



**DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES**

AMBIGUITY AND THE HISTORICAL EQUITY PREMIUM

Fabrice Collard, Sujoy Mukerji, Kevin Sheppard and Jean-Marc Tallon

Number 550
April 2011

Manor Road Building, Oxford OX1 3UQ

Ambiguity and the historical equity premium*

Fabrice Collard

Department of Economics, University of Bern
Schanzenneckstrasse 1, 3001 Bern

Sujoy Mukerji[†]

Department of Economics and University College
University of Oxford, Oxford OX1 3UQ

Kevin Sheppard

Department of Economics and Oxford-Man Institute of Quantitative Finance
University of Oxford, Oxford OX1 3UQ

Jean-Marc Tallon

Paris School of Economics, Université Paris 1 and CNRS
106-112 boulevard de l'Hôpital, 75013 Paris

May 9, 2011

Abstract

This paper assesses the quantitative impact of ambiguity on the historically observed equity premium. We consider a Lucas-tree pure-exchange economy with a single agent where we introduce two key non-standard assumptions. First, the agent's beliefs about the dividend/consumption process is ambiguous, i.e., she is uncertain about the exact probability distribution governing the realization of future dividends and consumption. Second, the agent's preferences are sensitive to this ambiguity, a property formalized using the smooth ambiguity model. The consumption and dividend process is assumed to evolve according to a hidden state model, popularized by Bansal and Yaron (2004), where a persistent latent state variable describes *temporary* shocks to the mean of consumption growth prospects. We further extend the model to allow for uncertainty about the magnitude of the persistence of the latent state. The agent's beliefs are ambiguous due to the uncertainty about the conditional mean of the probability distribution on consumption and dividends in the next period. We show that in this model ambiguity is *endogenously* dynamic, for example, increasing during recessions. This results in an endogenously volatile and (counter-)cyclical equity premium. We calibrate the level of ambiguity aversion to match *only* the first moment of the risk-free rate in data, and ambiguity to match the uncertainty conditional on the historical growth path, and evaluate the model using moderate levels of risk aversion. We find that this simple modification of a Lucas-tree model accounts for a large part of the historical equity premium, both in terms of its level and variation over time.

J.E.L. Codes: G12, E21, D81, C63

Keywords: Ambiguity Aversion, Asset pricing, Equity premium puzzle

*We thank Ravi Bansal, Paul Beaudry, Tim Cogley, Hui Chen, Hippolyte d'Albis, Vito Gala, Christian Gollier, Lars Hansen, Peter Klibanoff, Tarun Ramadorai and Raman Uppal for helpful discussions. We also thank seminar and conference participants at Adam Smith Asset Pricing conference, AEA, RUD (Heidelberg), Northwestern (MEDS), Warwick, Leicester, Transatlantic Theory Workshop (2010), EUI (Florence).

[†]Contact author: sujoy.mukerji@economics.ox.ac.uk

1 Introduction

This paper seeks to assess the *quantitative* impact of ambiguity about macroeconomic risk on financial asset prices and returns, in particular the equity premium. Ambiguity refers to uncertainty about the “true” probability distribution governing future outcomes. The decision maker’s ambiguity attitude determines how and to what extent that uncertainty affects her choices (e.g., whether she is averse to such uncertainty and if so, the level of aversion). Notions of ambiguity and its possible relevance to economics were discussed intuitively by Knight (1921) and Ellsberg (1961), and decision theoretic formulations by Schmeidler (1989) and Gilboa and Schmeidler (1989) presented a first set of tools to incorporate these ideas into formal economic analysis. Inspection and experimental evidence, typified by the Ellsberg examples, suggest that agents commonly adjust their behavior in response to such uncertainty (see, e.g., Camerer and Weber (1992)). In the economics literature, agents are typically posited as averse to ambiguity and inclined to choose actions whose consequences are more robust to the perceived ambiguity, e.g., a portfolio position whose (ex ante) value is relatively less affected by the uncertainty about probability distribution governing the future payoffs.¹ An important reason why ambiguity is pervasive in economic decision making environments is model uncertainty; robustness concerns in the face of such uncertainty may give rise to ambiguity averse behavior. For example, a typical professional investor may have different forecasting models for the same variable, or different parameter estimates for the same model, all of which are plausible on the basis of historical data. If the models make distinctly different forecasts about key variables of interest, it is natural to seek a portfolio that is robust across the range of forecasts rather than optimizing exclusively to the forecast from a single model as argued, e.g., in Hansen (2007). The formal model of ambiguity averse preferences we apply in this paper articulates one precise sense in which a decision maker may express her concern over robustness.

