcal{F}_t\right) - V_\theta, \tag{72}$$

where $a = -\delta/\gamma + \frac{1}{2}(\gamma-1)\sigma_D^2$ and $b = 1/\gamma$. Equation (72) defines a correctly specified theoretical linear regression equation if and only if the random term V_θ is 0. This in turn can happen only if no learning is present (Π is degenerate) or $\gamma = 1$ (log utility). In all other cases, the error term

$$d\varepsilon_t := dD/D - a - b \cdot E\left(dP/P + D/P|\,\mathcal{F}_t\right)$$

will be correlated with the regressor $dP/P + D/P$ in a least-squares regression of dD/D on $dP/P+D/P$. Under Assumption 3, such correlation induces a downward bias in the estimation of the EIS $1/\gamma$ in a least-squares regression of aggregate consumption growth dD/D on total equity returns $dP/P + D/P$. Since V_θ is decreasing in relative risk aversion, we can expect the bias to be larger for lower γ values. Figures 8 and 9 illustrate these features.

Insert Figures 8 and 9 about here

[6]See, e.g., Hall (1988), Vissing-Jorgensen (2002), and Attanasio and Vissing-Jorgensen (2003), among others.

In Figures 8 and 9, we observe a very large bias in the mean least-squares estimates of the EIS $1/\gamma$ in a regression of dD/D on $dP/P+D/P$. As expected, the bias is larger for lower values of γ. For instance, for $\gamma = 0.5$ the mean estimate of $1/\gamma$ is between 0.2 and 0.4, depending on the amount of ambiguity in the economy. This corresponds to a downward bias in the estimation of the EIS of about 80%. For $\gamma = 0.7$ mean EIS estimates range between about 0.35 and 0.6. Interestingly, such estimated values of the EIS are compatible with those obtained, e.g., in Vissing-Jorgensen (2002, Table 2A) and Attanasio and Vissing-Jorgensen (2003, Table 1A) for Euler equations including stock returns.

5. Conclusion

We derive asset prices in a continuous-time partial information Lucas economy with ambiguity aversion and time-additive power utility. In our model, ambiguity aversion implies only a partial asymptotic learning about a neighborhood of a priori statistically indistinguishable beliefs.

For low risk aversion, the joint presence of learning and ambiguity enforces large equity premia, already under a moderate amount of ambiguity in the economy. Additional model predictions are consistent also with the interest rate and the excess volatility puzzles. Introducing both learning and ambiguity aversion is crucial. Model settings in which learning is absent need unrealistically large amounts of pessimism to generate sizable equity premia. At the same time, they imply tiny equity volatilities. A further model implication is a highly time-varying true relation between excess returns and their conditional variances. This model feature generates estimated relations between excess returns and conditional variances with an undetermined sign and implies huge time-varying biases in the naively estimated risk-return trade-off. Finally, standard estimates of the elasticity of intertemporal substitution (EIS) based on Euler equations for equity returns are strongly downward biased under learning and ambiguity aversion. Therefore, an EIS well above one in the model is consistent with an observed (biased) estimated EIS well below one.

The time additive power utility function in our model allows us to obtain simple and easily interpretable closed form solutions for the desired equilibria, at the cost of constraining the relation between risk aversion and EIS. Disentangling risk aversion and EIS would allow for an additional degree of freedom which could be used, e.g., to generate higher worst case equity premia in our model. Such extensions are therefore interesting venues for further research. The specific relation between standard risk aversion and EIS could be weakened by using a setting of learning under ambiguity aversion with Epstein and Schneider (1989)-type preferences. Hayashi (2005) has provided a theoretical axiomatic framework of ambiguity aversion with recursive utility. To our knowledge, no concrete asset pricing setting with ambiguity and recursive utility has been studied so far in the literature. Adding learning to such settings will even improve the technical difficulties necessary to handle conveniently these models. Moreover, the basic intuition derived from our model is likely to hold also under more general preferences that disentangle risk aversion and the EIS. Investors with high relative risk aversions increase their hedging demand when they expect low consumption growth. This demand counterbalances the negative price pressure deriving from negative dividend news. Under ambiguity aversion, investors tend to understate actual consumption growth. Highly risk averse investors therefore increase further their hedging demand for equity. Since the supply of the risky asset is fixed and the riskless bond is in zero net supply, the higher demand increases the price of the risky asset relative to dividends. At the same time, for low EIS a lower expected consumption growth because of ambiguity aversion induces a large ceteris paribus a large substitution from today to tomorrow consumption, in order to smooth consumption out. Such an excess saving demand increases further the price of equity relative to dividends and lowers the equilibrium interest rate. From a more general perspective, the assumption of a low risk aversion in our model is just a condition ensuring that the elasticity of the total demand for risky assets with respect to changes in expected consumption growth is positive.

Appendix: Proofs

Proof of Corollary 1. The statement of the corollary follows easily by noting that under h_D-distorted dynamics we have:

$$d\widetilde{B}_D^h = dB_D + k_D\left(\theta + h_D\sigma_D - m_{\theta,h}\right)dt \quad , \quad d\widetilde{B}_e^h = dB_e + k_e\left(\theta + h_D\sigma_D - m_{\theta,h}\right)dt \quad .$$

■

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

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

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

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

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

$$dD = \left(\theta_i + h(\theta_i)\sigma_D\right)Ddt + \sigma_D D dB_D \quad . \qquad \text{(A3)}$$

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

$$(P_i) \quad : \quad \begin{cases} V^i(D) = \inf_{h(\theta_i)} E\left(\int_s^\infty e^{-\delta(t-s)}\frac{D(t)^{1-\gamma}}{1-\gamma}dt\right|\left. D(s) = D\right) \\ \frac{1}{2}h(\theta_i)^2 \leq \eta(\theta_i) \end{cases}$$

subject to the dividend dynamics

$$dD = (\theta_i + h(\theta_i)\sigma_D) Ddt + \sigma_D DdB_D \quad . \tag{A4}$$

The Hamilton Jacobi Bellman equation for this problem reads

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

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

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

Slackness then gives

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

implying

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

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

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

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

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

implying, under the given assumptions,

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

where

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

We thus obtain for the price of any risky asset with dividend process $(D(t))_{t\geq 0}$:

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

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:

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

i.e., 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 (24) under the worst-case likelihood $h^*(\theta) = -\sqrt{2\eta(\theta)}$. Statement (ii) follows by expressing the $R-$dynamics obtained in (i) with respect to the filtered Brownian motions $\widetilde{B}_D, \widetilde{B}_e$ under the reference model. ■

Proof of Proposition 3. Let $\theta := (\theta_1, .., \theta_n)'$ and

$$C(x) = \frac{1}{\delta + (\gamma - 1)x - \gamma(1-\gamma)\frac{\sigma_D^2}{2}} \quad .$$

To simplify notations, we define

$$V(\theta) = \sum_i \pi_i \theta_i \left[\frac{C(\theta_i)}{\sum_j \pi_j C(\theta_j)} - 1\right] \quad . \tag{A13}$$

