Working Paper
Series

National Centre of Competence in Research
Financial Valuation and Risk Management

Working Paper No. 249

Learning and Asset Prices under Ambiguous Information

Markus Leippold Fabio Trojani

Paolo Vanini

First version: September 2003
Current version: October 2004

This research has been carried out within the NCCR FINRISK project on "Conceptual Issues in Financial Risk Management"

Learning and Asset Prices under Ambiguous Information*

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

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

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

[c]*Zurich Cantonal Bank, Switzerland*

(First version: September 2003; This version October 12, 2004)

---

*We are grateful to Andrew Abel for many very valuable comments and suggestions on an earlier draft. We also thank Abraham Lioui, Pascal Maenhout, Alessandro Sbuelz and participants to the 2004 European Summer Symposium in Financial Markets in Gerzensee. The authors gratefully acknowledge the financial support of the Swiss National Science Foundation (NCCR FINRISK and grants 101312-103781/1 and 100012-105745/1).

*Correspondence address: Fabio Trojani, Institute of Finance, University of Lugano, Via Buffi 13, CH-6900 Lugano, e-mail: Fabio.Trojani@lu.unisi.ch.

Learning and Asset Prices under Ambiguous Information

ABSTRACT

We propose a new continuous time framework to study asset prices under learning and ambiguity aversion. In a partial information Lucas economy with time additive power utility, a discount for ambiguity arises if and only if the elasticity of intertemporal substitution (EIS) is above one. Then, ambiguity increases equity premia and volatilities, and lowers interest rates. Very low EIS estimates are consistent with EIS parameters above one, because of a downward bias in Euler-equations-based least-squares regressions. Ambiguity does not resolve asymptotically and, for high EIS, it is consistent with the equity premium, the low interest rate, and the excess volatility puzzles.

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

JEL Classification: C60, C61, G11.

This paper studies the equilibrium asset pricing implications of learning when the distinction between risk and ambiguity (Knightian uncertainty) aversion matters. Ambiguity refers to situations where investors do not rely on a single probability law to describe the relevant random variables. Ambiguity aversion means that investors dislike ambiguity about the probability law of asset returns. In 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, where economic agents have partial information about 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 the conditional and unconditional implications for equilibrium asset prices.

It is an open issue, whether ambiguity aversion gives a plausible explanation for many salient features of asset prices when learning is accounted for. For instance, can the equity premium puzzle 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 have to consider settings where a possible misspecification of beliefs and the corresponding learning dynamics is explicitly addressed. Our approach to learning under ambiguity aversion can be interpreted as a continuous-time extension of the axiomatic setting in Knox (2004). In our model, agents learn only some global ambiguous characteristics of the underlying endowment process, as parameterized by a finite set of relevant ambiguous states of the economy. Moreover, extending Epstein and Schneider (2002), 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 of the econ-

[1]E.g., Barberis (2000), Brennan (1998), Brennan and Xia (2001), Kandel and Stambaugh (1996), Pástor (2000), Pástor and Stambaugh (2000), Veronesi (1999, 2000) and Xia (2001), among others.

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

Using the exchange economy framework, we are able to compute analytically equilibrium equity premia, equity expected returns and volatilities, interest rates and price dividend ratios. Since we allow for exogenous signals about the unobservable expected growth rate of the aggregate endowment, we can also study the relation between asset prices, information noisiness and ambiguity. However, our main focus will be on studying how learning under ambiguity aversion affects the functional form of the equilibrium variables and, more specifically, if it worsens existing asset pricing puzzles. For instance, while there is now plenty of evidence that settings of ambiguity aversion do help in explaining the equity premium and the low interest rate puzzles (see the related literature in Section 1), we also know that in a pure setting of learning the equity premium can be even more than a puzzle (see, e.g., Veronesi (2000)). Does the combination of learning and ambiguity aversion help in giving a reasonable explanations for the equity premium puzzle? Similarly, we 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). Does the combination of learning and ambiguity aversion still generate excess volatility and volatility clustering?

All above questions can be addressed directly in our model. First, we find that learning under ambiguity aversion implies an equilibrium discount for ambiguity, if and only if relative risk aversion is low (below one) or, equivalently, if the elasticity of intertemporal substitution (EIS) is large (above one). Under low risk aversion, learning and ambiguity aversion increase conditional equity premia and volatilities. Second, learning and ambiguity aversion imply lower equilibrium interest rates, irrespective of risk aversion. Thus, with low risk aversion, we get both a higher equity premium and a lower interest rate. This is a promising feature of our setting, from the perspective of explaining simultaneously the equity premium and the risk free rate puzzles without an ad hoc use of preference parameters. Third, in our model no stable relation between excess returns and conditional variances exists. This feature generates estimated relations between excess returns and equity conditional variances with an indeterminate sign over time. Fourth, we show that estimates of the EIS based on standard Euler equations for equity returns are strongly downward biased in a setting of learning and ambiguity aversion. Therefore, under learning and ambiguity aversion EIS above one can be consistent with observed estimated EIS clearly below one. Moreover, since in our setting ambiguity does not resolve asymptotically, we can 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. Finally, in a setting with low risk aversions below one and moderate ambiguity, we obtain asset pricing predictions consistent with those of the equity premium, the low interest rate and the excess volatility puzzles.

The paper is organized as follows. The next section reviews the relevant literature on learning and ambiguity. Section II introduces our setting of learning under ambiguity aversion. The properties of the optimal learning dynamics are studied in Section III. Section IV characterizes and discusses conditional asset pricing relations under our setting of learning and ambiguity aversion. Section V concludes and summarizes.

I. Background

Distinguishing between ambiguity aversion and risk aversion is both economically and behaviorally important. As the Ellsberg (1961) paradox illustrates, investors behave inherently different 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 where preferences are represented by Max-Min expected utility over a set of multiple prior distributions. More 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. A discrete-time axiomatic foundation for that model has been provided later in Epstein and Schneider (2003), showing that a dynamically consistent conditional version of

Gilboa and Schmeidler (1989) preferences is represented by means of a recursive Max-Min expected utility criterion over a set of multiple distributions. Chen and Epstein (2002) extend that setting to continuous time. A second setting of intertemporal ambiguity aversion based on an alternative form of Max-Min expected utility preferences is proposed in Hansen, Sargent and Tallarini (1999, in discrete time) and Anderson, Hansen and Sargent (2003, in continuous time). Their setting applies robust control theory to economic problems.

Continuous time models of full information economies with ambiguity aversion have been recently proposed to give plausible explanations for several important characteristics of asset prices. Examples of such models include, among others, Gagliardini et al. (2004; term structure of interest rates), Epstein and Miao (2003; home bias), Liu et al. (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, the above models exclude any form of learning. Investors observe completely the state variables determining their 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 the issue of learning under ambiguity aversion has been addressed by a few authors. In a production economy subject to exogenous regime shifts, Cagetti at. al. (2002) apply robust filtering theory to study the impact of learning and ambiguity aversion on the aggregate capital stock, equity premia and price dividend ratios. Extending the discrete-time linear-quadratic setting in Hansen, Sargent and Wang (2002) to continuous time, they analyze an economy with a power utility representative agent and nonlinear state evolution. Using numerical methods, they show that ambiguity aversion increases precautionary saving in a way that is similar to the effect of an increased subjective time preference rate, leading to an increase in the capital stock. Moreover, the equity premium increases substantially due to ambiguity aversion. Price dividend ratios turn out to be lower. Epstein and Schneider (2002) highlight in a simple discrete time setting 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) construct a similar learning model under ambiguity to study the impact of an ambiguous signal precision on asset prices. They show that an ambiguous quality of information, defined in terms of a set of possible values of the signal precision, can generate skewed asset returns and returns volatility. Knox (2004) proposes an axiomatic setting of learning about a model parameter under ambiguity aversion. He extends previous settings of ambiguity aversion by weakening the axiom of consequentialism. Consequentialism is the property that counterfactuals are neglected in the determination of conditional preferences. It is an axiom related to the separability of conditional preferences relatively to a partition of relevant conditioning events in the model. Epstein and Schneider (2002) assume consequentialist conditional preferences with respect to the partitions generated by the whole history of asset prices. Knox (2004), instead, allows generically for a less consequentialist behavior, which is consistent with a less structured partition of events. Separability of preferences is then assumed only with respect to a less structured partition. Since separability of preferences across a partition of events is inherently related to a restricted independence axiom (see Theorem 1 of Knox (2004)), assuming a too consequentialist preference structure with respect to a richly structured partition may happen to be inappropriate in a context of ambiguity aversion. Knox (2004) provides some asset pricing examples where consequentialism with respect to a partition generated by asset returns may be inappropriate in conjunction with ambiguity aversion; see also Machina (1989) for a general discussion about the usefulness of consequentialism in the context of non-expected utility models.

We can think of our model as a continuous time extension of the axiomatic setting in Knox (2004). Our representative agent is able to learn only some global ambiguous characteristics of asset prices in dependence of a finite set of fuzzy macroeconomic conditions. Therefore, our representative agent neglects counterfactuals in the determination of conditional preferences only to the extent that such counterfactuals can be determined based on the given relevant set of states of the economy.

II. The Model

We start with a simple continuous time Lucas economy. The drift rate in the diffusion process for the dividend dynamics is unobservable. Investors learn about the "true" drift through the observation of 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. In this sense, the meaning of the signal is clear, even if it is noisy. For ambiguous signals, the distribution conditional on a given parameter value is unknown or at least not uniquely identified, as in our model. 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. This feature is modeled by a set of multiple likelihoods on the underlying dividend dynamics. The size of such sets 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. For example, recessions are less well studied and understood than expansionary economic phases, because recessions are typically relatively rare events with nonhomogeneous properties over time. This pattern can be easily incorporated in our model by means of a higher degree of ambiguity, i.e., by a broader set of multiple likelihoods conditional on bad economic conditions.

Our objective is to characterize conditional and unconditional equilibrium asset returns under different assumptions on the quality of a signal. We measure quality in terms of its noisiness and ambiguity. To this end, we develop an equilibrium model of learning under ambiguity aversion consisting of the following key ingredients:

1. A parametric *reference model* dynamics for the underlying dividend process and the unobservable dividend drift. The reference model is explicitly treated as an approximation

of the reality, rather than as 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. In our setting where agents have to learn the unobservable drift of the dividend dynamics, we believe that taking into account such specification doubts is an important modeling device. We introduce the reference model in Section A.

2. A set of *multiple likelihoods* on the dynamics of the unobservable dividend drift. We use these multiple likelihoods to compute a set of multiple ahead beliefs about the unknown dividend drift dynamics. This set of multiple ahead beliefs represents the investor's ambiguity on the dynamic structure of the unobservable expected dividend growth rate. The set of multiple likelihoods can also be interpreted as a description of a class of alternative specifications to the reference model, which are statistically close and therefore difficult to distinguish from it. We introduce the set of multiple likelihoods in Sections B and C.

3. 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. We formulate the optimization problem in Section D.

Given the three key ingredients above, a set of standard market clearing conditions on good and financial markets closes the model and allows to determine 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 (2004).

A. The Reference Model Dynamics

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

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

where $\gamma < 1$. The representative investor has a parametric reference model that describes in an approximate way the dynamics of dividends D

$$\frac{dD}{D} = E_t\left(\frac{dD}{D}\right) + \sigma_D dB_D \quad , \tag{1}$$

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

$$de = E_t\left(\frac{dD}{D}\right) + \sigma_e dB_e \quad , \tag{2}$$

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 of the reality. It implies a simple geometric Brownian motion dynamics for dividends with a constant drift that can take one of a finite number of candidate values.

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

$$\frac{1}{dt}E_t(dD/D) = \theta \quad , \tag{3}$$

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, Assumption 1 is a correct specification assumption on the prevailing dividend dynamics. It 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 (1) 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. Then, the equilibrium asset returns dynamics can be determined and the pricing impact of learning can be studied analytically.

In the next sections, we weaken Assumption 1 to account for specification doubts about the unobservable dividend drift dynamics in equations (1) and (3). In contrast to Veronesi (1999, 2000) and Cagetti et al. (2002), we explicitly avoid switching regimes in Assumption 1. Our setting can be extended to include also changes in regime. However, to compare the findings obtained in our setting of ambiguity aversion to those under Bayesian learning when complete asymptotic learning is possible, we confine ourselves to Assumption 1. Reference models with changes in regimes are left as a topic for future research.

B. Multiple Likelihoods

In reality, a correct specification hypothesis of the type given in Assumption 1 is very restrictive. It assumes that even when dividend drifts are 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 where 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 well 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, it is de-

[3] The importance of this issue has been early 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), Mancini et al. (2003), Gagliardini et al. (2004) and Ortelli and Trojani (2004) for some recent work in the field.

sirable 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.

We address explicitly specification doubts by modelling agents' beliefs, conditional on any possible reference model drift θ, by means of a set of multiple likelihoods. Multiplicity of likelihoods reflects agents ambiguity on the reference model specification. To define these sets, we restrict ourselves to absolutely continuous misspecifications of the geometric Brownian motion processes in equations (1) and (3). 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 (1) and (3).

Let $h(\theta)\sigma_D$ be an adapted 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 that will be defined more precisely later on (see Assumption 3 below). Under such a likelihood, the prevailing dividend dynamics are

$$\frac{dD}{D} = E_t^{h(\theta)}\left(\frac{dD}{D}\right) + \sigma_D dB_D \quad , \tag{4}$$

with signal dynamics

$$de = E_t^{h(\theta)}\left(\frac{dD}{D}\right) + \sigma_e dB_e \quad . \tag{5}$$

In our model, 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, several possible functional forms of the drift in equation (4) are considered, together with the reference model dynamics in equations (1) and (3). The set of possible drifts implied by the multiple likelihoods in $\Xi(\theta)$ represents, in a more realistic way, the relevant beliefs of an agent who does not trust completely the reference model dynamics.

C. A Specific Set of Multiple Likelihoods

Compared with the single likelihood specification in a Bayesian context, an agent with multiple likelihood beliefs can assume a weaker form of correct specification. In a natural way, the agent can assume that, in the relevant set of multiple likelihoods, at least one likelihood fits the unknown dividend drift dynamics. Therefore, we replace Assumption 1 by a weaker assumption.