This paper considers a single agent, Lucas-tree, pure-exchange economy with two innovations. First, the agent’s belief about the consumption and dividend process is ambiguous, i.e., in each period she is uncertain about the exact probability distribution governing the realization of consumption and dividends in the following period; the uncertainty is not static but dynamic, evolving, as the agent learns from history. Second, the agent’s preferences are sensitive to this ambiguity. What makes the agent’s belief ambiguous is her conditional uncertainty about the *mean* of the probability distribution on dividends next period. This mean parameter is not directly observable, but inference can be made on its current value. However, it can never be fully pinned down through inference because it is ever evolving, determined in part by *temporary* shocks (whose magnitude and sign depend on where the economy is in business cycle terms). Inference is further complicated by the fact that the agent is also uncertain about the magnitude of the parameter determining the persistence of the temporary shock. This reflects the difficulty in determining, through observation, whether the true growth

¹See Dow and Werlang (1992), Epstein and Wang (1994), Mukerji and Tallon (2001), Caballero and Krishnamurthy (2008), Chen, Ju, and Miao (2009), Gollier (2009), Boyle, Garlappi, Uppal, and Wang (2010), Hansen and Sargent (2010), Maccheroni, Marinacci, and Ruffino (2010) and Uhlig (2010), *inter alia*.

process is highly persistent with the persistent component having a small variance or, moderately persistent with greater variance of the persistent component. Learning ensures, following a period of stable growth, uncertainty about the conditional mean diminishes since forecasts based on alternative conjectures about the persistence will (endogenously) tend to agree. However, in the aftermath of a significant shock alternative formulations may disagree considerably about growth prospects, causing uncertainty about the conditional mean to increase temporarily.² Moreover, because of uncertainty about persistence, given a strong negative shock, its worst effects will be expected to last longer than they would be if the shock were positive. As a consequence, the extent of uncertainty and anxiety perceived by an ambiguity averse agent is different depending on the sign of the shock. This quantitative model of ambiguity about macroeconomic risk, where the ambiguity waxes and wanes *endogenously*, as a function of the publicly observed history of aggregate consumption and dividend, is a key part of the paper and underpins its measurement of the impact of ambiguity on asset returns.³

Formally, we endow the agent with smooth ambiguity preferences (Klibanoff, Marinacci, and Mukerji (2005, 2009), henceforth KMM2005, KMM2009). In this preference representation ambiguity is modeled as a subjective second order probability over first order probability distributions on (payoff determining) contingencies deemed possible by the decision maker. The representation functional is an expectation of expectations. The inner expectations evaluate the expected utilities corresponding to first order probabilities deemed subjectively possible while the outer expectation aggregates a *transformation* of these expected utilities with respect to the second order prior. The transformation of the expected utilities is to capture the agent’s ambiguity attitude; in particular, if the transformation is concave then the agent is ambiguity averse. Moreover, when the transformation is affine then the agent is ambiguity neutral, a subjective expected utility maximizer.

We calibrate the agent’s belief to the maintained assumption that the economy evolves according to (a modified version of) the hidden state model analyzed in Bansal and Yaron (2004). In Bansal and Yaron’s model, a hidden (latent) state variable describes the evolving economic potential of the economy by determining the extent of the temporary departure of the mean of the consumption and growth distribution from the trend. If the latent state were known to the agent, then the mean rate of growth of both consumption and dividends would be known. Since it is assumed that the growth distribution is Gaussian with a given (time invariant) volatility parameter, knowing the mean would be enough to completely characterize the distribution. The hidden state is not constant over time and evolves according to an *AR(1)* process. The agent, starting from a prior, updates beliefs about the latent state using Bayes rule. The updated belief on the date t hidden state is what, essentially, constitutes the ambiguity faced by the representative agent at date t . This kind of uncertainty about the data generating process is an example of “statistical ambiguity”, a term coined in Hansen (2007); here, it is the

²Two kinds of uncertainties follow an adverse shock, expressed by the following questions: “Are we in a recession? If so, how long will it last?” Questions, that certainly would have had some resonance in recent times.