To prove the first and the second statement of the proposition, we compute the gradient $\nabla V := (V_{\theta_1}, .., V_{\theta_n})'$, where for brevity $V_{\theta_k} := \partial V/\partial\theta_k$, $k = 1, .., n$. It then follows that

$$\begin{aligned} V_{\theta_k}(\theta) &= \pi_k \left[\frac{C(\theta_k)}{\sum_j \pi_j C(\theta_j)} - 1 \right] + \sum_i \pi_i \theta_i \left[\frac{C'(\theta_i)\delta_{ki}}{\sum_j \pi_j C(\theta_j)} - C(\theta_i) \frac{\pi_k C'(\theta_k)}{(\sum_j \pi_j C(\theta_j))^2} \right] \\ &= \frac{\pi_k C(\theta_k)}{\sum_j \pi_j C(\theta_j)} - \pi_k + \theta_k \frac{\pi_k C'(\theta_k)}{\sum_j \pi_j C(\theta_j)} - \frac{\sum_i \pi_i \theta_i C(\theta_i)}{\sum_j \pi_j C(\theta_j)} \cdot \frac{\pi_k C'(\theta_k)}{\sum_j \pi_j C(\theta_j)} \\ &= \frac{\pi_k C(\theta_k)}{\sum_j \pi_j C(\theta_j)} - \pi_k + \frac{\pi_k C'(\theta_k)}{\sum_j \pi_j C(\theta_j)} [\theta_k - m_\theta - V(\theta)] \quad , \end{aligned}$$

where $\delta_{ki} = 1$ if $k = i$ and $\delta_{ki} = 0$ else. From the explicit expression for $C(x)$,

$$C'(\theta_k) = (1 - \gamma)C(\theta_k)^2 \quad ,$$

implying

$$V_{\theta_k}(\theta) = \frac{\pi_k C(\theta_k)}{\sum_j \pi_j C(\theta_j)} - \pi_k + (1 - \gamma) \frac{\pi_k C(\theta_k)^2}{\sum_j \pi_j C(\theta_j)} [\theta_k - m_\theta - V(\theta)] \quad .$$

Under Assumption 3, the conditions

$$\frac{\pi_k C(\theta_k)}{\sum_j \pi_j C(\theta_j)} \gtrless \pi_k \quad ; \quad \theta_k - m_\theta \gtrless V(\theta) \tag{A14}$$

imply $V_{\theta_k} \gtrless 0$. Since

$$\left. \frac{\partial V_{\theta,h^*}}{\partial \sqrt{2\eta(\theta_k)}} \right|_{\eta(\theta)=0} = \left. \frac{\partial V(\theta - \sqrt{2\eta(\theta)}\sigma_D)}{\partial \sqrt{2\eta(\theta_k)}} \right|_{\eta(\theta)=0} = -V_{\theta_k}(\theta)\sigma_D, \tag{A15}$$

where $\sqrt{2\eta(\theta)} := (\sqrt{2\eta(\theta_1)}, .., \sqrt{2\eta(\theta_n)})'$, condition (A14) also implies the sign of (A15). To prove the third statement, we calculate

$$\begin{aligned}
\left.\frac{\partial V_{\theta,h^*}}{\partial\sqrt{2\eta}}\right|_{\eta=0} &= \sum_{k=1}^{n} \left.\frac{\partial V(\theta - \sqrt{2\eta(\theta)}\sigma_D)}{\partial\sqrt{2\eta(\theta_k)}}\right|_{\eta(\theta)=0} \\
&= -\sum_{k=1}^{n}\left\{\frac{\pi_k C(\theta_k)}{\sum_j \pi_j C(\theta_j)} - \pi_k + (1-\gamma)\frac{\pi_k C(\theta_k)^2}{\sum_j \pi_j C(\theta_j)}[\theta_k - m_\theta - V(\theta)]\right\}\sigma_D \\
&= -(1-\gamma)\sum_{k=1}^{n}\frac{\pi_k C(\theta_k)^2}{\sum_j \pi_j C(\theta_j)}[\theta_k - m_\theta - V(\theta)] \quad . \qquad \text{(A16)}
\end{aligned}$$

Therefore, it is sufficient to study the sign of

$$\sum_{k=1}^{n}\frac{\pi_k C(\theta_k)^2}{\sum_{j=1}^{n}\pi_j C(\theta_j)^2}[\theta_k - m_\theta - V(\theta)] \quad = \quad \sum_{k=1}^{n}\frac{\pi_k C(\theta_k)^2}{\sum_{j=1}^{n}\pi_j C(\theta_j)^2}[\theta_k - m_\theta] - V(\theta) \quad .$$

Since $C(x)$ is increasing and convex,

$$\sum_{k=1}^{n}\frac{\pi_k C(\theta_k)^2}{\sum_{j=1}^{n}\pi_j C(\theta_j)^2}[\theta_k - m_\theta] - V(\theta) > 0 \qquad \text{(A17)}$$

and

$$\left.\frac{\partial V_{\theta,h^*}}{\partial\sqrt{2\eta}}\right|_{\eta=0} < 0 \quad , \qquad \text{(A18)}$$

under the given assumptions, concluding the proof. ■

Proof of Lemma 3. We compute second-order asymptotics for the function

$$H(\eta(\theta_1), .., \eta(\theta_n)) := H(\eta(\theta)) := -m_{h^*}V_{\theta,h^*} \quad . \qquad \text{(A19)}$$

We first have:

$$\partial_k H = -V_{\theta,h^*}\partial_k m_{h^*} - m_{h^*}\partial_k V_{\theta,h^*} \quad , \qquad \text{(A20)}$$

and

$$\partial_i \partial_k H = -V_{\theta,h^*} \partial_i \partial_k m_{h^*} - \partial_k m_{h^*} \partial_i V_{\theta,h^*} - \partial_i m_{h^*} \partial_k V_{\theta,h^*} - m_{h^*} \partial_i \partial_k V_{\theta,h^*} \quad , \tag{A21}$$

where subscripts k, i denote partial derivatives for the arguments $\sqrt{2\eta(\theta_k)}$, $\sqrt{2\eta(\theta_i)}$. Using the explicit expression of m_{h^*}, and since $m_{h^*} = 0$ for $\eta(\theta) = 0$, it follows that

$$\begin{aligned}
\partial_k H|_{\eta(\theta)=0} &= -V_\theta \partial_k m_{h^*}|_{\eta(\theta)=0} = V_\theta \pi_k \sigma_D \\
\partial_i \partial_k H|_{\eta(\theta)=0} &= -\partial_k m_{h^*} \partial_i V_{\theta,h^*}|_{\eta(\theta)=0} - \partial_i m_{h^*} \partial_k V_{\theta,h^*}|_{\eta(\theta)=0} \\
&= [\pi_k \partial_i V_{\theta,h^*}|_{\eta(\theta)=0} + \pi_i \partial_k V_{\theta,h^*}|_{\eta(\theta)=0}]\sigma_D \quad .
\end{aligned} \tag{A22}$$