Assumption 2 *The "true" dividend drift specification is given by*

$$\frac{1}{dt} E_t^{h(\theta)} \left(\frac{dD}{D} \right) = \theta + h(\theta, t)\, \sigma_D \quad , \tag{6}$$

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 2, 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 Assumption 1. We note that a pure Bayesian assumption arises as soon as $\Xi(\theta) = \{0\}$ for all $\theta \in \Theta$. Then, agents would be concerned only with the pure noisyness of a signal about the parameter value θ. Therefore, the distinction between ambiguity and noisyness is absent in a pure Bayesian setting. However, when $\Xi(\theta) \neq \{0\}$, Assumption 2 implies a whole set of likelihoods that represent absolutely continuous misspecifications of the distributions under the reference model.

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 which 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 by a learning model about the relevant neighborhood $\Xi(\theta)$, rather than by 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, and tries to infer some more *global* characteristics of asset returns in dependence of such ambiguous macroeconomic states. We could also easily model ambiguity about the set of initial priors $(\widehat{\pi}_1, .., \widehat{\pi}_n)$ by introducing a corresponding set of multiple initial priors. However, the main implications from our analysis would not change, because the choice of the initial prior only affects the initial condition in the relevant dynamics of Π. Finally, since the size of the set $\Xi(\theta)$ can depend on the specific value of θ, our setting allows also for degrees of ambiguity that depend on economic conditions. This feature takes into account heterogenous degrees of ambiguity aversion about the processes for the relevant state variable.

We next specify the set $\Xi(\theta)$ of multiple likelihoods relevant for our setting. From a general perspective, $\Xi(\theta)$ should satisfy the following requirement:

- The set $\Xi(\theta)$ 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 requirement makes more precise the general principle that $\Xi(\theta)$ should contain only models for which agents have some well motivated specification doubt, relatively 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. Therefore, this property

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

defines a whole neighborhood of slight but otherwise arbitrary misspecifications of the reference model distributions. This is the starting point to develop optimal consumption/investment policies that are robust to any possible small misspecification of the given state dynamics. A set of multiple likelihoods satisfying the above requirement is given below.

Assumption 3 *For any $\theta \in \Theta$ we define $\Xi(\theta)$ by:*

$$\Xi(\theta) := \left\{ h(\theta) : \frac{1}{2} h^2(\theta, t) \leq \eta(\theta) \ \textit{for all } t \geq 0 \right\} \quad , \tag{7}$$

where $\eta(\theta_1), .., \eta(\theta_n) \geq 0$. Moreover,

$$(i) \sup_{h(\theta_i) \in \Xi(\theta_i)} h(\theta_i) < \sup_{h(\theta_j) \in \Xi(\theta_j)} h(\theta_j) \quad , \quad (ii) \inf_{h(\theta_i) \in \Xi(\theta_i)} h(\theta_i) < \inf_{h(\theta_j) \in \Xi(\theta_j)} h(\theta_j) \quad , \tag{8}$$

for any $i < j$.

Under Assumption 3 the discrepancy between the reference model distributions under a drift θ and those under any model implied by a drift distortion process $h(\theta) \in \Xi(\theta)$ can be constrained to be statistically small. Anderson, Hansen and Sargent (2003) show that the relative entropy between the reference model probability law and the one under any alternative candidate specifications can be constrained to be small. Relative entropy is a statistical measure of model discrepancy that can be used to bound model detection error probabilities to imply a relatively high probability of an error in model choice. In this sense, a moderate bound $\eta(\theta)$ implies for any likelihood in the set $\Xi(\theta)$ a small statistical discrepancy relative to a reference model dynamics with drift θ. Moreover, since definition (7) 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 defined by (7).

Finally, condition (8) is a monotonicity condition for the correspondence $\theta \longmapsto \Xi(\theta)$. It restricts the admissible set of multiple likelihoods to imply best and worst case dividend drifts of any admissible model neighborhood $\Xi(\theta)$ to be ranked in the same way as the reference model drifts. Condition (8) is a partial identifiability condition on the set $\Xi(\theta)$ of multiple likelihoods, because it does not exclude the case where an admissible dividend drift process is an element at the same time of two different candidate model neighborhoods. A stronger identifiability condition on a multiple likelihoods learning model would require that the classes of drift processes for D implied by two different neighborhoods $\Xi(\theta_i)$ and $\Xi(\theta_j)$, $i \neq j$, are disjoint.

Assumption 4 *The sets of $\Xi(\theta_1), .., \Xi(\theta_n)$ of candidate drift processes are such that*

$$\theta_i + \sigma_D h(\theta_i) \neq \theta_j + \sigma_D h(\theta_j) \tag{9}$$

for any $h(\theta_i) \in \Xi(\theta_i)$ and $h(\theta_j) \in \Xi(\theta_j)$ such that $i \neq j$.

Assumption 4 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. In contrast to that, condition (8) incorporates situations where the same drift process can be realistically generated under different macroeconomic conditions. Such a situation arises, e.g., when the degree of ambiguity $\eta(\theta)$ in the economy is high relatively to the distance between reference model drifts θ.

D. Ambiguity Aversion and Intertemporal Max-Min Expected Utility

Let $\mathcal{F}(t)$ denote the information available to investors at time t. This contains all possible realizations of dividends and signals. Investors learn about the dividend dynamics (4) by considering explicitly the ambiguity represented by the sets of multiple likelihoods $\Xi(\theta)$ in Assumption 3, given the prior probabilities $(\widehat{\pi}_1, .., \widehat{\pi}_n)$ at time 0. This learning mechanism will require the computation of "likelihood by likelihood" Bayesian ahead beliefs for $E_t^{h(\theta)}\left(\frac{dD}{D}\right)$ as functions of any likelihood model $h(\theta) \in \Xi(\theta)$ and given the filtration $\{\mathcal{F}(t)\}$; see also Epstein and Schneider (2002) and Miao (2001).

Let P be the price of a risky asset in the economy, r be the instantaneous interest rate and $\eta(\theta)$ be the function that describes the amount of ambiguity relevant to investors in dependence of a reference model drift θ. The investor determines optimal consumption and investment plans $C(t)$ and $w(t)$. She solves the continuous time intertemporal Max-Min expected utility optimization problem

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

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 3 holds.

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 $C(t) = D(t)$.

III. Multiple Filtering Dynamics under Ambiguity

Learning under ambiguity consists in constructing a set of standard Bayesian ahead beliefs for $E_t^{h(\theta)}\left(\frac{dD}{D}\right)$ in dependence of any likelihood $h(\theta) \in \Xi(\theta)$. Given such multiple beliefs, the solution of problem (P) in equation (10) can be found by solving an equivalent full information problem, where the dividend drift dynamics are defined in terms of the filtration $\{\mathcal{F}(t)\}$ generated by dividends and signals. In this section, we study the dynamic properties of such Bayesian ahead beliefs under different hypotheses on the relation between the underlying true dividend drift dynamics and any corresponding Bayesian prediction for $E_t^{h(\theta)}\left(\frac{dD}{D}\right)$, where $h(\theta) \in \Xi(\theta)$.

A. Bayesian Learning "Likelihood by Likelihood"

For a given admissible likelihood model $h(\theta) \in \Xi(\theta)$ let $\pi_i(t)$ be investors belief that the drift rate is $\theta_i + h(\theta_i)\sigma_D$, conditionally on past dividend and signal realizations:

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

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:

$$\frac{1}{dt} E^{h(\theta)} \left(\frac{dD}{D} \middle| \mathcal{F}(t) \right) = \sum_{i=1}^{n} (\theta_i + h(\theta_i)\sigma_D)\pi_i(t) = m_{\theta,h} \quad , \tag{11}$$

where

$$m_{\theta,h} = m_\theta + m_{h(\theta)} \quad , \quad m_\theta = \sum_{i=1}^{n} \theta_i \pi_i(t) \quad , \quad m_{h(\theta)} = \sum_{i=1}^{n} h(\theta_i)\pi_i(t)\sigma_D \quad . \tag{12}$$

The filtering equations implied by a given likelihood $h(\theta)$ are given next.

Lemma 1 *Suppose that at time zero investors beliefs are represented by the prior probabilities* $\widehat{\pi}_1, .., \widehat{\pi}_n$. *Under a likelihood* $h(\theta) \in \Xi(\theta)$ *it follows:*

1. *The dynamics of the optimal filtering probabilities vector* $\pi_1, .., \pi_n$ *is given by*

$$d\pi_i = \pi_i (\theta_i + h(\theta_i)\sigma_D - m_{\theta,h}) \left(k_D d\widetilde{B}_D^h + k_e d\widetilde{B}_e^h \right) \quad ; \quad i = 1, .., n \quad , \tag{13}$$

   where

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

   $k_D = 1/\sigma_D$, $k_e = 1/\sigma_e$. *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* $\{\mathcal{F}(t)\}$.

2. *If* $\pi_i(0) > 0$, *for every finite* t *it follows*

$$\Pr(\pi_i(t) > 0) = 1$$

for any probability Pr *equivalent to the one under the reference model.*

The dynamics (13) describe the optimal filtering probabilities of an investor using an admissible likelihood model $h(\theta) \in \Xi(\theta)$ for prediction purposes. Under such a likelihood, the filtered error process $\left(\widetilde{B}_D^h, \widetilde{B}_e^h\right)$ is a Brownian motion with respect to $\{\mathcal{F}(t)\}$. $d\widetilde{B}_D^h$ and $d\widetilde{B}_e^h$ are the normalized innovations of dividend and signal realizations under the likelihood $h(\theta)$. They enter in (13) normalized by the corresponding precision parameters k_D, k_e. Therefore, signal innovations have a higher impact than dividend innovations in agents posterior distributions when they are more precise, i.e., when $k_e > k_d$. This intuition is the same as in Veronesi (2000). The loading factor of $k_D d\widetilde{B}_D^h + k_e d\widetilde{B}_e^h$ in (13) depends on the likelihood $h(\theta)$ used to build a prediction belief out of the set $\Xi(\theta)$ of admissible likelihoods. Therefore, the choice of the likelihood affects the subjective variability of the posterior probabilities Π over time. In the special case where the likelihood $h(\theta) \in \Xi(\theta)$ is such that $h(\theta_1) = ... = h(\theta_n)$, i.e., when perceived reference model misspecifications are $\theta-$independent, this effect disappears.

B. Bayesian Learning and Model Misspecification

To gain more intuition about the implications of the above results, we express (13) in terms of the original Brownian motions B_D and B_e. This exercise yields the description of the process $d\pi_i$ from the perspective of an outside observer knowing exactly the underlying dynamics of dividends, i.e., knowing the true reference model dividend drift θ and the true local distortion, $h_D\sigma_D$, that define the underlying dividend drift process. We can then gauge how, in a standard Bayesian learning process, a likelihood misspecification affects the dispersion and the dynamics of the perceived beliefs.

Corollary 1 *Let $h(\theta) \in \Xi(\theta)$ be an admissible likelihood. If the true dividend dynamics are given by a drift distortion process h_D then*

$$d\pi_i = \pi_i \left(\theta_i + h(\theta_i)\sigma_D - m_{\theta,h}\right)\left[k\left(\theta + h_D\sigma_D - m_{\theta,h}\right)dt + k_D dB_D + k_e dB_e\right] \quad , \tag{14}$$

where $k = k_D^2 + k_e^2$. In particular, if the likelihood model $h(\theta)$ is correctly specified in a standard Bayesian sense, i.e., if $\theta + h_D \sigma_D = \theta_l + h(\theta_l)\sigma_D$ for some $\theta_l \in \Theta$, then

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

Expressions (14), (15) give the dynamics of the posterior probability π_i under different assumptions on the correct specification of the likelihood $h(\theta)$. Equation (15) gives the learning dynamics for the case where the likelihood model $h(\theta)$ is correctly specified. If θ_l defines the "true" dividend drift $\theta_l + h(\theta_l)\sigma_D$, then the drift of π_l in the dynamics (15) is always positive and it is quadratically increasing in the distance between the true drift $\theta + h(\theta)\sigma_D$ and the posterior expectation $m_{\theta,h}$. This features tends to increase the posterior probability π_l over time and to narrow the distance between the true drift and the posterior expectation $m_{\theta,h}$. As this happens, the variance of $d\pi_l/\pi_l$ shrinks so that eventually the probability π_l will converge to 1 and agents will fully learn the dynamic structure of the dividend drift process $\theta_l + h(\theta_l)\sigma_D$. This argument gives the next corollary.

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

Under a correctly specified likelihood model $h(\theta) \in \Xi(\theta)$ agents will thus learn asymptotically the correct process $\theta_l + h(\theta_l)\sigma_D$ for the underlying dividend drift. Intuitively, the same cannot be expected generally for a Bayesian learning process based on a misspecified likelihood $h(\theta)$. To highlight the basic point we can study the learning dynamics for the simplified setting with only two possible reference model dividend drift values. Equation (14) gives the relevant learning dynamics for the case where the likelihood is misspecified.

Example 1 Consider the following simplified model structure:

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

Let $\theta_1 + h_D \sigma_D$ be the true underlying dividend drift process. Then, equation (1) implies the learning dynamics:

$$
\begin{aligned}
d\pi_1 &= \pi_1 \left(\theta_1 - m_{\theta,h}\right) \left[k \left(\theta_1 + h_D \sigma_D - m_{h,\theta}\right) dt + k_D dB_D + k_e dB_e\right] \\
&= \pi_1 \left(1 - \pi_1\right) \left(\theta_1 - \theta_2\right) \left[k \left(\theta_1 + h_D \sigma_D - m_{h,\theta}\right) dt + k_D dB_D + k_e dB_e\right] \quad . \qquad (16)
\end{aligned}
$$

From Example 1, we see immediately that if

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

then $\pi_1 \to 1$ ($\pi_1 \to 0$) almost surely as $T \to \infty$. Under these conditions investors will therefore "learn" asymptotically a constant dividend drift process θ_1 (θ_2) even if the true one $\theta_1 + h_D \sigma_D$ is possibly time varying in a nontrivial and unpredictable way. In particular, this remark implies that we will always have $\pi_1 \to 1$ ($\pi_1 \to 0$) as $T \to \infty$ for all settings where the true drift $\theta_1 + h_D \sigma_D$ is uniformly lower than θ_1 (higher than θ_2). In the more general case where $\theta_1 + h_D \sigma_D$ is 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

We plot two different trajectories of π_1 under a dividend drift process such that

$$
\theta_1 + h_D(t)\, \sigma_D = \begin{cases} \left(\theta_1 + \theta_2\right)/2 + a & t \in [k, k+1)\,, \\ \left(\theta_1 + \theta_2\right)/2 - a & t \in [k+1, k+2) \end{cases} \quad , \qquad (17)
$$

where $k \in \mathbb{N}$ is even. The process (17) describes a simple deterministic and piecewise constant dividend drift misspecification. More complex (possibly nonparametric) misspecifications can be also considered. However, the main message of Figure 1 would not change.