³That learning may actually increase ambiguity is not a novel observation; see e.g., Epstein and Schneider (2008). However, in the present paper a signal does not cause ambiguity to increase because it is (exogenously) assumed to be of dubious quality. The model of beliefs here includes a theory that shows how the news of a growth outcome may or may not increase uncertainty depending on the run of history it follows.

uncertainty (at date t) about the probability distribution over future growth outcomes given the best statistical (rational, Bayesian) inference from (publicly available data) history of observations on growth outcomes up until time t . In our main model, dubbed the “two- ρ ” model, the value of the persistence parameter of the $AR(1)$ process describing the evolution of the latent state is not completely known by the agent: she is free to infer from growth outcomes, as the economy evolves, whether the true magnitude of the persistence factor, ρ , is moderate or high. This leads to an endogenously time-varying ambiguity, an apposite dynamic property. In essence, learning about the true persistence induces heteroskedasticity (of beliefs) since forecasts about near future growth prospects, predicated on the two possible levels of persistence, may credibly disagree, making the agent’s belief about these prospects more uncertain time to time, depending on history. In our model, any significant shock that comes against the run of recent history will sow seeds of *increased* doubt and ambiguity in beliefs. Following a positive shock, the uncertainty is whether “green shoots” have actually arrived and whether they will take root. After a negative shock, the uncertainty involves the possibility of a recession, including one that might take quite long to recover from. We find the ambiguity averse agent will react *asymmetrically* to these uncertainties; she will be more anxious with the latter, seeking a more robust and secure portfolio.

The ambiguity is measured from data, as just described, while ambiguity aversion is calibrated to match the first moment of the risk-free rate in data. Relative risk aversion is constrained to lie between one and three. This yields predictions of the equilibrium rates of return from the (amended) Lucas-tree model. We provide arguments why the magnitudes of ambiguity aversion so invoked are plausible and consistent with experimental data, and find that the level, volatility *and* movements of the predicted equity premium, in particular its counter-cyclical, closely matches the data.

We describe next how the analysis here relates to other explanations in the literature (of the observed behavior of equity premium) based on *aggregate* uncertainty in representative agent frameworks. Three papers closest to ours are Bansal and Yaron (2004), Hansen and Sargent (2010) and Ju and Miao (2010).

Bansal and Yaron (2004) pioneered the use of the (basic) stochastic model we apply. They used the model to show how long run risk (LRR) and aversion to such risk (while allowing a Kreps and Porteus (1978)/Epstein and Zin (1991) like separation of intertemporal elasticity of substitution (IES) from risk aversion) could explain aspects of the observed equity premium. The new perspective developed in our paper is that the same stochastic model with minimal changes can serve as a tractable and interesting model of ambiguity about macroeconomic risk with beliefs substantially tied to data. The changes we introduce are: (1) letting the belief about the latent state be the full Bayes posterior, instead of degenerate, probability-one-belief on the filtered state; (2) letting the agent face uncertainty about the level of the persistence parameter, updating beliefs about the level through observation of growth outcomes, instead of assuming that the agent believes with probability one that the persistence parameter has a high value. We also assume that volatility of innovations to consumption is constant, as in their CASE I model, and not endowed with an exogenous, stochastic volatility as in their CASE II. We show, (1) and (2) are actually enough to yield a models of beliefs where the uncertainty and ambiguity

endogenously varies over time, in interesting and intuitive ways. In this way, we believe, the analysis in the paper demonstrates a broader scope of application of the LRR framework.