A second-order Taylor expansion of H at $\eta(\theta) = 0$ then gives, up to term or order $o(\|\eta(\theta)\|)$,

$$\begin{aligned}
H(\eta(\theta)) &= \sum_{k=1}^n V_\theta \pi_k \sqrt{2\eta(\theta_k)}\sigma_D \\
&\quad + \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^n \left\{\pi_k \partial_i V_{\theta,h^*}|_{\eta(\theta)=0} + \pi_i \partial_k V_{\theta,h^*}|_{\eta(\theta)=0}\right\} \sqrt{2\eta(\theta_k)}\sqrt{2\eta(\theta_i)}\sigma_D^2 \\
&= -m_{h^*} V_\theta + \sum_{i=1}^n \sum_{k=1}^n \pi_k \partial_i V_{\theta,h^*}|_{\eta(\theta)=0} \sqrt{2\eta(\theta_k)}\sqrt{2\eta(\theta_i)}\sigma_D^2 \\
&= -m_{h^*} V_\theta + \sum_{i=1}^n \partial_i V_{\theta,h^*}|_{\eta(\theta)=0} \sqrt{2\eta(\theta_i)}\sigma_D \sum_{k=1}^n \pi_k \sqrt{\eta(\theta_k)}\sigma_D \\
&= -m_{h^*}[V_\theta + \sum_{i=1}^n \partial_i V_{\theta,h^*}|_{\eta(\theta)=0} \sqrt{2\eta(\theta_i)}\sigma_D]
\end{aligned} \tag{A23}$$

■

Proof of Corollary 6. From Lemma 2, V_{θ,h^*} is increasing in mean-preserving spreads $\widetilde{\Pi}$, under the given assumptions. Moreover,

$$-m_{h^*} = \sum_{i=1}^n \pi_i \sqrt{\eta(\theta_i)}\sigma_D \tag{A24}$$

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

References

Aït-Sahalia, Y., J. A. Parker, and M. Yogo, 2004, "Luxury Goods and the Equity Premium," *Journal of Finance*, 52, 2959–3004.

Anderson, E. W., L. P. Hansen, and T. J. Sargent, 2003, "Robustness, Detection and the Price of Risk," Working paper, University of Chicago.

Attanasio, P., and G. Weber, 1989, "Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption," *Economic Journal*, 99, 59–73.

Attanasio, P. O., J. Banks, and S. Tanner, 2002, "Asset Holding and Consumption Volatility," *Journal of Political Economy*, 110, 771–792.

Attanasio, P. O., and A. Vissing-Jorgensen, 2003, "Stock Market Participation, Intertemporal Substitution, and Risk Aversion," working paper, AEA Papers and Proceedings.

Bansal, R., and A. Yaron, 2004, "Risks for the Long Run: a Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Barberis, N., 2000, "Investing for the Long Run when Returns are Predictable," *Journal of Finance*, 55, 225–264.

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, 249–283.

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., 1999, "Asset Prices, Consumption and the Business Cycl," in *Handbook of Macroeconomics, Volume I*, ed. by J. B. Taylor, and M. Woodford. Elsevier Science, North Holland.

Chen, Z., and L. G. Epstein, 2002, "Ambiguity, Risk and Asset Returns in Continuous Time," *Econometrica*, 70, 1403–1443.

Detemple, J., 1986, "Asset Pricing in a Production Economy with Incomplete Information," *Journal of Finance*, 41, 383–391.

Dothan, M., and D. Feldman, 1986, "Equilibrium Interest Rates and Multiperiod Bonds in a Partially Observable Economy," *Journal of Finance*, 41, 369–382.

Ellsberg, D., 1961, "Risk, Ambiguity and the Savage Axioms," *Quarterly Journal of Economics*, 25, 643–669.

Epstein, L. G., and J. Miao, 2003, "A two-Person Dynamic Equilibrium under Ambiguity," *Journal of Economic Dynamics and Control*, 27, 1253–1288.

Epstein, L. G., and M. Schneider, 1989, "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: a Theoretical Framework," *Econometrica*, 57, 937–969.

——— , 2002, "Learning under Ambiguity," Working paper, University of Rochester.

——— , 2003, "Recursive Multiple Priors," *Journal of Economic Theory*, 113, 32–50.

——— , 2004, "Ambiguity, Information Quality and Asset Pricing," Working paper, University of Rochester.

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

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

Gagliardini, P., P. Porchia, and F. Trojani, 2004, "Ambiguity Aversion, Bond Pricing and the Non-Robustness of some Affine Term Structures," Working paper, University of St.Gallen.

Gagliardini, P., F. Trojani, and G. Urga, 2005, "Robust GMM Tests for Structural Breaks," *Journal of Econometrics*, forthcoming.

Gennotte, G., 1986, "Optimal Portfolio Choice Under Incomplete Information," *Journal of Finance*, 41, 733–746.

Gilboa, I., and D. Schmeidler, 1989, "Maxmin Expected Utility with Non-unique Prior," *Journal of Mathematical Economics*, 18, 141–153.

Hall, R. E., 1988, "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96, 339–357.

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

Hansen, L. P., and K. J. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

Hayashi, 2005, "Intertemporal substitution, risk aversion and ambiguity aversion," *Economic Theory*, 25(4), 933–956.

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

Knox, T., 2005, "Foundations for Learning How to Invest when Returns are Uncertain," *Journal of Economic Theory*, forthcoming.

Liptser, R., and A. N. Shiryaev, 2001, *Statistics of Random Processes, Vol. I: General Theory.* Spinger, 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), 131–164.

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), 951–983.

Mancini, L., E. Ronchetti, and F. Trojani, 2005, "Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models," *Journal of the American Statistical Association*, 100(470), 628–641.

Mehra, R., and E. C. Prescott, 1985, "The Equity Premium: a Puzzle," *Journal of Monetary Economics*, 15, 145–161.

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

Ortelli, C., and F. Trojani, 2005, "Robust Efficient Method of Moments," *Journal of Econometrics*, 128(1), 69–97.

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, 335–381.

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.

Trojani, F., and P. Vanini, 2002, "A Note on Robustness in Merton's Model of Intertemporal Consumption and Portfolio Choice," *Journal of Economic Dynamics and Control*, 26, 423–435.

——— , 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, 2465–2486.

Veronesi, P., 1999, "Stock market overreaction to bad news in good times: A rational expectations equilibrium model," *Review of Financial Studies*, 12, 975–1007.

——— , 2000, "How does information quality affect stock returns?," *Journal of Finance*, 55, 807–837.

Vissing-Jorgensen, A., 2002, "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110, 825–853.

Wakker, P., and D. Deneffe, 1996, "Eliciting von Neumann-Morgenstern Utilities when Probabilities are Distorted or Unknown," *Management Science*, 42, 1131–1150.