Figure 1 shows that under a dividend drift process (17) 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 (16) are the points $\pi_1 = 1$ and $\pi_1 = 0$. Any value $\pi_1 \in (0, 1)$ such that

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

makes the drift, but not the diffusion, equal to zero in the dynamics (16). Consequently, π_1 will never stabilize asymptotically in regions such that $m_{h,\theta} \approx \theta_1 + h_D \sigma_D$. An asymptotic behavior such that $m_{h,\theta} \approx \theta_1 + h_D \sigma_D$ would be ideally more natural, if the goal is to approximate adequately $\theta_1 + h_D \sigma_D$ by means of $m_{h,\theta}$, even under a misspecified likelihood. However, under the given misspecified likelihood it will never arise. Richer, but qualitatively similar, patterns emerge when the set of possible states of the economy is enlarged.

Figure 2 presents the prevailing posterior probabilities dynamics in a setting where dividends indeed follow a geometric Brownian motion and the given learning model is misspecified in a very simplified way.

Insert Figure 2 about here

In that case, we observe convergence of different posterior probabilities to 1 (Panel (A) and (C)) and, in some cases, a dynamics that does not converge over the given time horizon (Panel (B)).

The above discussion highlights in a simple setting that under a possibly slightly misspecified likelihood a Bayesian investor will not be able to evaluate exactly the utility of a consumption/investment strategy, because she will never identify exactly the underlying dividend drift process, even asymptotically. Even if the amount of misspecification is moderate, it is then highly possible that it affects significantly the realized utility of optimal policies under the reference model's assumptions. We therefore work with a setting of learning where investors explicitly exhibit some well founded specification doubts about the given reference model. These misspecification doubts are described by means of our sets $\Xi(\theta)$ of indistinguishable multiple likelihoods for the dividend drift dynamics. Such sets of multiple likelihoods depict investor's ambiguity about all finite dimensional distributions of dividends.

C. Learning under Ambiguity

Under a likelihood misspecification a standard Bayesian learning process does not generally lead to learn the correct dividend drift dynamics. Moreover, candidate drift distortions $h(\theta) \in \Xi(\theta)$ are hardly statistically distinguishable from the reference model drift dynamics using observations on D and e. Which learning behavior should agents adopt in this case? 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

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

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

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

with the Brownian motions $\widetilde{B}_D^h$, $\widetilde{B}_e^h$, with respect to the filtration $\{\mathcal{F}(t)\}$ and under the likelihood $h(\theta)$. 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, the continuous time optimization problem (10) can be written as a full information problem where the relevant dynamics are defined in terms of the filtration $\{\mathcal{F}(t)\}$. The relevant problem reads

$$(P) \quad : \quad \max_{C,w} \inf_{h(\theta)} E\left[\int_0^\infty u(C,s)\,ds \,\middle|\, \mathcal{F}(0)\right] \tag{18}$$

subject to the dividend and wealth dynamics

$$dD = m_{\theta,h} D dt + \sigma_D D d\widetilde{B}_D^h \quad , \tag{19}$$

$$d\pi_i = \pi_i \left(\theta_i + h(\theta_i)\sigma_D - m_{\theta,h}\right)\left(k_D d\widetilde{B}_D^h + k_e d\widetilde{B}_e^h\right) \quad , \tag{20}$$

$$dW = W\left[w\left(\frac{dP + Ddt}{P}\right) + (1-w)\,rdt\right] - Cdt \tag{21}$$

where for any $\theta \in \Theta$ the standardized drift distortion is such that $h(\theta) \in \Xi(\theta)$ and under Assumption 3. In contrast to a standard (single-likelihood) Bayesian setting of learning, in (18) investors are requested to select optimally both the forecast procedure for the unknown dividend drift *and* the associated consumption/investment policies.

D. Intertemporal Max-Min Expected Utility and Asset Prices

In equilibrium, the optimization problem (18) of our representative agent reads:

$$(P) \quad : \quad J(\Pi, D) = \inf_{h(\theta)} E\left[\int_0^\infty e^{-\delta t}\frac{D_t^{1-\gamma}}{1-\gamma}dt \,\middle|\, \mathcal{F}(0)\right] \quad , \tag{22}$$

subject to the dynamics

$$dD = m_{\theta,h} D dt + \sigma_D D d\widetilde{B}_D^h \quad , \tag{23}$$

$$d\pi_i = \pi_i \left(\theta_i + h(\theta_i)\sigma_D - m_{\theta,h}\right)\left(k_D d\widetilde{B}_D^h + k_e d\widetilde{B}_e^h\right) \quad , \tag{24}$$

where for any $\theta \in \Theta$ the standardized drift distortion is such that $h(\theta) \in \Xi(\theta)$ and under Assumption 3. The solution of this problem is presented in the next proposition.

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

$$\widehat{\theta}_i + (1-\gamma)\,\sqrt{2\eta\,(\theta_i)}\sigma_D > 0 \quad , \quad i = 1, .., n \quad . \tag{25}$$

Then, we have:

1. *The normalized misspecification* $h^*(\theta)$ *solving problem (22) is given by*

$$h^*(\theta_i) = -\sqrt{2\eta\,(\theta_i)} \quad , \quad i = 1, .., n \quad . \tag{26}$$

2. *The value function* J *to problem (22) is given by*

$$J(\Pi, D) = \frac{D^{1-\gamma}}{1-\gamma}\sum_{i=1}^{n} \pi_i C_i \quad , \tag{27}$$

where

$$C_i = \frac{1}{\widehat{\theta}_i + (1-\gamma)\,\sqrt{2\eta\,(\theta_i)}\sigma_D} \quad , \quad i = 1, .., n \quad . \tag{28}$$

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

$$P(\Pi, D) = D\sum_{i=1}^{n} \pi_i C_i \tag{29}$$

4. *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

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

In Proposition 1, each constant of the form (28) represents investors expectation of discounted lifetime dividends, conditional on a constant dividend drift process $\theta_i - \sqrt{2\eta\,(\theta_i)}\sigma_D$ and normalized to make it independent of the current level of dividends. 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 $\Xi(\theta_i)$. The discounting factor is the intertemporal marginal substitution rate of aggregate consumption, given by:

$$\frac{u_c\left(D\left(t\right),t\right)}{u_c\left(D\left(s\right),s\right)}=e^{-\delta(t-s)}\left(\frac{D\left(t\right)}{D\left(s\right)}\right)^{-\gamma} \quad . \tag{32}$$

More specifically, we thus have:

$$C_i=E^{h^*(\theta_i)}\left[\int_s^\infty e^{-\delta(t-s)}\left(\frac{D\left(t\right)}{D\left(s\right)}\right)^{1-\gamma}dt\right]=\frac{1}{D\left(s\right)}E^{h^*(\theta_i)}\left[\int_s^\infty \frac{u_c\left(D\left(t\right),t\right)}{u_c\left(D\left(s\right),s\right)}D\left(t\right)dt\right] \quad , \tag{33}$$

where $E^{h^*(\theta_i)}\left[\cdot\right]$ denotes expectations under a geometric Brownian motion process for D with drift $\theta_i-\sqrt{2\eta\left(\theta_i\right)}\sigma_D$. A high C_i implies that investors are willing to pay a high price for the ambiguous state $\Xi\left(\theta_i\right)$. Since the state is not observable, they weight each C_i by the posterior probability π_i to get the price (29) of the risky asset under learning and ambiguity aversion. We remark that C_i is a function of both investors' ambiguity aversion, via the parameter $\eta\left(\theta_i\right)$, and investor's relative risk aversion γ. To make this dependence more explicit, it is easy to see that (33) can be equivalently written as:

$$\begin{aligned} C_i &= \frac{1}{D\left(s\right)}E\left[\int_s^\infty e^{-(1-\gamma)\sqrt{2\eta(\theta_i)}\sigma_D(t-s)}\frac{u_c\left(D\left(t\right),t\right)}{u_c\left(D\left(s\right),s\right)}D\left(t\right)dt|\theta=\theta_i\right] \\ &= \frac{1}{D\left(s\right)}E\left[\int_s^\infty e^{-\left[\delta+(1-\gamma)\sqrt{2\eta(\theta_i)}\sigma_D\right](t-s)}\left(\frac{D\left(t\right)}{D\left(s\right)}\right)^{1-\gamma}dt|\theta=\theta_i\right] \quad , \end{aligned} \tag{34}$$

where $E\left[\cdot|\theta=\theta_i\right]$ 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 $\Xi\left(\theta_i\right)$ is equivalent to the one implied by a corrected time preference rate

$$\delta\longrightarrow\delta+\left(1-\gamma\right)\sqrt{2\eta\left(\theta_i\right)}\sigma_D \tag{35}$$

under the reference model dynamics. The adjustment (35) depends on the amount of ambiguity of the ambiguous state $\Xi\left(\theta_i\right)$, relative risk aversion γ and dividend growth volatility σ_D. Cagetti et al. (2002) observe that ambiguity aversion decreases the aggregate capital stock in way that is similar to the effect of an increased subjective discount rate. The adjustment (35) suggests that their finding can be rationalized analytically in our setting. However, notice that in our general case of an heterogeneous degree of ambiguity $\eta\left(\theta\right)$ the final effect of ambiguity on equity prices cannot be mapped into an adjustment of one *single* time preference rate. The above remarks imply directly the next result.

Corollary 3 *The following statements hold:*

1. *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.
2. *The price of any ambiguous state* $\Xi(\theta)$ *is an increasing function in the degree of ambiguity* $\eta(\theta)$ *if and only if* $\gamma > 1$,
3. *The price of any ambiguous state* $\Xi(\theta)$ *is independent of the degree of ambiguity* $\eta(\theta)$ *for* $\gamma = 1$.

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 $\gamma > 1$, one obtains the somewhat counterintuitive implication[5] that the price of an ambiguous state is higher than the one of an unambiguous one. This is illustrated in Figure 3 where we plot C_i as a function of $\eta(\theta_i)$ for different risk aversion levels $\gamma = \{0.1, 0.5, 1, 3, 5\}$.

Insert Figure 3 about here

As stated, for $\gamma < 1$ function C_i is a decreasing convex function of $\eta(\theta_i)$. The convexity of such function is higher for lower risk aversion parameters. For $\gamma > 1$ function C_i is an increasing function of $\eta(\theta_i)$ with can be both locally convex and locally concave over some different zones of the parameter space (see, e.g., the plot for $\gamma = 3$).

To understand this finding, recall that in the determination of C_i the representative investor discounts future dividends trough their marginal utility (32). Therefore, in equilibrium a higher dividend growth rate implies a higher expected future consumption growth and a higher discount rate. Since the effect on the discount rate dominates for $\gamma > 1$, a lower expected

[5] Relatively, e.g., to the basic intuition provided by the standard (static) Ellsberg (1961) paradox.

dividend growth (due to a conservative concern for ambiguity) implies a lower discount rate and a higher price for ambiguous states. In other words, investors with high relative risk aversion increase their hedging demand for equity when they expect low consumption growth. Under ambiguity aversion, such highly risk averse investors tend to understate actual consumption growth and increase further their hedging demand for risky assets. Since the supply of the risky asset is fixed and the riskless bond is in zero net supply, such an excess demand increases the price of the risky asset relative to dividends. Therefore, in the presence of intertemporal hedging against bad consumption news, the equilibrium impact of ambiguity aversion can be different from the one expected under a simple constant opportunity set setting as, e.g., in Maenhout (2004) or Trojani and Vanini (2002).

There is some recent work in the literature on the question of which are typical risk aversion parameters of agents that are confronted with an ambiguous environment. Suggestive in this context is the evidence collected by the experimental work of Wakker and Deneffe (1996; e.g., the graphs on p. 1143), who estimate a virtually linear utility function with a utilities elicitation procedure that is robust to the presence of ambiguity. In the same experiments, the utility functions estimated by procedures that are not robust to the presence of ambiguity were clearly concave. These findings suggest that the high risk aversions estimated in some experimental research can be also due to some pronounced deviations from expected utility. In our setting, such deviation from expected utility arises from Max-Min expected utility preferences reflecting ambiguity aversion. Therefore, excess returns will incorporate both risk aversion and ambiguity aversion in equilibrium. For $\gamma > 1$, the equilibrium impact of ambiguity is dominated by the effect on the equilibrium stochastic discount factor. As we show below, this implies that settings of learning and ambiguity aversion with high risk aversions will deliver low (negative) equity premia and low volatilities, together with high and highly variable interest rates. That is, imposing high risk aversions worsens the well known asset pricing puzzles when learning under ambiguity aversion is considered. Therefore, we focus in the sequel on settings with moderate risk aversions.

Assumption 5 *The representative agent in the model has a relative risk aversion parameter* $\gamma < 1$.

Since we adopted a setting with power utility of consumption, Assumption 5 is equivalent to assuming an elasticity of intertemporal substitution (EIS) $1/\gamma > 1$. Therefore, we focus in Assumption 5 on EIS larger than one. Such interpretation is perfectly consistent with the idea that excess returns are going to reflect mainly some premium for ambiguity, rather than a premium for risk. 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 much larger EIS than individuals which 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 et al. (2002) find with UK data 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 et al. (2004) estimate EIS above one using Euler equations for treasury bills where consumption is measured by consumption of luxury goods. Typically, in such empirical studies the EIS estimated for US data in Euler equations including 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 et al. (2004, p. 22) such a 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 Zin (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 estimates of the 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.

We note that, in our setting of learning under ambiguity, the downward biases in the EIS estimates are particularly large for Euler equations using equity returns; see Section D below.

D.1. Price Dividend Ratios

Under Assumption 5 we have from Proposition 1 a few nice and simple implications for the behavior of the price dividend ratio P/D in the model. They are summarized by the next result.

Corollary 4 *Under Assumption 5 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) The price dividend ratio P/D is an increasing convex function of the reference model expected growth rates $(\theta_1, .., \theta_n)$ in the economy.

c) 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 of the economy.

Finding *a)* in Corollary 4 is a direct implication of (29) and (34). The convexity statement follows from the convexity of the price function C_i in (28) as a function of the ambiguity parameter $\eta(\theta_i)$. Findings *b)* and *c)* follow from the fact that under Assumption 5 the constants C_i in (29) are increasing convex functions of the reference model expected growth rate of the economy. Under Assumption 5, 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. Moreover, price dividend ratios are increasing in the economy's expected growth rate (Finding *b)*), a well documented empirical fact.