Hansen and Sargent (2010) study the effect of model uncertainty and robustness concerns on price of macroeconomic risk. They focus on model uncertainty between a Bansal-Yaron style LRR model and an i.i.d. specification (which may be thought of as a trivial LRR model, with the persistence parameter set to zero). In their model the agent's second order uncertainty is a two point belief. One point of the support is the i.i.d. growth parameter while the other is the filtered latent state in the long run risk model (i.e., conditional on the LRR model, the agent's belief is degenerate on the filtered state). In our model, the agent's second order beliefs is a mixture distribution of two non-degenerate posteriors on latent states from two *non-trivial* long run risk models. (Even if our agent were sure which of the two LRR models is the correct, she would still face ambiguity, as the expected growth rate would remain uncertain.) Their model of beliefs does not incorporate the two modifications we introduce to the LRR framework. The first allows for a more complete representation of the ambiguity arising from the mean uncertainty since every value of the mean parameter the agent thinks is possible is included in the support of the belief. The second is motivated by empirical considerations. For instance, over the sample period we consider the i.i.d. model is very largely dominated (in terms of likelihood) by the LRR model. As a formulation of ambiguity aversion their model of agent's preferences corresponds to a particular parametrization of the smooth ambiguity model with the ambiguity attitude specified as an exponential function and the risk attitude as the log function. This particular specification is somewhat limiting and atypical since in this (knife-edge) case learning and ambiguity has no effect on equity return (it only affects the risk-free return, see section 3.1.6). An advantage of the Hansen and Sargent (2010) model is that it incorporates effects of having a separate IES parameter uncoupled from risk aversion parameter unlike ours, which focuses on the effects of ambiguity sensitivity. Thus one may view this paper as building on and extending the approach of Hansen and Sargent in ways that allow us to obtain a fuller account of the pure effect of ambiguity on rates of return. Of course, our two directions of generalization meant we had a model that fell outside the class where their methods of solution applied.

Ju and Miao (2010) use the smooth ambiguity framework to assess the effect of ambiguity on dynamics of asset prices. The model of stochastic evolution of belief is different from the one used in this paper and the parametric specification of ambiguity attitude is also different. The hidden/latent state variable driving the (mean) growth rate in the economy may take a continuum of values in our model, while it may take only two possible values (and thus only two possible growth rates) in their model. The ambiguity aversion function in this paper is of the exponential form, whereas in Ju and Miao it is of the power variety, with a further incorporation of a separate IES parameter as in Hansen and Sargent. The exponential form has the advantage that it makes the mechanism by which ambiguity aversion affects the rates of return in equilibrium more transparent. As we shall see, this transparency is facilitated by the fact that this formulation, unlike the power-power specification, lends itself to a (node-specific) change of measure interpretation. Given the richer description of uncertainty

in our model of beliefs, we do not need a *separate* IES parameter (separate from risk aversion) to match data. In their model without the separate IES effect added in, the level of equity premium matches data but the volatility is well off. Moreover, our results are based on the actual history of consumption and are not expected values based solely on the assumed process of growth, obtained by averaging over counter-factual (simulated) sample paths. This, we believe, is important since we want to know whether the ambiguity in beliefs conditional on the *observed* history can explain the observed statistics of rates of return.

Gollier (2009) shows analytically, using a (static) smooth ambiguity model, that an increase in ambiguity aversion may not, in general, increase the equity premium. The finding makes a good case for empirical investigation of the question of the connection between ambiguity and equity premium. Abel (2002), Cecchetti, Lam, and Nelson (2000), Giordani and Soderlind (2006), Jouini and Napp (2006), show that exogenously introducing pessimism and doubt in beliefs can generate a realistic equity premium and risk-free rate. Our results are driven by similar elements of pessimism and doubt, but in our framework these arise endogenously. Barro (2006), and Weitzman (2007) show that rare risks and/or heavy tails may contribute to the large equity premium and low risk-free rate observed in the data. Our contribution focuses on “common” uncertainty near the current growth rate rather than on “rare” uncertainty, and so is easier to relate to observed consumption data. Constantinides (1990) and Campbell and Cochrane (1999) study a model with habits in consumption which can match the level, variation and counter-cyclicity of the equity premia. Habits effectively allow the risk aversion to vary endogenously over the business cycle. The crucial difference to our paper is that we have constant aversion (to ambiguity and risk) but our agent faces time-varying uncertainty and it is that variation in uncertainty, rather than in the aversion to it, which causes the returns and premia to vary.

The rest of the paper is organized as follows. Section 2 introduces the relevant details of smooth ambiguity preferences, describes and analyzes the amended Lucas tree economy. In particular, we show how the presence of ambiguity aversion affects the Euler equations. Section 3, the heart of the paper, develops the specifics of our quantitative model of ambiguous beliefs and derives and explains the quantitative implications of such beliefs on level and time variation of rates of returns. Section 4 addresses the question whether the magnitude of ambiguity aversion we invoke is reasonable. A last section concludes.