Xia, Y., 2001, "Learning About Predictability: The Effects of Parameter Uncertainty on Dynamic Asset Allocation," *Journal of Finance*, 56, 205–246.

Figures and Tables

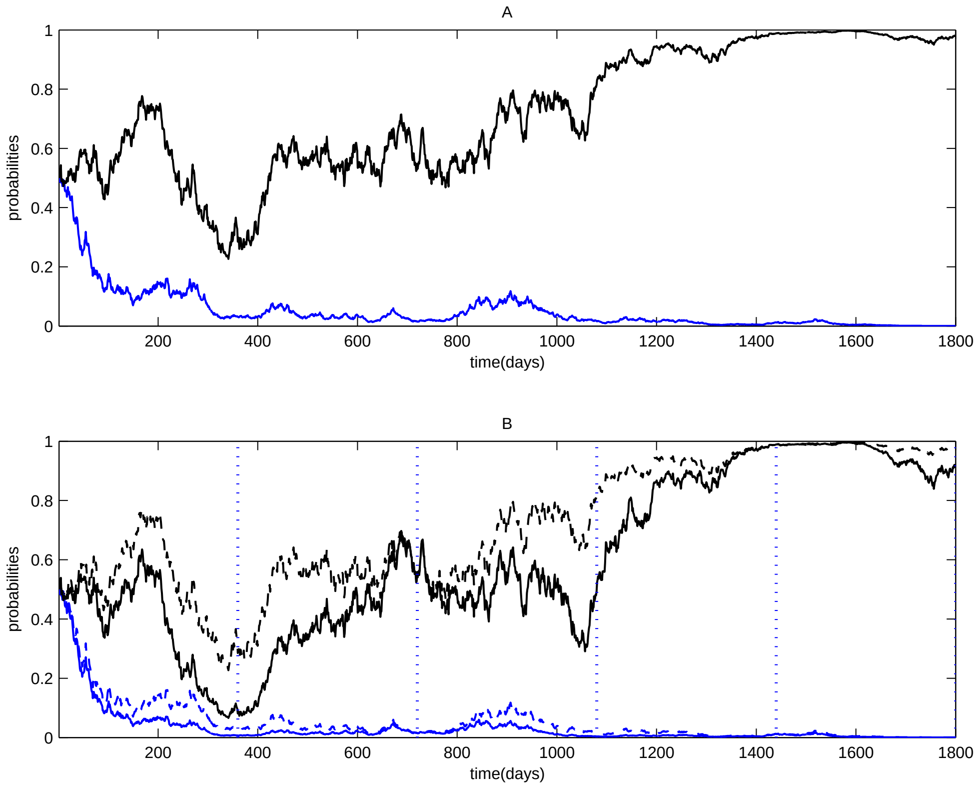

Fig. 1. Posterior probabilities dynamics. The panels display trajectories for the probability π_1 given in equation (17) of Example 1. Panel A shows two trajectories for π_1 with $a = 0$ in equation (19). We plot the same trajectories with the same random seed in Panel B (dashed lines) and add the trajectories (solid lines) where we assume $a = 0.015$. The switching in equation (19) is deterministic and occurs every year (see the dotted vertical lines in Panel B). Further parameters are $\Theta = \{0.0075, 0.0275\}$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$.

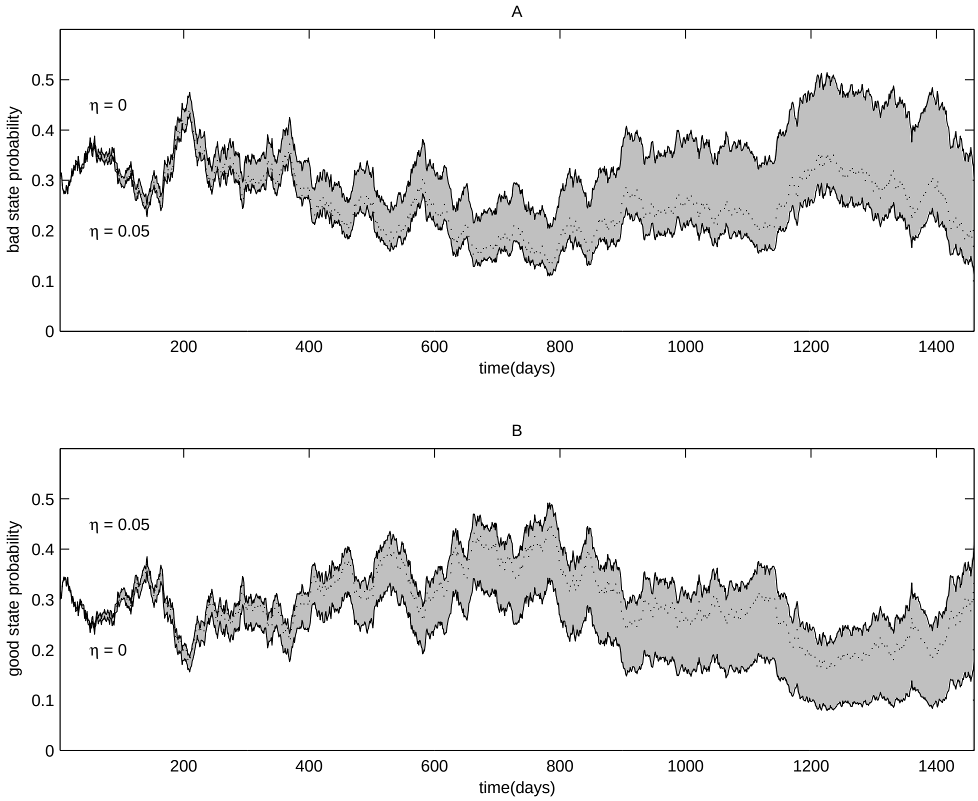

Fig. 2. The effect of ambiguity aversion on the prevailing posterior probabilities dynamics. We assume three possible states and the filtered probabilities dynamics in equation (39) with parameters set equal to $\sigma_D = 0.0375$, $\sigma_e = 0.015$, $\Theta = \{0.0023, 0.0173, 0.0323\}$, $\theta = 0.0173$, and a set of discretized normal priors $\Pi(0) = \{0.3085, 0.3829, 0.3085\}$. Panel A plots the probability dynamics of the "bad" state θ_1 for three different levels of a homogeneous ambiguity parameter $\eta = \{0, 0.025, 0.05\}$. The dotted line represents the dynamics under the intermediate level of ambiguity $\eta = 0.025$. In Panel B, we plot the dynamics of the posterior probabilities for the "good" state θ_3 for the same levels of ambiguity (these graphs are based on the same random seed as the one used in Panel A).