Insert Figure 4 about here

In Figure 4, we plot in Panels (A) through (C) for different values of γ the price dividend ratio as a function of the amount of ambiguity in the economy. For simplicity of exposition, we assume in all graphs a homogenous degree of ambiguity $\eta(\theta) = \eta$ across the different states $\theta_1, .., \theta_n$. Panels (D) and (E) display the different distributions Π used to compute the price dividend functions in Panels (A) through (C). Panel (F) presents a mean preserving spread of the prior distribution plotted in Panel (D), while Panel (E) presents a mean preserving spread of the prior distribution plotted in Panel (F). For any level of risk aversion γ, we see that when going from Panel (A) to Panel (C) the price dividend ratio function is shifted upward, according to the mean preserving spreads in Panel (F) and Panel (E). The shift is larger for moderate risk aversions because of the higher convexity of the functions C_i as functions of θ_i. Moreover, for any given prior distribution and level γ of risk aversion we see in Panel (A) through (C) that P/D is a convex decreasing function of the amount of ambiguity $\eta(\theta)$ in the economy. The convexity of such function is higher for lower risk aversion parameters.

D.2. Equilibrium Interest Rate

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 $(\eta(\theta_1), .., \eta(\theta_n))$. The equilibrium interest rate r_{NA} in the absence of ambiguity is obtained by setting $\eta(\theta) = 0$ for all $\theta \in \Theta$ in (30),

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

Hence,

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

Figure 5 illustrates the prevailing equilibrium interest rate under learning and ambiguity aversion as a function of the size of ambiguity in the economy.

Insert Figure 5 about here

The decreasing convex pattern of interest rates as a function of the ambiguity size is more pronounced for higher risk aversions. As expected, higher risk aversions increase equilibrium interest rates via a lower EIS. The ambiguity parameter, instead, has a sizable reduction effect on interest rates.

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

$$r = \delta + \gamma\left(\theta_l - \sqrt{2\eta(\theta_l)}\sigma_D\right) - \frac{1}{2}\gamma(\gamma+1)\sigma_D^2 \quad . \tag{36}$$

The interest rate (36) 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 which is the precise drift $\theta_l + h(\theta_l)\sigma_D \in \Xi(\theta_l)$ that generated the dividend dynamics, even if she learned that the relevant model neighborhood is $\Xi(\theta_l)$. Therefore, a premium for this residual ambiguity persists asymptotically.

For the case where 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 $\Xi(\theta_1)$,.., $\Xi(\theta_n)$ and is time varying. The weights are given by the posterior probabilities Π. Settings of ambiguity aversion but no learning have been studied in Sbuelz and Trojani (2002) under an *exogenous* time varying degree of ambiguity. The key difference between such settings of ambiguity aversion and our setting lies in the fact that the state variables dynamics driving the time varying ambiguity in our economy is endogenously determined, because it depends itself on the degree of ambiguity $\eta(\theta)$ in the economy.

D.3. Endogenous Learning Dynamics

The normalized worst case drift distortion (26) in 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 knowing:

a) that the dividend dynamics indeed satisfies the reference model in (1) and (2).

b) 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 a different learning dynamics than the optimal Bayesian one under the reference model's likelihood. This is highlighted by the next Corollary.

Corollary 5 *Under the reference model in (1) and (2), the filtered probabilities dynamics of a representative agent solving the equilibrium optimization problem (22) are:*

$$d\pi_i = \pi_i \left(\theta_i - \sqrt{2\eta\left(\theta_i\right)}\sigma_D - m_{\theta,h^*}\right)\left[k\left(\theta - m_{\theta,h^*}\right)dt + k_D dB_D + k_e dB_e\right] \tag{37}$$

Equation (37) 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 a learning dynamics of a Bayesian investor. To highlight this point, we consider for simplicity the case of a constant ambiguity aversion $\eta\left(\theta_1\right) = ... = \eta\left(\theta_n\right) = \eta$. This gives the dynamics

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

For $\eta = 0$, the dynamics (38) are those of a standard (single likelihood) Bayesian learner. Under ambiguity aversion, the drift

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

is larger for above average reference model drifts θ $(\theta - m_\theta > 0)$ and lower for below average reference model drifts θ $(\theta - m_\theta < 0)$. Therefore, investors subject to ambiguity aversion will tend to "learn" more rapidly a large reference model drift than a low reference model drift. Unconditionally, this will imply learning dynamics where the a posteriori expected reference model drift m_θ under ambiguity aversion will be higher than the one of a Bayesian investors, i.e., the learning dynamics under ambiguity aversion will imply an *optimistic* tendency to overstate the a posteriori reference model drifts relatively to a standard Bayesian prediction. Such a tendency will be more apparent for large precision parameters k.

Figure 6 illustrates the above 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).

Insert Figure 6 about here

In Panel (A), the uniformly higher probabilities π_1 arise in the absence of ambiguity (the straight line corresponding 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 with opposite direction, arise for the probabilities π_3 of the "good" state θ_3 in Panel (B).

IV. 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

$$dR = \frac{dP + Ddt}{P} - rdt \quad . \tag{39}$$

We summarize our findings in the next proposition.

Proposition 2 *Under the reference model dynamics, the equilibrium excess return process R under ambiguity aversion has dynamics*

$$dR = \mu_R dt + \sigma_D d\widetilde{B}_D + V_{\theta,h^*}\left(k_D d\widetilde{B}_D + k_e d\widetilde{B}_e\right) \quad , \tag{40}$$

where

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

$$V_{\theta,h^*} = \sum_{i=1}^{n} \frac{\pi_i C_i\left(\theta_i - \sqrt{2\eta\left(\theta_i\right)}\sigma_D\right)}{\sum_{i=1}^{n}\pi_i C_i} - m_{\theta,h^*} \quad , \tag{41}$$

and with the Brownian motions increments

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

with respect to the filtration $\{\mathcal{F}_t\}$.

We can now analyze in more detail how learning under ambiguity aversion affects the conditional structure of asset returns.

A. Investors' Uncertainty on Ambiguous Model Neighborhoods

The reference model excess return process (40) is characterized by the quantities V_{θ,h^*} and $m_{h^*(\theta)}$. The quantity $m_{h^*(\theta)} = m_{\theta,h^*} - m_\theta$ is a conservative correction to the reference model's a posteriori expectations m_θ. Such quantity accounts for the possibility of misspecification doubts when computing a posteriori expectations for the growth rate of the economy and is always negative. The quantity V_{θ,h^*} reflects the difference between the worst case expected growth rate of the economy, m_{θ,h^*}, and its value adjusted counterpart. More precisely, define a value adjusted distribution $\overline{\Pi} = (\overline{\pi}_1, .., \overline{\pi}_n)$ by

$$\overline{\pi}_i = \frac{\pi_i C_i}{\sum_{j=1}^{n}\pi_j C_j} \quad ,$$

and let

$$\overline{m}_{\theta,h^*} := \overline{m}_\theta + \overline{m}_{h^*(\theta)} := \overline{E}\left(\theta - \sqrt{2\eta\left(\theta\right)}\sigma_D \middle| \mathcal{F}_t\right) = \sum_{j=1}^{n} \overline{\pi}_j\left(\theta_j - \sqrt{2\eta\left(\theta_j\right)}\sigma_D\right) \quad ,$$

be the worst case expected growth rate of the economy under the value adjusted distribution $\overline{\Pi}$. It then follows

$$V_{\theta,h^*} = \overline{m}_{\theta,h^*} - m_{\theta,h^*} \quad ,$$

showing that V_{θ,h^*} gives the distance between the value adjusted worst case expected growth rate and the worst case expected growth rate of the economy. Hence, V_{θ,h^*} is a summary of the investor's uncertainty about the future worst case growth rate of the economy, as well as of the ambiguity intrinsic in their own valuation of the asset. For example, if no uncertainty about the true model neighborhood is present (i.e., if Π is degenerate at $\Xi\left(\theta_l\right)$, say) or if all worst case states $\theta - \sqrt{2\eta\left(\theta\right)}\sigma_D$ are valued identically, then

$$V_{\theta,h^*} = 0 \quad .$$

Notice, however, that under Assumption 5 different worst case states have to be valued differently. Therefore, under Assumption 5 we can have $V_{\theta,h^*} = 0$ if and only if Π is degenerate, i.e., if and only if no uncertainty about the ambiguous model neighborhoods $\Xi\left(\theta_1\right),..,\Xi\left(\theta_n\right)$ is present. In fact, V_{θ,h^*} will be larger either when agents have more diffuse beliefs about $\Xi\left(\theta_1\right),..,\Xi\left(\theta_n\right)$ or when they value the asset very differently across the different states. These differences in valuation depend on the heterogeneity on the worst case growth rate $\theta - \sqrt{2\eta\left(\theta\right)}\sigma_D$ across such states. In fact, under Assumption 3 we can characterize the sign of V_{θ,h^*} as follows.

Lemma 2 *Under Assumption 3 we have:*

1. *(i)* $V_{\theta,h^*} > 0$ *and* $\overline{m}_\theta - m_\theta > 0$ *if and only if* $\gamma < 1$. *(ii)* $V_{\theta,h^*} < 0$ *and* $\overline{m}_\theta - m_\theta < 0$ *if and only if* $\gamma > 1$. *(iii)* $V_{\theta,h^*} = 0$ *and* $\overline{m}_\theta - m_\theta = 0$ *if and only if* $\gamma = 1$.
2. *Let* $\widetilde{\Pi}$ *be a mean reserving spread of* Π. *Then, under Assumption 5:*

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

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

3. V_{θ,h^*} is decreasing in γ.

Given Assumption 5, part 1 of Lemma 2 says that the value adjusted distribution gives more weight to high worst case model growth states $\theta - \sqrt{2\eta(\theta)}\sigma_D$. This is a direct implication of the fact that the weights C_i are increasing convex functions of θ_i and $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$. The opposite holds for $\gamma > 1$.

Part 2 of Lemma 2 shows that an increase in the uncertainty about the worst case growth rate of the economy increases V_{θ,h^*}. This finding is again a consequence of the convexity of the weights C_i as functions of $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$ for $\gamma < 1$. In such a case, the value adjusted distribution $\widetilde{\Pi}$ becomes even more skewed to high-value worst case drift states. Part 3 follows from the less convex behavior of C_i as a function of $\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$ under a higher risk aversion γ.

B. Equilibrium Equity Premia

From equation (41) the equilibrium equity premium μ_R perceived by the representative agent in our economy is given by

$$\mu_R = \gamma\left(\sigma_D^2 + V_{\theta,h^*}\right) - m_{h^*(\theta)}\left(1 + kV_{\theta,h^*}\right) \quad , \tag{42}$$

and is the sum of two terms. The term

$$\mu_R^{wc} = \gamma\left(\sigma_D^2 + V_{\theta,h^*}\right) \quad , \tag{43}$$

is the equity premium perceived by an investors under the worst case learning dynamics implied by the worst case likelihood $h^*(\theta)$ in Proposition 1. The term $\gamma\sigma_D^2$ is the equilibrium worst case equity premium of an economy with ambiguity aversion but no learning (see Maenhout (2004) and Trojani and Vanini (2002)). The direct contribution of learning and ambiguity aversion to worst case equity premia μ_R^{wc} can be studied by considering equation (43) for a fixed posterior probability vector Π. More specifically, part 1 of Lemma 2 implies that $V_{\theta,h^*} > 0$

under Assumption 5. Hence, in that case the contribution to worst case equity premia μ_R^{wc} is positive. The impact of ambiguity aversion on V_{θ,h^*} works through the weights C_i which reflect the joint price for risk and ambiguity of a model neighborhood $\Xi(\theta_i)$. From part 2 of Lemma 2, a more diffuse posterior distribution on the neighborhoods $\Xi(\theta)$ increases worst case equity premia under Assumption 5. Thus, a low signal precision in the economy can be expected to imply higher worst case equity premia, and vice versa. In that case, the relation between signal noisiness and worst case equity premia in our setting of learning under ambiguity aversion is positive.

When comparing with a setting without ambiguity ($\eta(\theta) = 0$ for all $\theta \in \Theta$), it is interesting to remark that the actual equity premium μ_R is bounded from above as a function of risk aversion γ; Veronesi (2000), Proposition 3. Hence, the Mehra and Prescott (1985) equity premium puzzle becomes even more puzzling under a noisy but unambiguous information, because the actual premium cannot be matched by a high risk aversion. In our setting of ambiguity aversion, this feature implies directly that μ_R^{wc} will be bounded as a function of risk aversion. However, fitting worst case premia μ_R^{wc} by means of moderate risk aversions is a less ambitious task than fitting actual premia μ_R. Since V_{θ,h^*} is continuous and decreasing in γ (recall Lemma 2, statement 3) also the actual equity premium μ_R is bounded as a function of risk aversion. However, for any given risk aversion parameter, actual equity premia μ_R depend strongly on the ambiguity index $\eta(\theta)$. Therefore, the ambiguity parameter $\eta(\theta)$ is the one that can be used to fit $\mu_R - \mu_R^{wc}$ (and hence μ_R) to the data. The dependence of μ_R and μ_R^{wc} on γ is illustrated by Figure 7 for some parameter choices in the model under an homogeneous degree of ambiguity $\eta(\theta) = 0.01$.

Insert Figure 7 about here

From Figure 7 we observe in Panel (C) a concave dependence of μ_R on γ, with a maximum at $\gamma = 0.5$.[6] For such a risk aversion level the worst case premium is about 0.2%. The actual equity premium μ_R in Panel (D) is, instead, a *monotonically decreasing* function of risk aversion. However, in our setting such a pattern is quantitatively well compatible with the

[6]This finding is consistent with Proposition 3 in Veronesi (2000).

empirical predictions of the equity premium puzzle. E.g., for moderate risk aversions γ between 0.3 and 0.5, the actual premium ranges from about 11% to about 7%. This effect arises despite the small size of the ambiguity parameters used; see Panel (B) of Figure 7. Panels (B) and (D) of Figure 8 highlight the effect on equity premia and worst case premia for some examples of heterogenous degrees of ambiguity in the economy.