2 Smooth ambiguity and the Lucas tree

We describe here the analytical framework used. It consists of a usual Lucas tree economy but for the agent’s preferences and beliefs. The former are of the smooth ambiguity type posited in (Klibanoff, Marinacci, and Mukerji, 2005, 2009). We first recall the basics of static smooth ambiguity preferences and its recursive extension to a dynamic setting. Next we describe the general economy under consideration. Finally, we derive equilibrium Euler equations for interest rates and asset returns. More specific assumptions imposed on agent’s beliefs about the stochastic process generating consumption and dividend outcomes are introduced in section 3; here we assume a fairly general form of beliefs.

2.1 Agent's preferences: the smooth ambiguity model and its recursive formulation

KMM2005 proposed and axiomatized the smooth ambiguity model of preferences over acts, which map states to payoffs, such that the decision maker prefers act f to act g if and only if $E_\mu \phi (E_{\pi_\theta} u \circ f) \geq E_\mu \phi (E_{\pi_\theta} u \circ g)$, where E is an expectation operator, u is a vN-M utility function, ϕ is an increasing transformation, and μ is a subjective probability over a set Θ of parameters, an element of which identifies a probability measure π_θ over the outcomes. Thus μ represents the decision maker's subjective uncertainty about the "true" probability distribution governing outcomes. A key feature of this model is that it separates ambiguity, identified as a characteristic of the decision maker's subjective beliefs, and ambiguity attitude, a characteristic of the decision maker's tastes. Attitudes towards pure risk are characterized by the shape of u , as usual, while attitudes towards ambiguity are characterized by the shape of ϕ , given u . In particular, a concave ϕ characterizes ambiguity aversion, which is defined to be an aversion to mean preserving spreads in the distribution over expected utility values, $E_{\pi_\theta} u \circ f$, induced by μ and f . This distribution represents the uncertainty about ex ante evaluation; it shows the probabilities of different evaluations of the act f under the different (first-order) probabilities subjectively deemed as relevant. Intuitively, ambiguity averse decision makers prefer acts whose evaluation is more robust to the possible variation in probabilities. This preference model does not, in general, impose reduction between μ and the π_θ 's in the support of μ . Such reduction only occurs when ϕ is linear, a situation identified with ambiguity neutrality and wherein the preferences are observationally equivalent to that of a subjective expected utility maximizer with a subjective prior μ (over parameters).

In the present paper we follow KMM2009, which develops a dynamic, recursive version of the smooth ambiguity model. In KMM2009 the basis of the dynamic model is the state space S , the set of all observation paths generated by an event tree, a graph of decision/observation nodes. The root node of the tree, s^0 , branches out into a set of immediate successor nodes, $S^1 \ni s^1 \equiv (s^0, s_1)$ where $s_1 \in \mathcal{S}_1$, the set of possible observations at time period 1. The decision maker chooses between (consumption) plans f , each of which associates a payoff to a node s^t in the event tree. The decision maker is uncertain about which stochastic process governs the probabilities on the event tree. The domain of this uncertainty is given by a parameter space Θ , the set of (unobservable) parameters, over which the decision maker makes inference at each s^t . We denote by $\pi_\theta (s_{t+1} | s^t)$ the probability under distribution π_θ that the next observation will be s_{t+1} , given that node s^t is reached. The decision maker's prior on the parameter space Θ is denoted by μ . KMM2009 give assumptions such that *recursive smooth ambiguity preferences* over plans f at a node s^t , are updated and represented as:

$$V_{s^t}(f) = u(f(s^t)) + \beta \phi^{-1} \left[\int_{\Theta} \phi \left(\int_{\mathcal{S}_{t+1}} V_{(s^t, s_{t+1})}(f) d\pi_\theta(s_{t+1} | s^t) \right) d\mu(\theta | s^t) \right], \quad (1)$$

where $V_{s^t}(f)$ is a recursively defined (*direct*) value function, u is a vN-M utility index, β is a discount factor and ϕ a function whose shape characterizes the decision maker's ambiguity attitude, while $\mu(\cdot | s^t)$, denotes the Bayesian posterior, describing the decision maker's updated belief at s^t .

2.2 A Lucas-tree economy and Euler equations

There is an infinitely-lived agent, with recursive smooth ambiguity preferences, consuming a single good. She can trade in a risk-free asset, whose holding and price at time t are denoted b_t and p_t^f respectively. There is also an asset (whose quantity is normalized to 1 unit) that yields a stochastic dividend at each period, D_t . The asset with uncertain dividend (usually dubbed, the “risky” asset) has a price p_t at time t , and its holding is denoted e_t . Consumption at time t is denoted C_t .