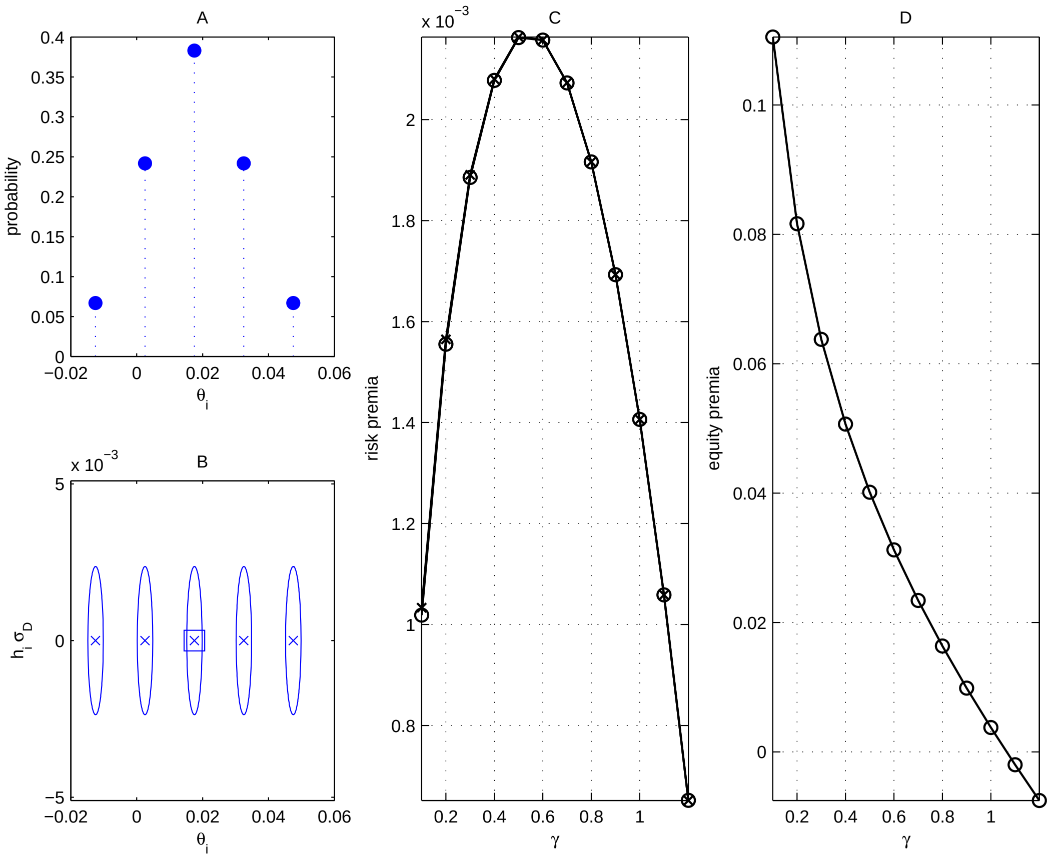

Fig. 3. Risk premium and ambiguity premium under homogeneous ambiguity. Panel A plots the set of probabilities Π relevant for the figure. Panel B plots the different relevant reference model states $\theta_1, .., \theta_n$. The true reference model dividend drift state is marked with a square and is equal to the posterior expected value $\sum \pi_i \theta_i$. We use a small amount of homogeneous ambiguity $\eta = 0.0012$. The size of the ambiguous neighborhoods $\Xi(\theta_1), .., \Xi(\theta_n)$ is highlighted by the ellipses centered at $\theta_1, .., \theta_n$ in Panel B. Further, we set $\delta = 0.05$, $\sigma_D = 0.0375$ and $\sigma_e = 0.015$. With these parameters, we plot the resulting risk premium μ_R^{wc} and the equity premium μ_R in Panels C and D as functions of γ.

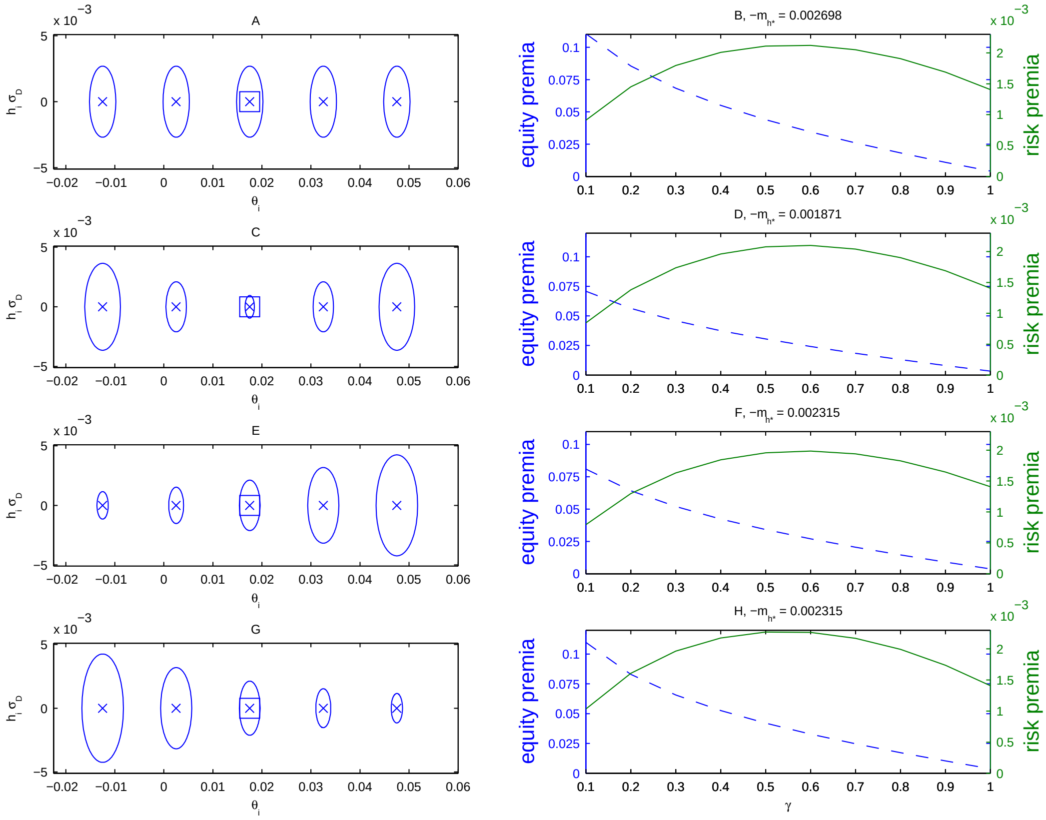

Fig. 4. Risk premium and ambiguity premium under heterogeneous ambiguity. Panels A, C, E, and G plot different entropy preserving distributions of ambiguity around the reference model dividend drift states $\theta_1, .., \theta_5$, i.e., Panels A, C, E, and G are such that the weighted entropy measure $\sum \pi_i \eta(\theta_i)$ is equal to 0.0012, as in Figure 3. Panels B, D, F, and H plot the equity premium μ_R (dashed lines) and the risk premium μ_R^{wc} (solid line) implied by the different distributions of ambiguity in Panels A, C, E, and G as a function of risk aversion $\gamma \in [0.1, 1]$. For comparability, we also give the size of $-m^*$ implied by each plot. Further, we set $\delta = 0.05$, $\sigma_D = 0.0375$ and $\sigma_e = 0.015$.