Insert Figure 8 about here

In Figure 8, parameters and prior structure Π are equal to those in Figure 7. Panels (A-1) through (A-3) highlight the different forms of function $\eta(\theta)$ underlying all graphs. For comparison, we choose these functions in a way that preserves the same weighted entropy measure $\frac{1}{2}\sum \pi_i h^2(\theta_i)$ as in Figure 7. We observe the highest equity premia and worst case premia for the asymmetric function $\eta(\theta)$ in Panel (A-2) (the lines in Panel (B), (D) marked with '$*$'). Therefore, those ambiguity structures yield the highest premia precisely in the case where the ambiguity associated with "bad" economic states is higher. In fact, the worst case premia in Panel (D) implied by the ambiguity function in Panel (A-2) are even larger than the actual premia of a setting with no ambiguity (the dotted line in Panel (C)).

To understand the findings in Figure 7 and 8 we can study the reference model equity premium μ_R relative to the worst case premium μ_R^{wc}, as given by

$$\mu_R - \mu_R^{wc} = -m_{h^*(\theta)}\left(1 + kV_{\theta,h^*}\right) \quad . \tag{44}$$

Notice, that such a quantity depends on the risk aversion parameter γ indirectly through V_{θ,h^*}. From Lemma 2, V_{θ,h^*} is positive under Assumption 5 and decreasing in γ. The equity premium component

$$-m_{h^*(\theta)}kV_{\theta,h^*} \tag{45}$$

characterizes the marginal impact of learning under ambiguity aversion on equity premia. Therefore, under Assumption 5 the difference $\mu_R - \mu_R^{wc}$ is positive and decreasing in γ. Since for moderate risk aversions the excess premium $\mu_R - \mu_R^{wc}$ is positive, an accurate choice of $\eta(\theta)$ can help in fitting more satisfactorily the equity premia implied by the observed data for a given signal precision parameter k. Furthermore, from (45) we also observe that a less noisy signal (i.e., a higher signal precision parameter k) implies a higher reference model equity premium, given a posterior probability vector Π and taking V_{θ,h^*} constant.

The above observation gives a striking implication of learning under ambiguity aversion. In our model, a precise (i.e., a non-noisy) signal on an ambiguous model neighborhood can imply an excess reward for being exposed to such ambiguity. Therefore, noisiness and ambiguity aversion both influence the broad quality of a signal and its effects on asset prices.[7] We note that this feature makes the sign of the full impact of a higher signal precision on reference model premia μ_R indeterminate. Indeed, while a higher signal precision k tends to reduce V_{θ,h^*} (see part 3 of Lemma 2), the precision k also impacts directly (45) in a positive way. The equilibrium excess premium under ambiguity but no learning arises by setting $\widehat{\pi}_l = 1$ for the true model neighborhood $\Xi\left(\theta_l\right)$, to get

$$\mu_R - \mu_R^{wc} = \sqrt{2\eta\left(\theta_l\right)}\sigma_D \quad .$$

The contribution of ambiguity aversion to the asymptotic equity premium is given by a first order effect of ambiguity that is proportional to dividend volatilities σ_D. In particular, such an equity premium for ambiguity does not disappear asymptotically, even under an asymptotic learning about $\Xi\left(\theta_l\right)$. Such an asymptotic premium rewards the representative agent for the residual ambiguity about the precise drift, which generated the dividend dynamics out of a relevant neighborhood $\Xi\left(\theta_l\right)$. The equilibrium excess premium under ambiguity and without learning arises as the limit of a sequence of equity premia in partial information economies with ambiguity aversion.

[7] In an economy with no ambiguity one has $\mu_R = \mu_R^{wc}$. In this case, worst case and reference model equity premia are identical and the signal precision parameter k does not affect directly conditional equity premia.

C. Risky Asset Volatilities and their Relation to Equilibrium Equity Premia

From Proposition 2 the volatility of stock returns is given by

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

and under Assumption 5 it is higher in the presence of learning under ambiguity aversion than in the absence of an uncertainty about the ambiguous model neighborhood $\Xi(\theta)$. Moreover, σ_R^2 is an increasing function of the uncertainty about the true model neighborhood $\Xi(\theta)$ since mean preserving spread increase V_{θ,h^*} under Assumption 5.

Panel (C) of Figure 8 highlights the effect on equity volatilities for the heterogenous degrees of ambiguity studied in Figure 8. As in Veronesi (2000), volatilities are U-shaped functions of risk aversion, which attain a minimum at $\gamma = 1$. However, different structures $\eta(\theta)$ of ambiguity can imply higher or lower volatilities than in the pure Bayesian learning case. In particular, we observe that the asymmetric function $\eta(\theta)$ in Panel (A-2) yields the highest equity volatilities (the lines in Panel (C) marked by '$*$'). Hence, ambiguity structures, for which the ambiguity associated with "bad" economic states is higher, can produce even more excess volatility than the one implied by a pure setting of learning with no ambiguity.

From (42), (46) the relations between worst case and reference model equity premia and the conditional variance of returns are

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

and

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

respectively. In particular, when Π is non-degenerate several patterns for a time varying relation between μ_R and σ_R^2 arise. More precisely, equation (48) implies a *true positive but time varying relation* between $\mu_R - \mu_R^{wc}$ and σ_R^2 while a *true linear relation* between μ_R^{wc} and σ_R^2 is implied by (47). In (47) such a linear relation is biased by the time varying stochastic term $-\gamma V_{\theta,h^*}\left(1 + kV_{\theta,h^*}\right)$, which will tend to bias downwards estimates of γ based on a least squares regression of μ_R^{wc} on σ_R^2. Moreover, for realistic parameter choices in the model the actual premium μ_R will be dominated by the term $\mu_R - \mu_R^{wc}$, implying that there is no true and simple time invariant relation between equity premia and conditional variances in a setting of learning under ambiguity.

Figure 9 highlights the above observations by plotting the time series of estimated parameters in a sequence of rolling regressions of R on σ_R^2.

Insert Figure 9 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.

D. 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 in our setting have to be associated with EIS above one. However, in our model 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. Similarly 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.[8]

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

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

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

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

$$E\left(dD/D|\,\mathcal{F}_t\right) = a + b \cdot E\left(dP/P + D/P|\,\mathcal{F}_t\right) - V_\theta, \tag{49}$$

where $a = -\delta/\gamma + \frac{1}{2}(\gamma-1)\sigma_D^2$ and $b = 1/\gamma$. Equation (49) 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 5, 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. Figure 10 and 11 illustrate these features.

Insert Figures 10,11 about here

In Figures 10 and 11, 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.

V. Conclusions and Outlook

We derive asset prices in a simple continuous time partial information Lucas economy with ambiguity aversion and time additive power utility. In our model, learning and ambiguity aversion imply an equilibrium discount for ambiguity only for moderate relative risk aversions below one or, equivalently, elasticities of intertemporal substitution (EIS) above one. Equilibrium interest rates are lower irrespective of risk aversion. For low risk aversions, we observe higher conditional equity premia and volatilities relatively to (i) a comparable partial information, rational expectations, Lucas economy and (ii) a comparable full information Lucas economy with ambiguity aversion. Ambiguity aversion implies only a partial asymptotic learning about a neighborhood of a priori statistically indistinguishable beliefs. This result motivates explicitly settings of ambiguity aversion but no learning as the limit of equilibria under learning and ambiguity aversion. In a setting with low relative risk aversions below one and moderate ambiguity, we obtain asset pricing predictions consistent with the equity premium, the low interest rate, and the excess volatility puzzles. We also find that no time invariant relation between excess returns and conditional variances exists in equilibrium under learning and ambiguity aversion. This feature generates estimated relations between excess returns and equity conditional variances with an undetermined sign over time. Moreover, standard EIS estimates based on Euler equations for equity returns are strongly downward biased in a setting of learning and ambiguity aversion. Therefore, EIS well above one in the model are consistent with observed (biased) estimated EIS well below one.

We end with two final comments. First, the time additive power utility function in our model allows us to obtain simple closed form solutions for the desired equilibria at the cost of constraining the relation between risk aversion and EIS. The specific relation between standard risk aversion and EIS in our framework could be weakened by using a setting of learning under ambiguity aversion with Epstein and Zin (1989)-type preferences. 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. However, such an extension departs from the axiomatic setting of learning under ambiguity aversion in Knox (2004). Our setting of learning and ambiguity aversion can be interpreted as a continuous time extension of such an axiomatic framework and has therefore a clear behavioral interpretation. Moreover, it is intriguing to note that under ambiguity aversion relative risk aversion parameters too far away from risk neutrality may even be behaviorally inappropriate. Provocative in this context is the experimental evidence collected by Wakker and Deneffe (1996) who estimated a virtually linear utility function when using a utilities 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 were clearly concave. Nevertheless, the basic intuition derived from our model is likely to hold also under more general preferences that disentangle risk aversion and 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 lower expected consumption growth because of ambiguity aversion induces a large substitution from today's to tomorrow's consumption which smooths out consumption. 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, we can thus interpret the assumption of a high EIS and a low risk aversion in our model just as a condition ensuring that the elasticity of the total demand for risky assets with respect to changes in expected consumption growth is positive.

A second simplifying building block of our model of learning and ambiguity aversion is a geometric Brownian motion reference model dynamics for the underlying dividend process. This choice allowed us to highlight in a simple framework important issues related to (i) the non convergence of a Bayesian learning process under a likelihood misspecification and (ii) the asymptotic persistence of ambiguity under learning and ambiguity aversion. The bias of standard Bayesian learning procedures under a misspecification of the underlying dividend dynamics gave us a natural motivation for the conservative Max-Min expected utility learning approach followed in the paper. Richer learning dynamics could be studied, including for instance reference models with business cycles and regime changes in the underlying dividend process. Such extensions are interesting venues for future research on learning under ambiguity aversion.

VI. Appendix

In this Appendix we provide all proofs to the propositions in the paper.

Proof of Lemma 1. The Lemma is a direct consequence of Lemma 1 in Veronesi (2000), when applied to the likelihood $h(\theta)$. ■

Proof of Corollary 1. The statement of the corollary follows by noting that under a h_D-distorted dynamics it follows

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

■

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

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

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

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

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

$$dD = (\theta_i + h(\theta_i)\sigma_D) D dt + \sigma_D D d\widetilde{B}_D \quad .$$

Therefore, Assumption 3 implies that we can focus on solving the problem

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

subject to the dividend dynamics

$$dD = (\theta_i + h(\theta_i)\sigma_D) D dt + \sigma_D D d\widetilde{B}_D^h \quad .$$

The Hamilton Jacobi Bellman (HJB) equation for this problem reads

$$0 = \inf_{h(\theta_i)} \left\{ -\delta V^i + \frac{D^{1-\gamma}}{1-\gamma} + (\theta_i + h(\theta_i)\sigma_D) D \cdot V_D^i + \frac{1}{2}\sigma_D^2 D^2 V_{DD}^i + \lambda \left(\frac{1}{2} h(\theta_i)^2 - \eta(\theta_i) \right) \right\} \quad , \tag{50}$$

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

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

Slackness then gives

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

implying

$$h(\theta_i) = \pm\sqrt{2\eta(\theta_i)} \quad .$$

Inserting back this expression in (50) we get two possible differential equations for V^i:

$$(i) \quad : \quad 0 = -\delta V^i + \frac{D^{1-\gamma}}{1-\gamma} + \left(\theta_i + \sqrt{2\eta\left(\theta_i\right)}\sigma_D\right) D \cdot V_D^i + \frac{1}{2}\sigma_D^2 D^2 V_{DD}^i \quad .$$

$$(ii) \quad : \quad 0 = -\delta V^i + \frac{D^{1-\gamma}}{1-\gamma} + \left(\theta_i - \sqrt{2\eta\left(\theta_i\right)}\sigma_D\right) D \cdot V_D^i + \frac{1}{2}\sigma_D^2 D^2 V_{DD}^i \quad .$$

However, it is clear that the infimum in (50) can be only characterized by equation (ii). Thus, the worst case drift solution is characterized by

$$h^*\left(\theta_i\right) = -\sqrt{2\eta\left(\theta_i\right)} \quad .$$

This proves the first statement. To prove the second statements, we first note that, conditional on $\widetilde{\theta} = \theta_i - \sqrt{2\eta\left(\theta_i\right)}\sigma_D$, we have

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

Then, since $C_i = (1-\gamma)V_i/D(t)^{1-\gamma}$, we get

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

where

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

This proves the second statement. To prove the last two statements notice that given the set of worst case drift misspecifications, we obtain for the price of any risky asset with dividend process $\left(D\left(t\right)\right)_{t\geq 0}$

$$\frac{P\left(t\right)}{D\left(t\right)} = \sum_{i=1}^n \pi_i E\left[\int_t^\infty e^{-\delta(s-t)}\left(\frac{D\left(s\right)}{D\left(t\right)}\right)^{1-\gamma} ds \middle| \widetilde{\theta} = \theta_i - \sqrt{2\eta\left(\theta_i\right)}\sigma_D\right] \quad , \qquad (51)$$

or equivalently

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

where $\rho(t) = u_c(D(t), t) = e^{-\delta t} D(t)^{-\gamma}$. Writing this equation in differential form and applying it to the risky asset paying a "dividend" $D = r$ we obtain:

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

■

Proof of Corollary 5. The Corollary is obtained by setting $h_D = 0$, $h(\theta_i) = -\sqrt{2\eta(\theta_i)}$ and $m_{h(\theta)} = -\sum_{i=1}^{n} \pi_i \sqrt{2\eta(\theta_i)}$ in Corollary 1. ■

Proof of Proposition 2. The result follows by applying the proof of Proposition 2 in Veronesi (2000) to the $D-$dynamics (23) with the worst case term $m_{h^*(\theta)} = -\sum_{i=1}^{n} \pi_i \sqrt{2\eta(\theta_i)}$ and by expressing them under the reference model $\Pi-$dynamics (13) with the Brownian motion $\widetilde{B}_D$ and $\widetilde{B}_e$. ■

Proof of Lemma 2. 1. By Assumption 3 we have

$$\theta_i - \sqrt{2\eta(\theta_i)}\sigma_D > \theta_j - \sqrt{2\eta(\theta_j)}\sigma_D \quad , \tag{52}$$

for any $i > j$. Therefore, for $\gamma < 1$ the sequence $\{C_i\}_{i=1,..n}$ is monotonically strictly increasing in i (see (28) for the definition of C_i). Hence, for fixed Π the sequence $\{C_i / \sum_{i=1}^{n} \pi_i C_i\}_{i=1,..n}$ is monotonically strictly increasing in i. In particular, this implies

$$C_1 / \sum_{i=1}^{n} \pi_i C_i < 1 < C_n / \sum_{i=1}^{n} \pi_i C_i \quad .$$

Since the sequences $\{\theta_i\}_{i=1,..,n}$ and $\left\{\widetilde{\theta}_i\right\}_{i=1,..,n}$, where $\widetilde{\theta}_i := \theta_i - \sqrt{2\eta(\theta_i)}\sigma_D$, are increasing in i this implies

$$\overline{m}_\theta = \sum_{i=1}^{n} \pi_i \frac{C_i \theta_i}{\sum_{i=1}^{n} \pi_i C_i} > \sum_{i=1}^{n} \pi_i \theta_i = m_\theta \quad .$$

Similarly,

$$\overline{m}_\theta + \overline{m}_{h^*(\theta)} = \sum_{i=1}^{n} \pi_i \frac{C_i \widetilde{\theta}_i}{\sum_{i=1}^{n} \pi_i C_i} > \sum_{i=1}^{n} \pi_i \widetilde{\theta}_i = m_\theta + m_{h^*(\theta)} \quad ,$$

implying $V_{\theta.h^*} > 0$. The proof for $\gamma = 1$ and $\gamma > 1$ follows in the same way. The proof of 2. and 3. follows with the same arguments as in the proof of Lemma 2 (b) and (c) in Veronesi (2000). ■

References

Aït-Sahalia, Y., J. A. Parker, and M. Yogo, 2004. Luxury Goods and the Equity Premium, Journal of Finance, forthcoming.