As in Bansal and Yaron (2004) and Campbell (1996) we will assume that dividend and consumption follow different stochastic processes, thus departing from the original Lucas tree economy. The gap between consumption and dividend is due to some (exogenously given) labor income l_t . Equilibrium will require that at each time $C_t = l_t + D_t$. It is thus equivalent to derive the stochastic process followed by C_t from the assumed processes for D_t and l_t as we do in this section or to assume directly a stochastic process for C_t and D_t , leaving the process for l_t implicit. Thus, we can indifferently view a node s^t in the tree describing the economy as an observed history of realizations given either by the list $\{(D_\tau, l_\tau)\}_{\tau=0}^t$ or by $\{(C_\tau, D_\tau)\}_{\tau=0}^t$.

Next, we derive Euler equations that (implicitly) define equilibrium prices in this economy. At each node, let μ_t denote the (second order) belief on parameters in Θ defining (first order) probability distributions on immediate successors (C_{t+1}, D_{t+1}) . Beliefs are updated as a function of the observed realizations of the consumption and dividend signals according to Bayes law. Wealth at time $t + 1$ is $W_{t+1} = (p_{t+1} + D_{t+1})e_t + b_t + l_{t+1}$, and the budget constraint in period t is given by $C_t = W_t - p_t e_t - p_t^f b_t$. The agent’s maximization problem may be described in terms of a recursive Bellman equation given by:

$$J(W_t, \mu_t) = \max_{C_t, b_t, e_t} u(C_t) + \beta \phi^{-1} [E_{\mu_t} (\phi(E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1}))))], \quad (2)$$

subject to the budget constraint and the law of motion of the two “state” variables (wealth and beliefs), where $J(W_t, \mu_t)$ denotes a recursively defined *indirect* value function. An equilibrium of this economy is given by $\{(p_\tau, p_\tau^f, e_\tau, b_\tau, C_\tau)\}_{\tau=0}^\infty$ such that the consumption and asset holding processes solve the maximization program and furthermore the market clears, i.e., $e_t = 1$, $b_t = 0$, $C_t = D_t + l_t$ for any t .

First order conditions are given by:

$$\beta \Upsilon_t E_{\mu_t} [\xi_t(\theta) E_{\pi_\theta} (u'(C_{t+1}))] = p_t^f u'(C_t) \quad (3)$$

$$\beta \Upsilon_t E_{\mu_t} [\xi_t(\theta) E_{\pi_\theta} ((p_{t+1} + D_{t+1}) u'(C_{t+1}))] = p_t u'(C_t) \quad (4)$$

where $\Upsilon_t = E_{\mu_t} [\phi'(E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1}))) \times (\phi^{-1})'] [E_{\mu_t} (\phi(E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1}))))]$ and where

$$\xi_t(\theta) = \frac{\phi'(E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1})))}{E_{\mu_t} [\phi'(E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1})))]} \quad (5)$$

The expressions are thus similar to those in an economy where the agent is an expected utility maximizer, but for the terms Υ_t and ξ_t . Both Υ_t and ξ_t depend on the ambiguity attitude, ϕ , and beliefs. The function ξ_t is

a Radon–Nikodym derivative, effecting a *node specific* change of measure, or “distortion”, on the posterior μ_t . The distortion is a function of the continuation values that are obtained at successor nodes. In this paper we assume an exponential ambiguity attitude, $\phi(x) = -\exp(-\alpha x)/\alpha$, where the parameter α represents ambiguity attitude. *This* specification allows us to simplify these expressions significantly, since we now have $\Upsilon_t = 1$, and the change of measure takes the form,

$$\xi_t(\theta; \alpha) \equiv \frac{\exp(-\alpha(E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1}))))}{E_{\mu_t}[\exp(-\alpha(E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1}))))]} \quad (6)$$

Further, assume the vN-M utility u takes the power form $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$. With these specifications, Euler equation determining the risk-free rate is:

$$\beta E_{\mu_t} \xi_t(\theta; \alpha) E_{\pi_\theta} \left[\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \right] = p_t^f \Leftrightarrow \quad (7)$$

$$\beta R_t^f E_{\mu_t} [\xi_t(\theta; \alpha) E_{\pi_\theta} E_{\pi_\theta} [\exp(-\gamma g_{t+1})]] = 1 \quad (8)$$