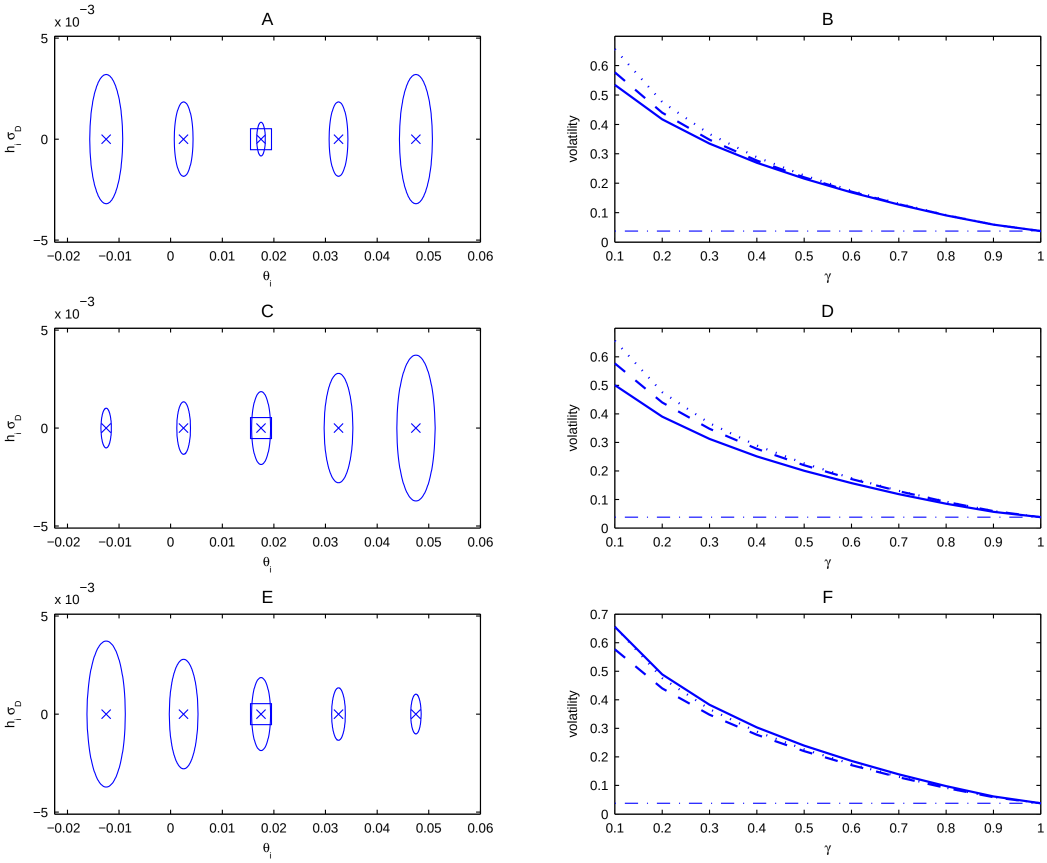

Fig. 5. Equity volatility. Panels A, C, and E plot different entropy preserving distributions of ambiguity around the reference model dividend drift states $\theta_1, .., \theta_5$, i.e., Panels A, C, and E are such that the weighted entropy measure $\sum \pi_i \eta(\theta_i)$ is equal to 0.0012, as in Figure 3. Panels B, D, and F plot the resulting volatility σ_R (solid curves) implied by the different distributions of ambiguity in Panels A, C, and E. For comparability, we plot in each graph the quantities prevailing under a homogeneous ambiguity parameter $\eta = 0.01$ (dashed curves), the pure learning setting, i.e, $\eta(\theta_i) = 0$, $i = 1, .., n$ (dotted curve), and the pure ambiguity case arising under a degenerate Π (dash dotted line). Further, we set $\delta = 0.05$, $\sigma_D = 0.0375$ and $\sigma_e = 0.015$.

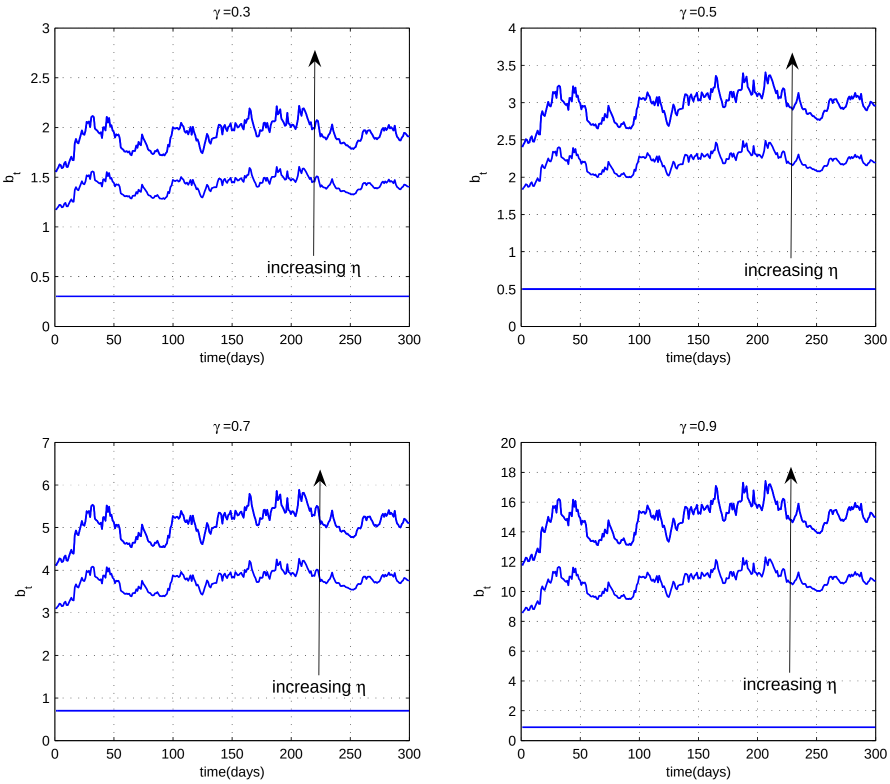

Fig. 6. Theoretical time-varying return/volatility trade-off. For different parameters $\gamma = 0.3, 0.5, 0.7, 0.9$, we plot the theoretical (time-varying) coefficient $b_t = \gamma - m_{h^*}(\theta)/V_{h^*,\theta}$ in the theoretical expected excess return and variance relation $\mu_R = b_t\sigma_R^2 + c_t$, where $c_t = -\gamma V_{\theta,h^*}(1 + kV_{\theta,h^*}) + m_{h^*}(1 + \sigma_D^2/V_{h^*,\theta})$. In all panels, we plot b_t as a function of time for homogeneous ambiguity structures $\eta(\theta) = \eta$, where $\eta = 0.0017, 0.0033$. The horizontal flat solid lines correspond to the (constant) risk premium coefficient γ.