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

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

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 G. Weber, 1989. Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption, Economic Journal, 99, 59-73.

Bansal, R. and A. Yaron, 2004. Risks for the Long Run: a Potential Resolution of Asset Pricing Puzzles, Journal of Finance, forthcoming.

Barberis, N., 2000. Investing in 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. Assessing Asset Pricing Anomalies, Review of Financial Studies, 14, 905-942.

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

Campbell, J. Y., 1999. Asset Prices, Consumption and the Business Cycle, in John B. Taylor and Michael Woodford, eds.: Handbook of Macroeconomics, Volume 1 (Elsevier Science, North Holland).

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

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

Epstein, L. 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, 2002. Learning under Ambiguity, Manuscript, University of Rochester.

Epstein, L., G. and M. Schneider, 2003. Recursive Multiple Priors, Journal of Economic Theory, 113, 32-50.

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

Epstein, L., and S. E. Zin, 1989. Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: a Theoretical Framework, Econometrica, 57, 937-969.

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 Lugano.

Gagliardini P., F. Trojani and G. Urga, 2004. Robust GMM Tests for Structural Breaks, Journal of Econometrics, forthcoming.

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, Jr., 1999. Robust Permanent Income and Pricing, Review of Economic Studies, 66, 873-907.

Hansen, L.P., T. J. Sargent, G. A. Turmuhambetova and N. Williams, 2001. Robustness and Uncertainty Aversion, Preprint University of Chicago.

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

Huber, P., 1981. Robust statistics, Wiley, New York.

Kandel, S., and R. F. Stambaugh, 1996. On the Predictability of Stock Returns: an Asset Allocation Perspective, Journal of Finance, 51, 385-424.

Knox, T., 2004. Foundations for Learning How to Invest when Returns are Uncertain, Manuscript, Journal of Economic Theory, forthcoming.

Krishnakumar, J. and E. Ronchetti, 1997. Krishnakumar, J. and E. Ronchetti, 1997. Robust Estimators for Simultaneous Equations Models, Journal of Econometrics, 78, 295-314.

Liu, J., J. Pan and T. Wang, 2004. An Equilibrium Model of Rare Events Premia, Review of Financial Studies, forthcoming.

Lucas, R. E. Jr., 1978. Asset Prices in an Exchange Economy, Econometrica 46, 1429-1446.

Maenhout, P., 2004. Robust Portfolio Rules and Asset Pricing, Review of Financial Studies, forthcoming.

Mancini, L., E. Ronchetti and F. Trojani, 2003. Optimal Conditionally Unbiased Bounded-Influence Inference in Dynamic Location and Scale Models, Journal of the American Statistical Association, forthcoming.

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

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

Ortelli, C. and F. Trojani, 2004. Robust Efficient Method of Moments, Journal of Econometrics, forthcoming.

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.

Ronchetti, E. and F. Trojani, 2001. Robust Inference with GMM estimators, Journal of Econometrics, 101, 37-69.

Routledge, B. R. and S. E. Zin, 2001. Model Uncertainty and Liquidity, NBER Working paper number w8683.

Sakata, S. and H. White, 1998. High Breakdown Point Conditional Dispersion Estimation with Application to S&P 500 Daily Returns Volatility, Econometrica 66, 529-567.

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

Shepard, N. G. and A. C. Harvey, 1990. On the Probability of Estimating a Deterministic Component in the Local Level Model, Journal of Time Series Analysis, 11, 339-347.

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, 26/3, p. 423-435.

Trojani, F. and P. Vanini, 2004. Robustness and Ambiguity Aversion in General Equilibrium, Review of Finance, 2, 279-324.

Uppal, R. and T. Wang, 2003. Model Misspecification and Under-Diversification, Journal of Finance, 58 (6), 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.

Veronesi, P., 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 Dynamics Asset Allocation, Journal of Finance, 56, 205-246.

Figures

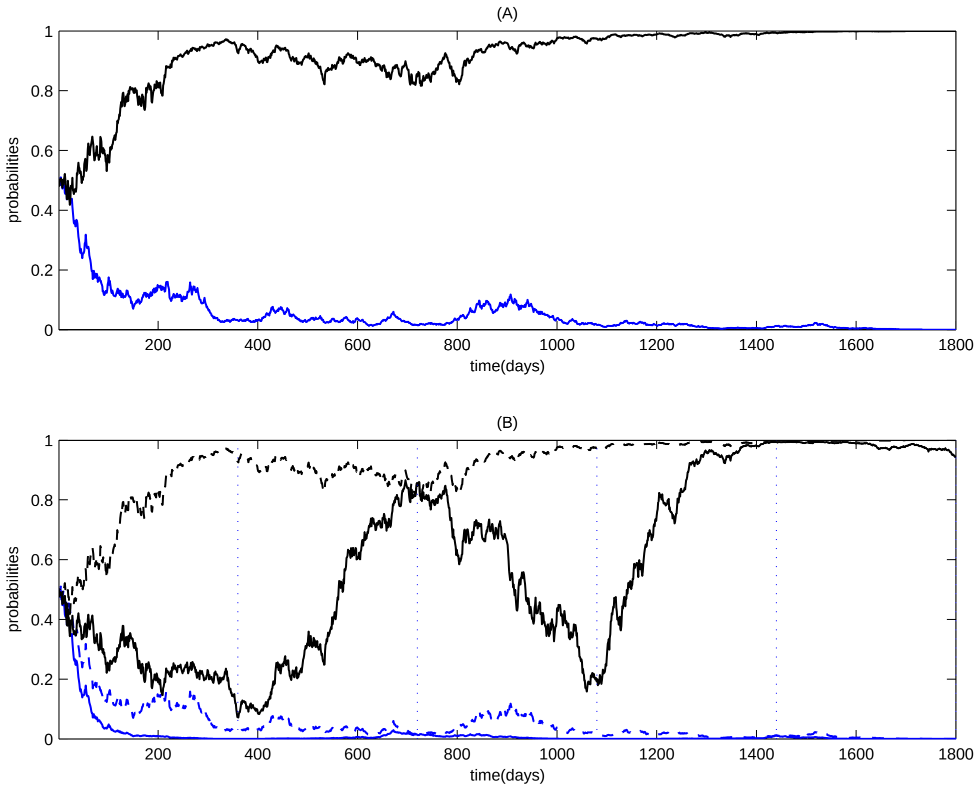

Figure 1. Posterior probabilities dynamics. The panels display trajectories for the probability π_1 given in equation (16) of Example 8. Panel (A) shows two trajectories for π_1 with $a = 0$ in equation (17). We plot the same trajectories with the same random seed in Panel (B) (dashed lines) and add the trajectories (solid lines) where we assume $a = 0.015$. The switching in equation (17) is deterministic and occurs every year (see the dotted vertical lines in Panel (B)). The further parameters are $\Theta = \{0.0075, 0.0275\}$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$.

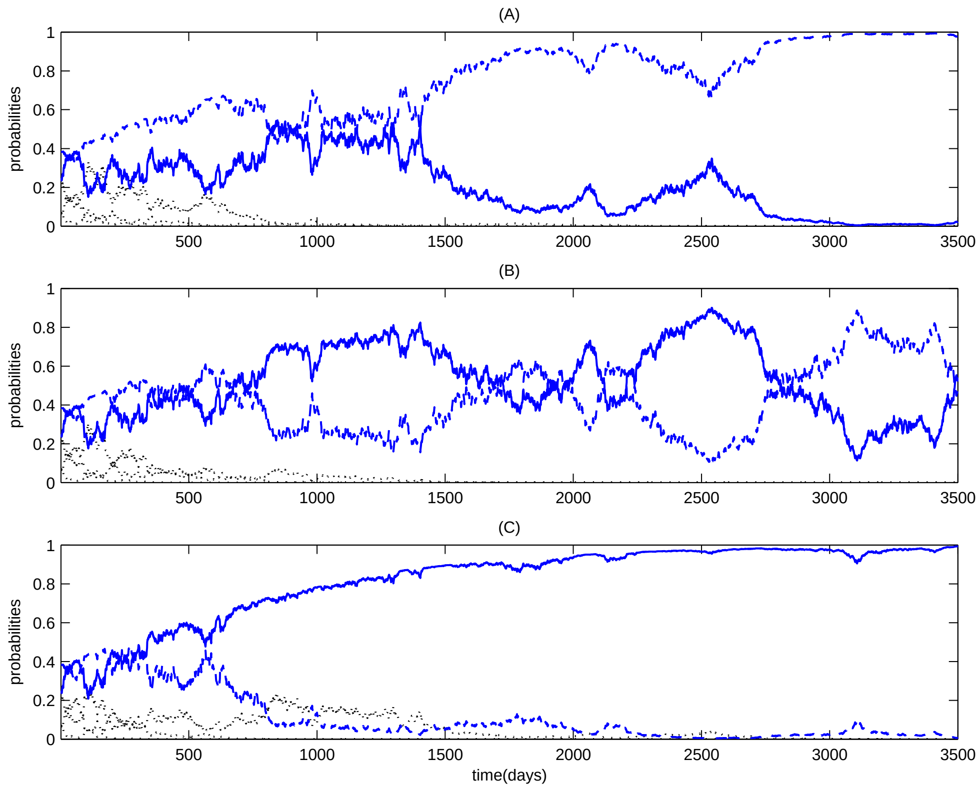

Figure 2. Posterior probabilities dynamics. Panels (A), (B), (C) plot for different random seeds the dynamics of the implied posterior probabilities when the learning model allows for five possible dividend drift states and the underlying dividend drift process is geometric Brownian motion with drift θ_2. In Panel (A), we assume that there is no likelihood misspecification, i.e., $h(\theta_i) = 0$. In Panel (B), we introduce a simple (constant) misspecification and set $h(\theta_i) = 0.141$, $i = 1, .., n$. In Panel (C), we increase the size of such misspecification to $h(\theta_i) = 0.316$, $i = 1, .., n$. It turns out that the learning dynamics in Panels (A), (B), and (C) differ substantially. In Panel (A), state θ_2 will eventually be correctly learned (see the dashed line in the graph). In Panel (B), the posterior probabilities does not converge after 3600 days. In Panel (C), state θ_3 will be eventually "learned" (see the solid line in the graph). Model parameters are $\Theta = \{-0.0125, 0.0025, 0.0175, 0.0325, 0.0475\}$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$, $h_D = 0$.

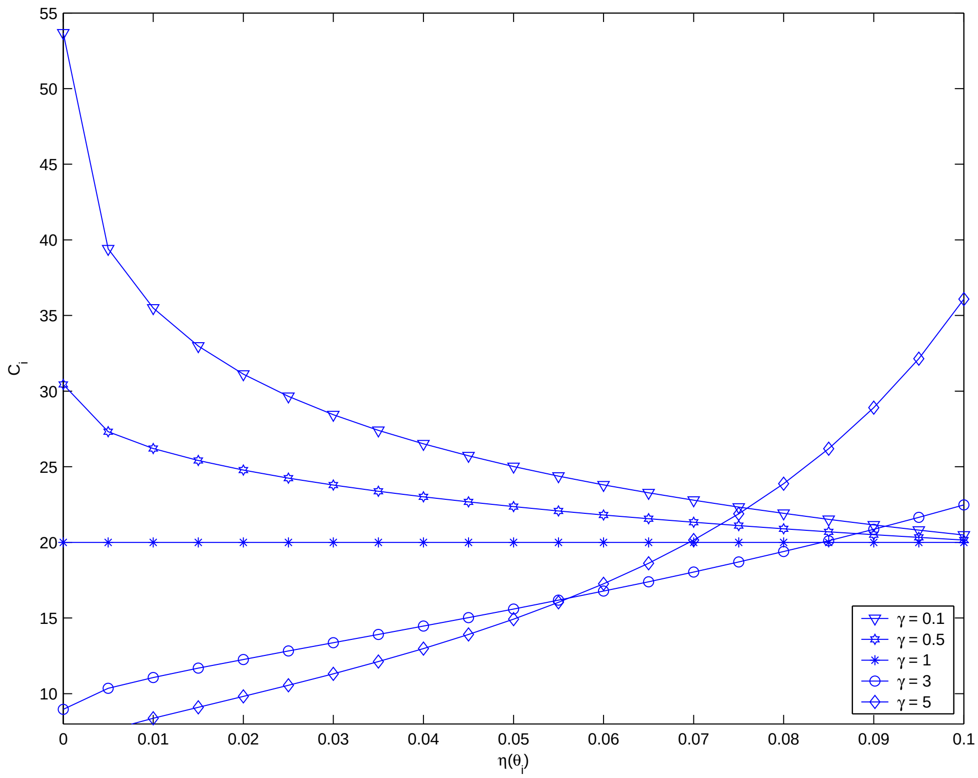

Figure 3. Function C_i in dependence of the ambiguity and risk aversion parameters $\eta_i(\theta_i)$ and γ. We assume $\delta = 0.025$, $\theta_i = 0.0175$, $\sigma_D = 0.0375$, and calculate C_i for different amounts of ambiguity, $\eta_i(\theta_i) \in [0, 0.1]$, and different degrees of risk aversion, $\gamma = \{0.1, 0.5, 1, 3, 5\}$.