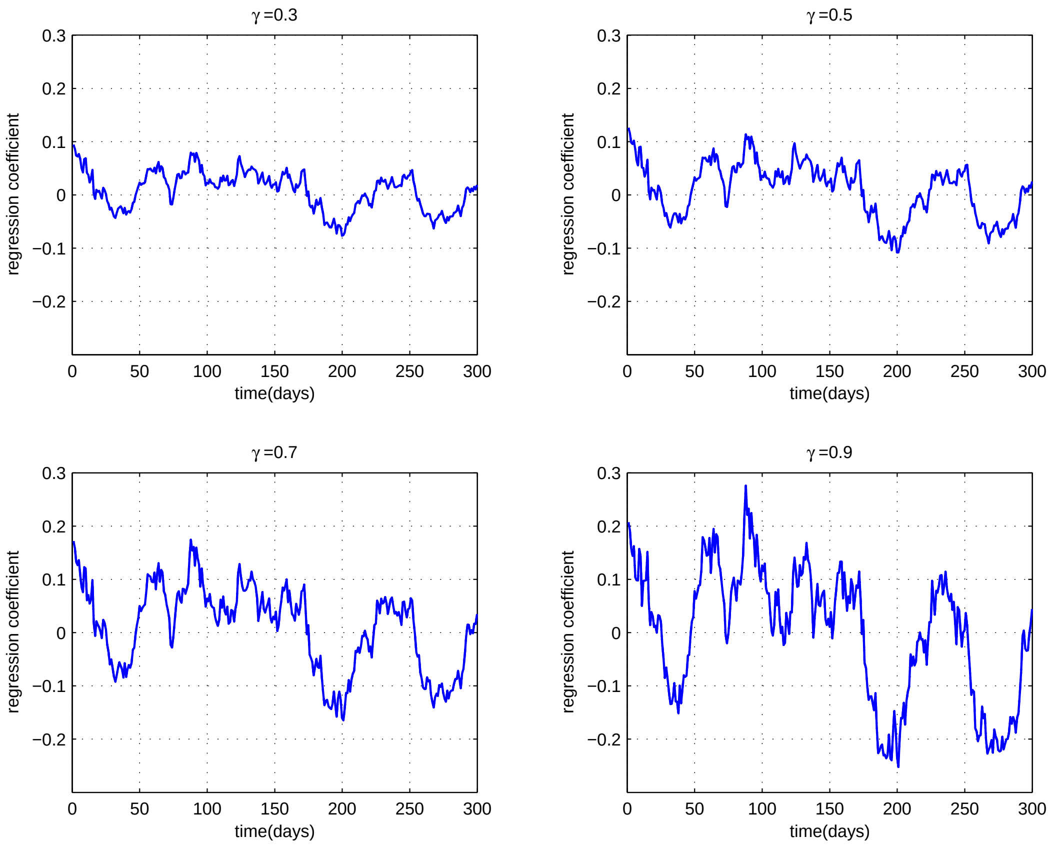

Fig. 7. Rolling regression analysis. For different parameters $\gamma = 0.3, 0.5, 0.7, 0.9$ we plot the time variation of the estimated parameter b in a rolling regression of R on σ_R^2. The rolling regressions are based on sample sizes of 50 observations simulated from a model with three reference model drift states $\Theta = \{0.0025, 0.0175, 0.0325\}$ and under a homogeneous degree of ambiguity $\eta = 0.001$. The true dividend is $\theta = 0.0175$. Further parameters are $\delta = 0.05$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$.

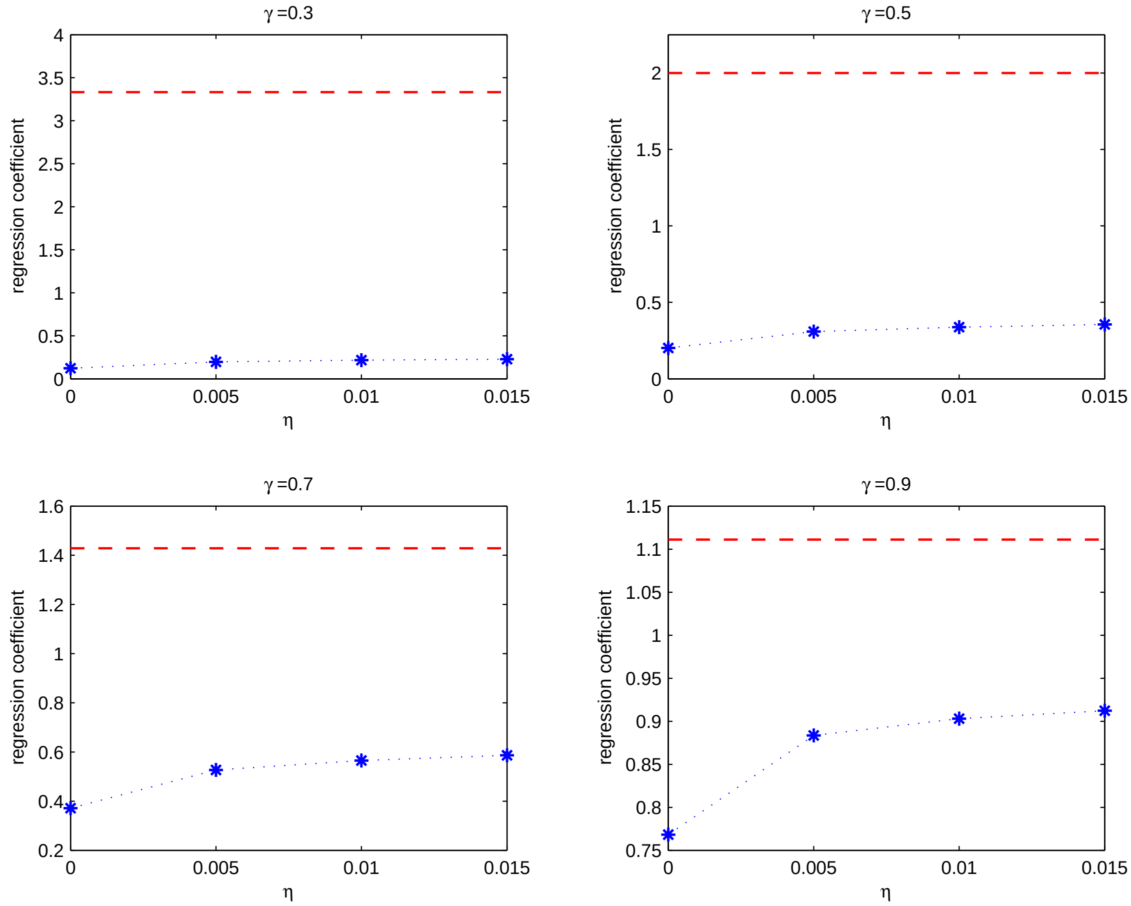

Fig. 8. Regression analysis. For different parameters $\gamma = 0.3, 0.5, 0.7, 0.9$, we plot the mean estimated parameter in 1000 regressions of dD/D on $dP/P + P/D$. The regressions are based on sample sizes of 365 observations simulated from a model with three reference model drift states $\Theta = \{0.0025, 0.0175, 0.0325\}$ and under three homogeneous degrees of ambiguity $\eta = 0, 0.005, 0.01, 0.015$. The true dividend is $\theta = 0.0175$. Further parameters are $\delta = 0.05$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$. In all panels, the dashed horizontal lines give the correct underlying value $1/\gamma$ of the EIS. The dotted lines give the resulting mean parameter estimates as a function of η.