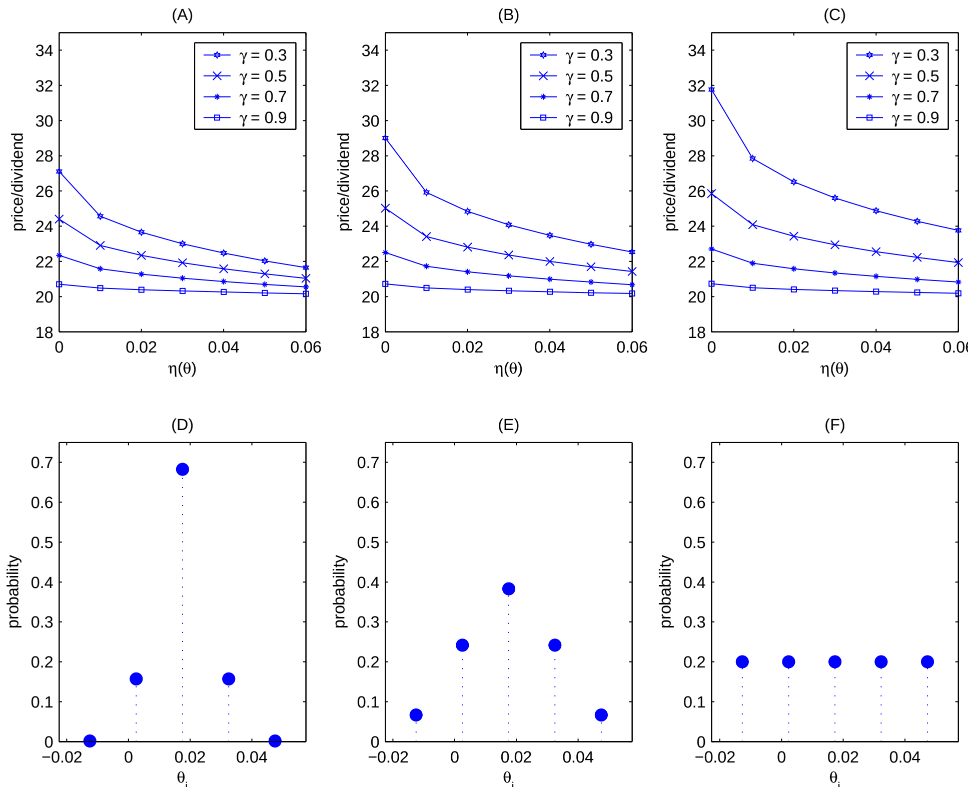

Figure 4. The impact of uncertainty and ambiguity aversion on price dividend ratios. In Panels (A)-(C) we explore the impact of mean preserving spreads on price dividend ratios. We display the prior distribution corresponding to each of those panels in the Panels (D)-(F). Given such prior distributions, we assume a setting with five possible states $\Theta = \{-0.0125, 0.0025, 0.0175, 0.0325, 0.0475\}$. We further set $\delta = 0.05$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$.

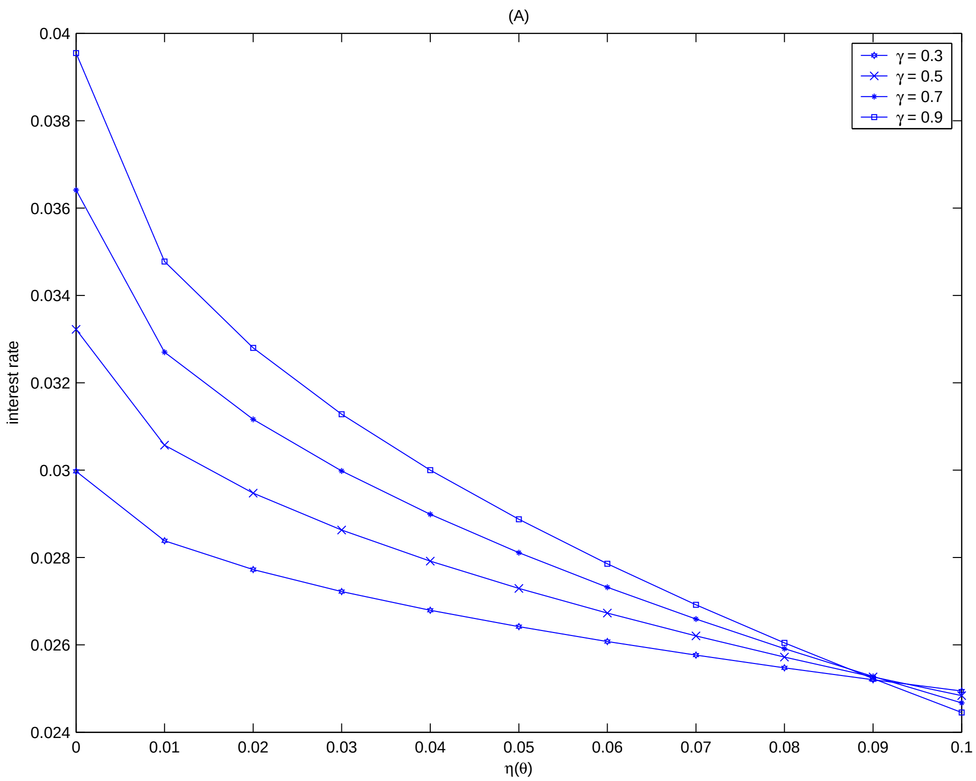

Figure 5. The equilibrium interest rate. We plot the equilibrium interest rate as displayed in Proposition 9 for different homogenous degrees of ambiguity $\eta(\theta)$ and different risk aversions $\gamma = \{0.3, 0.5, 0.7, 0.9\}$. We assume five states $\Theta = \{-0.0125, 0.0025, 0.0175, 0.0325, 0.0475\}$ and set $\delta = 0.025$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$ with a constellation of discretized normal priors $\Pi = \{0.0668, 0.2417, 0.3829, 0.2417, 0.0668\}$.

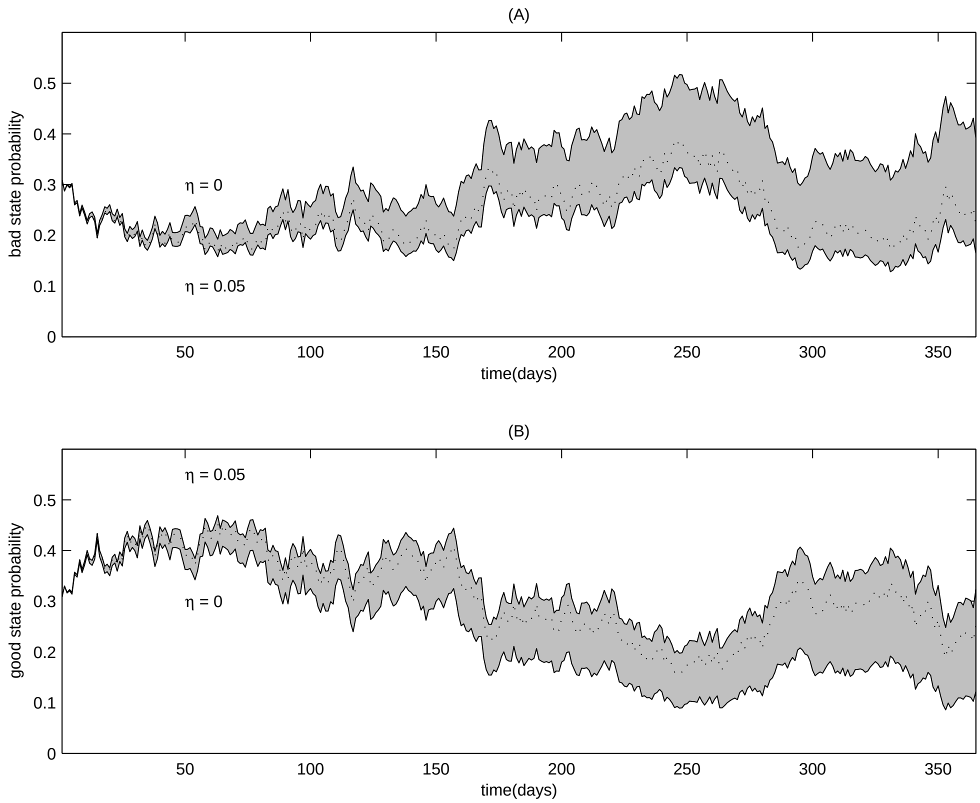

Figure 6. The effect of ambiguity aversion on the prevailing posterior probabilities dynamics. We assume three possible states and the filtered probabilities dynamics in equation (38) with parameters set equal to $\sigma_D = 0.0375$, $\sigma_e = 0.015$, $\Theta = \{0.0023, 0.0173, 0.0323\}$, $\theta = 0.0173$, 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 homogenous ambiguity parameter $\eta = \{0, 0.025, 0.05\}$. The dynamics under the intermediate level of ambiguity $\eta = 0.025$ are represented by the dotted line. 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)).

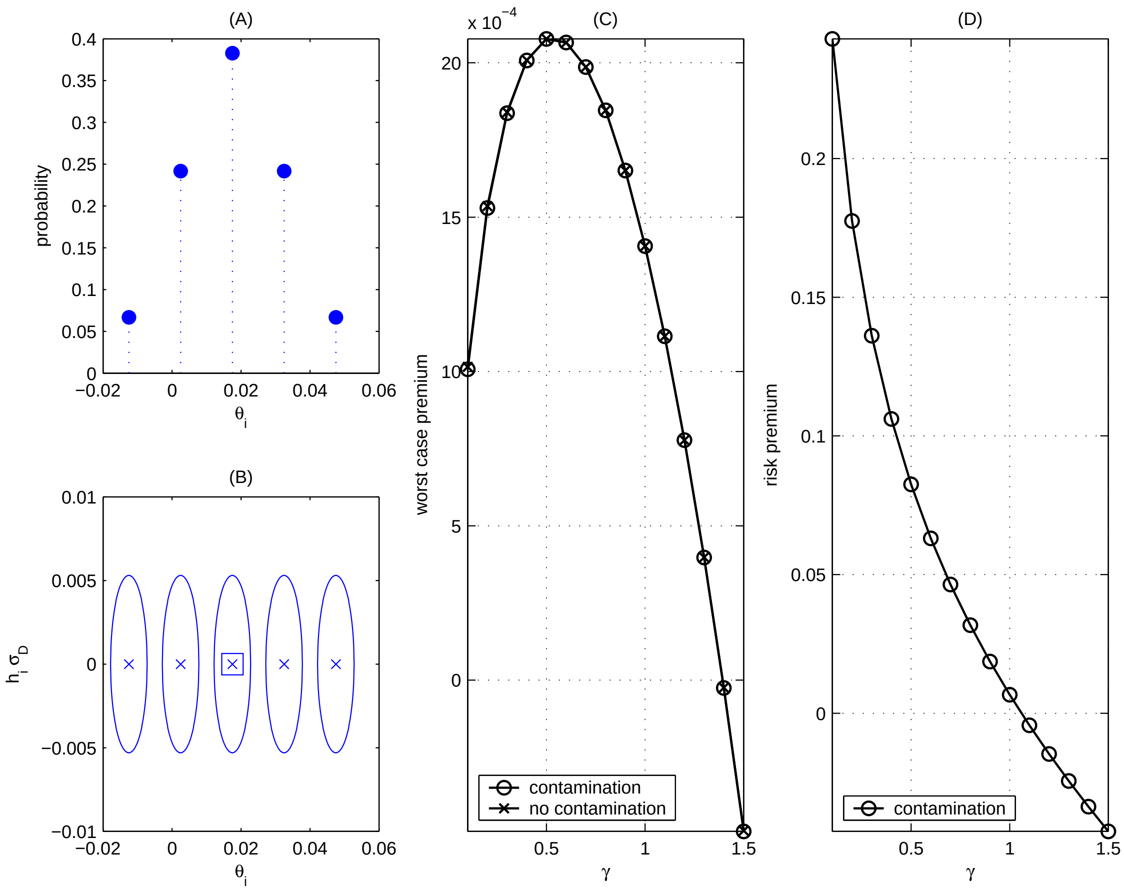

Figure 7. Panel (A) plots the set of probabilities Π relevant for the figure while Panel (B) plots the different relevant reference model states $\theta_1, .., \theta_n$. The true reference model dividend drift state is marked with a square and has been set equal to the posterior expected value $\sum \pi_i \theta_i$. We use a small amount of homogenous ambiguity $\eta = 0.01$. 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 the resulting worst case equity premium μ_R^{wc} and the actual equity premium μ_R are plotted in Panel (C) and (D) as functions of γ.