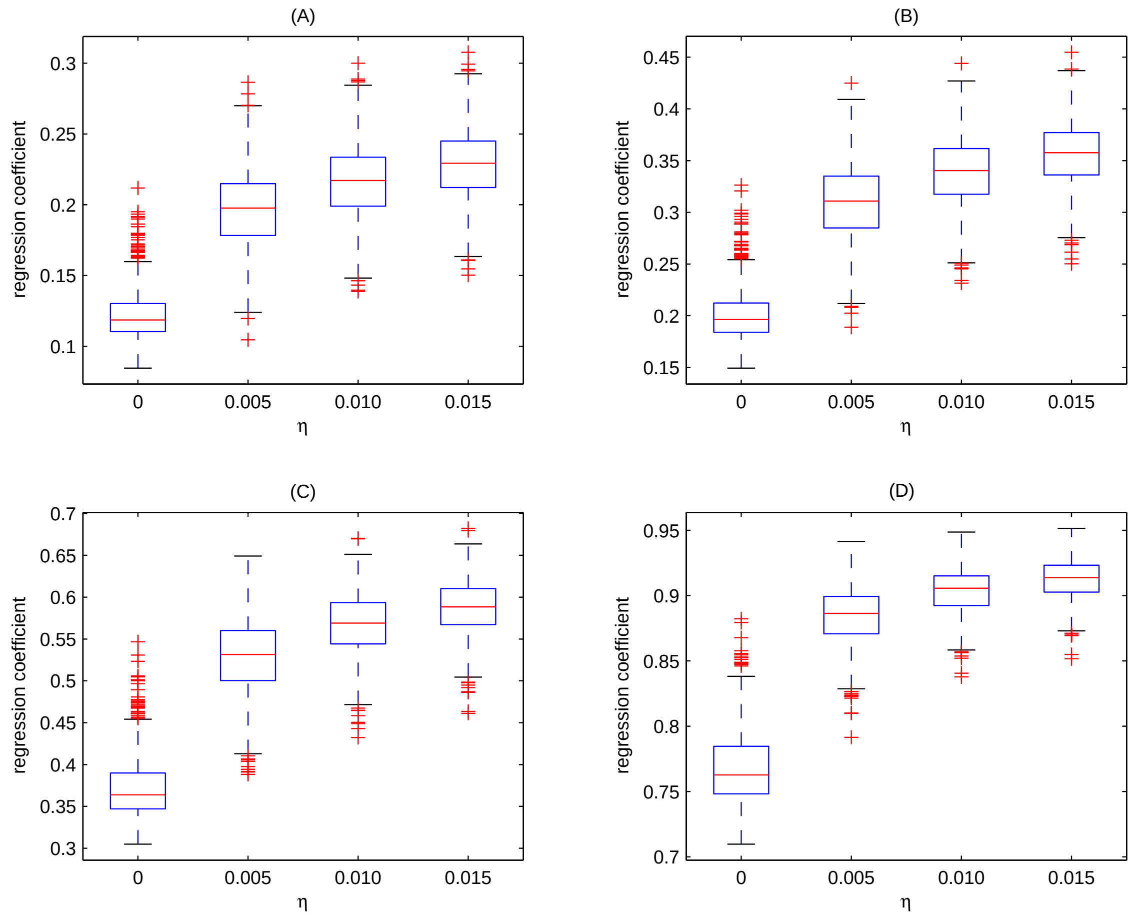

Fig. 9. Regression analysis. For different parameters $\gamma = 0.3, 0.5, 0.7, 0.9$, we present the box plots of the estimated parameters in 1000 regressions of dD/D on $dP/P + P/D$. The regressions are based on sample sizes of 365 observations simulated from a model with three reference model drift states $\Theta = \{0.0025, 0.0175, 0.0325\}$ and under three homogeneous degrees of ambiguity $\eta = 0, 0.005, 0.01, 0.015$. The true dividend is $\theta = 0.0175$. Further parameters are $\delta = 0.05$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$.

	interest rate (%)		P/D ratio	
η	NL	L	NL	L
0.000	2.8223	2.8223	87.5214	96.3729
0.001	2.7384	2.7384	81.5374	83.7536
0.005	2.6348	2.6348	75.1836	79.8640
0.010	2.5571	2.5571	71.0357	74.7932
0.020	2.4473	2.4473	65.8945	68.7587

Table 1

The table presents values of interest rates and price-dividend ratios for a risk aversion $\gamma = 0.5$ and different homogeneous levels of ambiguity η, in model settings without learning (NL) and with learning (L). Calculations are based on a model with five reference model drift states $\Theta = \{-0.0125, 0.0025, 0.0175, 0.0325, 0.0475\}$. The true dividend is $\theta = 0.0175$. Further parameters are $\delta = 0.02$, $\sigma_D = 0.0375$, $\sigma_e = 0.005$.

η	RP (%)		EP (%)		Vol (%)		EP/Vol (%)	
	NL	L	NL	L	NL	L	NL	L
0.000	0.0703	0.2033	0.0703	0.2033	3.7500	28.7165	1.8750	0.7079
0.001	0.0703	0.1952	0.2380	4.8484	3.7500	27.0534	6.3471	17.9216
0.005	0.0703	0.1871	0.4453	9.9459	3.7500	25.4045	11.8750	39.1502
0.010	0.0703	0.1819	0.6006	13.3940	3.7500	24.3429	16.0171	55.0222

Table 2

The table presents values of risk premia (RP), equity premia (EP), volatilities (Vol), and ratios of equity premia and volatilities (EP/Vol) for a risk aversion $\gamma = 0.5$ and different homogeneous levels of ambiguity η, under model settings without learning (NL) and with learning (L). Calculations are based on a model with five reference model drift states $\Theta = \{-0.0225, -0.0025, 0.0175, 0.0375, 0.0575\}$. The true dividend is $\theta = 0.0175$. Further parameters are $\delta = 0.03$, $\sigma_D = 0.0375$, $\sigma_e = 0.01$.