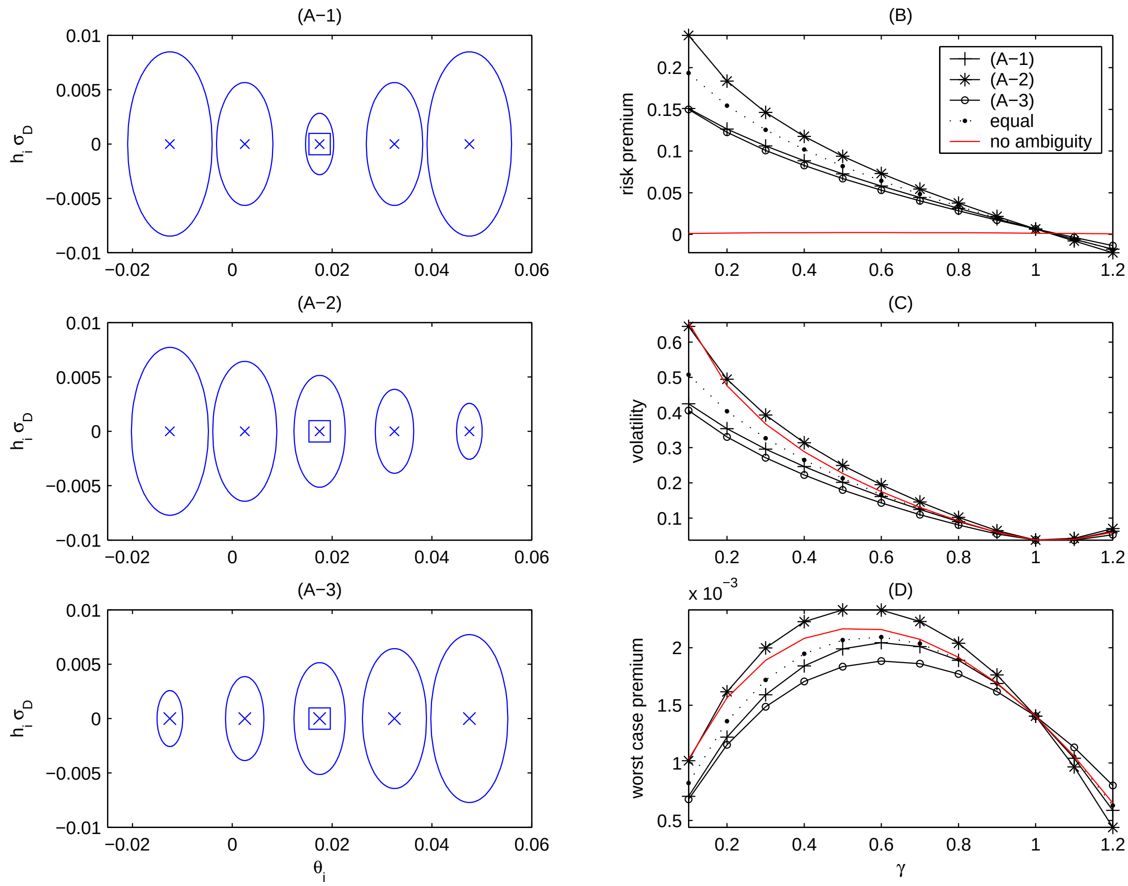

Figure 8. Panels (A-1)–(A-3) plot different entropy preserving distributions of ambiguity around the reference model dividend drift states $\theta_1, .., \theta_5$, i.e., Panels (A-1)–(A-3) are such that the weighted entropy measure $\frac{1}{2}\sum \pi_i h^2(\theta_i)$ is equal to 0.01, as in Figure 7. For Panel (A-1), we set $\eta(\theta_i)_{i=1,..,5} = \{0.0256, 0.0114, 0.0028, 0.0114, 0.0256\}$, for Panel (A-2), $\eta(\theta_i)_{i=1,..,5} = \{0.0212, 0.0147, 0.0094, 0.0053, 0.0024\}$ and for Panel (A-3), $\eta(\theta_i)_{i=1,..,5} = \{0.0024, 0.0053, 0.0094, 0.0147, 0.0212\}$. The set $\Pi = \{0.0668, 0.2417, 0.3829, 0.2417, 0.0668\}$ of prior distributions and the set $\Theta = \{-0.0125, 0.0025, 0.0175, 0.0325, 0.0475\}$ of reference model drifts relevant for all graphs are the same as those in Panel (A) of Figure 7. Panels (B), (C), and (D) plot, the equity premium μ_R, the volatility σ_R and the worst case premium μ_R^{wc} implied by the different distributions of ambiguity in Panels (A-1)–(A-3). For comparison, we plot in each graph the corresponding quantities prevailing under an homogenous (equal) ambiguity parameter $\eta = 0.01$ and in the absence of ambiguity ($\eta(\theta_i) = 0$, $i = 1, .., n$). Further parameters are $\delta = 0.05$, $\sigma_D = 0.0375$ and $\sigma_e = 0.015$.

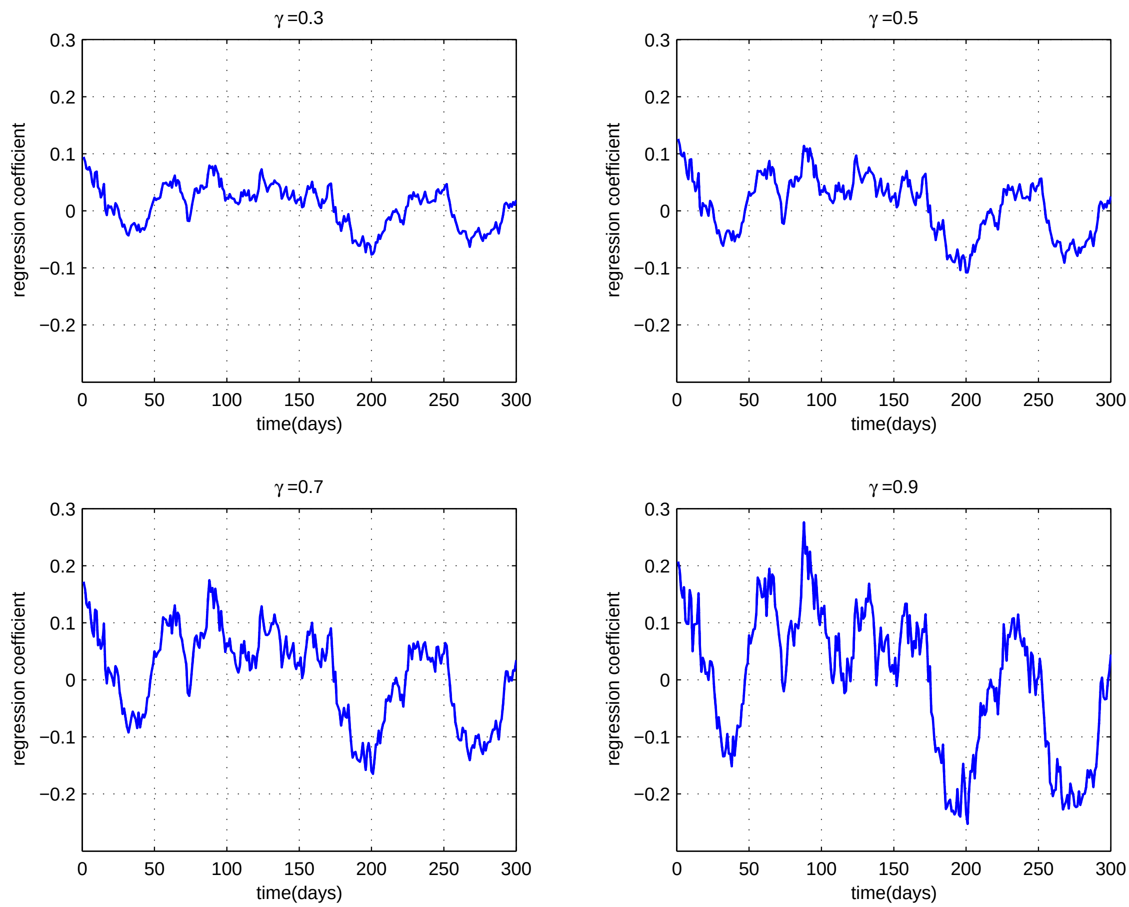

Figure 9. Regression analysis. For different parameters $\gamma = 0.3, 0.5, 0.7, 0.9$ we plot the time variation of the estimated parameter b in (49) for a rolling regression of R on σ_R^2. The rolling regressions are based on sample sizes of 50 observations simulated from a model with three reference model drift states $\Theta = \{0.0025, 0.0175, 0.0325\}$ and under an homogenous degree of ambiguity $\eta = 0.01$. The true dividend is $\theta = 0.0175$. Further parameters are $\delta = 0.05$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$.

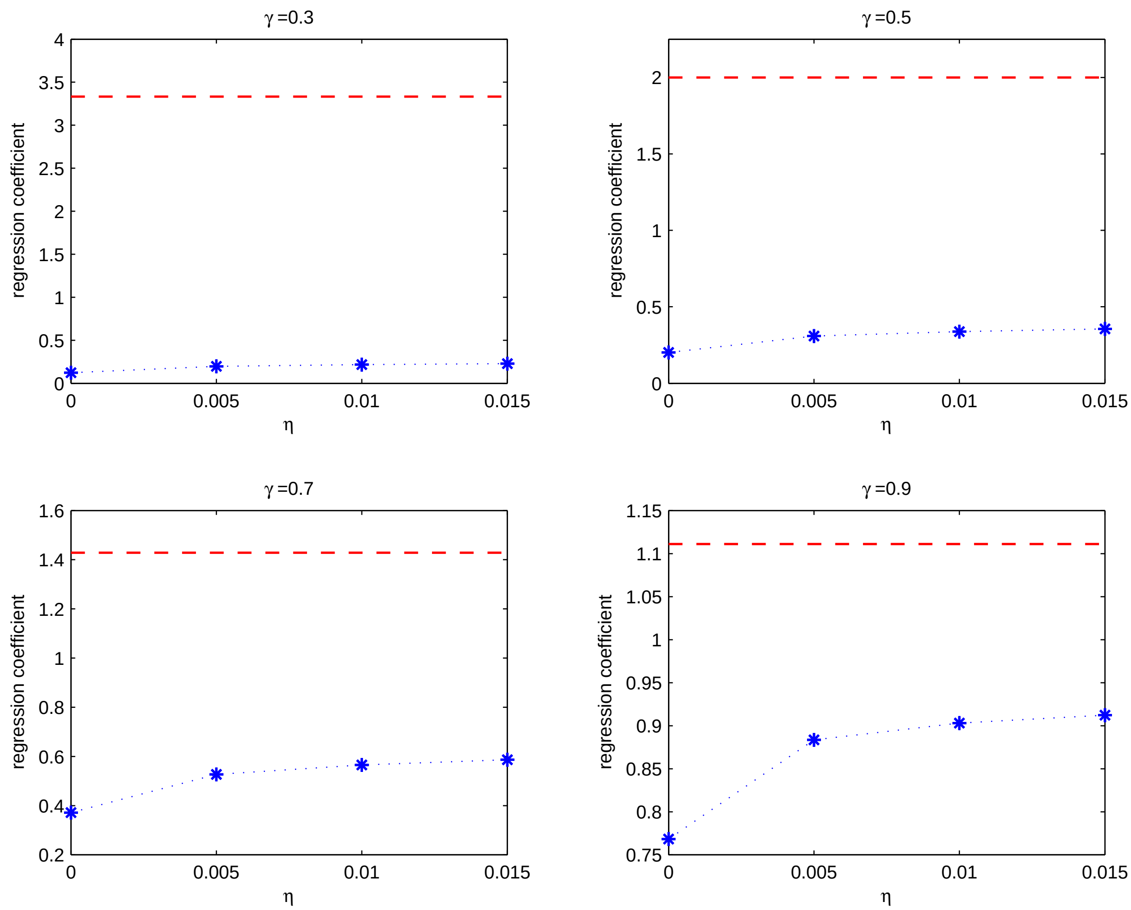

Figure 10. Regression analysis. For different parameters $\gamma = 0.3, 0.5, 0.7, 0.9$, we plot the mean estimated parameter in 1000 regressions of dD/D on $dP/P + P/D$. The regressions are based on sample sizes of 365 observations simulated from a model with three reference model drift states $\Theta = \{0.0025, 0.0175, 0.0325\}$ and under three homogenous 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.

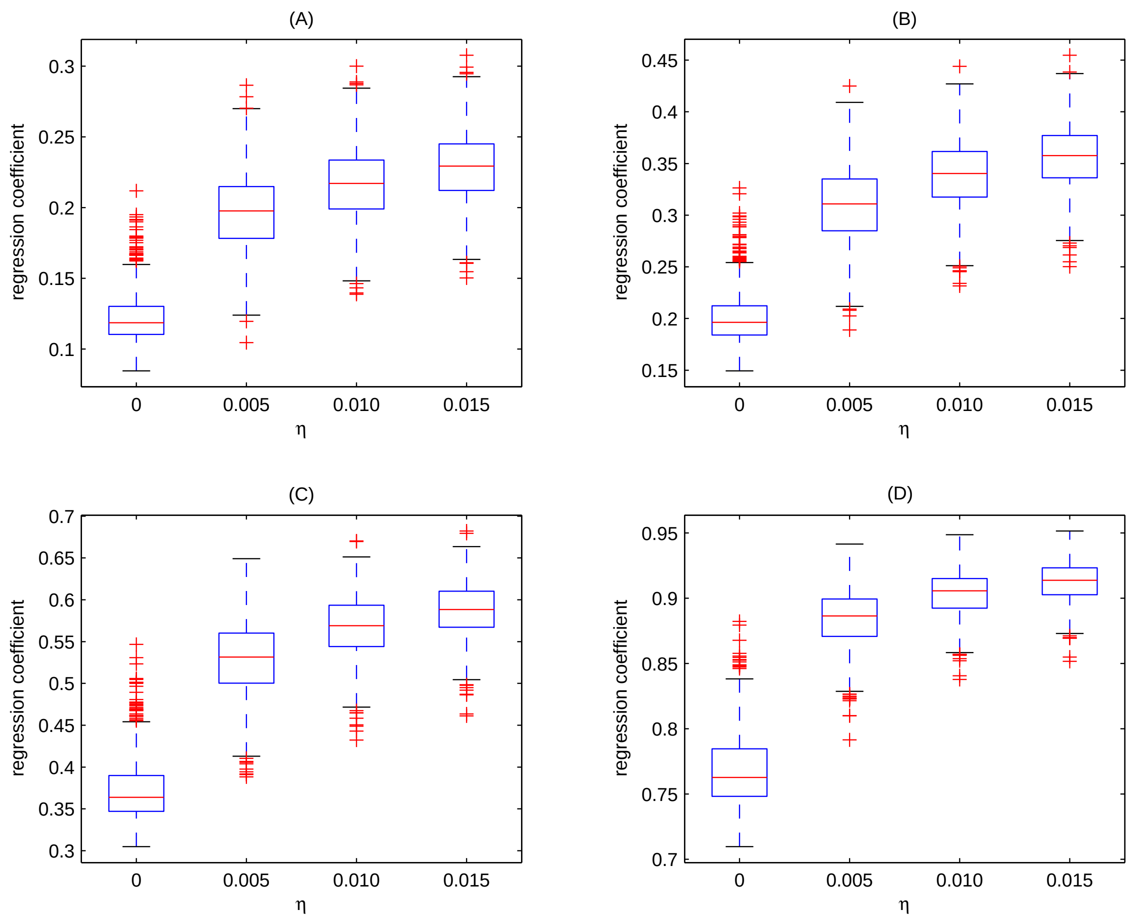

Figure 11. Regression analysis. For different parameters $\gamma = 0.3, 0.5, 0.7, 0.9$, we present the box plots of the estimated parameters in 1000 regressions of dD/D on $dP/P + P/D$. The regressions are based on sample sizes of 365 observations simulated from a model with three reference model drift states $\Theta = \{0.0025, 0.0175, 0.0325\}$ and under three homogenous degrees of ambiguity $\eta = 0, 0.005, 0.01, 0.015$. The true dividend is $\theta = 0.0175$. Further parameters are $\delta = 0.05$, $\sigma_D = 0.0375$, $\sigma_e = 0.015$.