And, the equation for price of the risky asset simplifies to:

$$\beta E_{\mu_t} \left[\xi_t(\theta; \alpha) E_{\pi_\theta} \left[\frac{p_{t+1} + D_{t+1}}{p_t} \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \right] \right] = 1 \Leftrightarrow \quad (9)$$

$$\beta E_{\mu_t} \left[\xi_t(\theta; \alpha) E_{\pi_\theta} \left[\left(\frac{\exp(z_{t+1}) + 1}{\exp(z_t)} \right) \exp(d_{t+1} - \gamma g_{t+1}) \right] \right] = 1 \Leftrightarrow \quad (10)$$

$$\beta E_{\mu_t} [\xi_t(\theta; \alpha) E_{\pi_\theta} [R_{t+1} \exp(-\gamma g_{t+1})]] = 1 \quad (11)$$

where $z_t = \log\left(\frac{p_t}{D_t}\right)$, $g_{t+1} = \log\left(\frac{C_{t+1}}{C_t}\right)$, $d_{t+1} = \log\left(\frac{D_{t+1}}{D_t}\right)$, the logarithm of price-dividend ratio, rates of growth of consumption and dividend, respectively, while $R_t^f = \frac{1}{p_t^f}$, $R_{t+1} = \frac{p_{t+1} + D_{t+1}}{p_t}$ denote the risk-free and risky rates of return.

Remark 1 *These Euler equations seem identical to ones obtained in a standard Bayesian model except for the inclusion of the distortion function, ξ_t . The distortion, in the case of ambiguity aversion, increases the (posterior) weight on one-period ahead probability distributions π_θ with lower expected continuation values, $E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1}))$. Hence, when considered as a one-step ahead problem, the marginal trade-offs encapsulated in the Euler equations are those of a Bayesian using a different, distorted, posterior. However, the distortion is generally distinct at each node and so it is not possible to ascribe an “as if” equivalent Bayesian prior for the entire event tree, and hence the full set of Euler equations (i.e., across all nodes in the tree) cannot be interpreted as arising from a Bayesian model.*

3 Asset prices with unobserved, persistent shocks to expected growth

This section applies the asset pricing model developed in the previous section to two related specifications of the agent’s belief about the stochastic evolution of the economy. The specifications share a key feature:

the ambiguity in the agent's belief about growth realizations arises purely from her uncertainty about *expected* growth and the expected growth is uncertain because it is subject to periodic shocks. The dynamics of the shocks are assumed to follow an autoregressive process, and though expected growth is uncertain, the agent may make inferences about the current state on the basis of observed history of growth realizations. We proceed with the analysis in two parts, each based on a particular belief specification. In the first, the agent knows the persistence parameters while in the second it is assumed the agent is uncertain about this value. As will be seen, the second assumption yields a richer and more realistic dynamic picture with the extent of uncertainty varying endogenously over time and is our main model of beliefs. The first, however, is useful in setting ideas and building intuition.

3.1 When there is certainty about the persistence: the single- ρ model

3.1.1 A simple model of beliefs

Here we assume the agent believes the growth rate of consumption and dividends are driven by a common, latent state, x_t , which evolves according to an AR(1) with known persistence. This is the CASE I model in Bansal and Yaron (2004).

$$\begin{aligned} g_{t+1} &= \bar{g} + x_{t+1} + \sigma_g \epsilon_{g,t+1} \\ d_{t+1} &= \bar{d} + \psi x_{t+1} + \sigma_d \epsilon_{d,t+1} \\ x_{t+1} &= \rho x_t + \sigma_x \epsilon_{x,t+1} \end{aligned} \tag{12}$$

where $(\epsilon_{g,t+1}, \epsilon_{d,t+1}, \epsilon_{x,t+1})' \sim N(0, I)$. The long-run growth rate of consumption and dividend are shown by \bar{g} and, \bar{d} , respectively. The shock, x , on the other hand, accounts for the temporary deviation from trend due to the effect of the business cycle, which we model as an autoregressive process with a persistence factor denoted by ρ . The factor ψ accounts for the empirically observed volatility of dividend relative to that of consumption. This modeling device was introduced in Abel (1999) and is followed widely in the finance literature, including in Bansal and Yaron (2004), and may be interpreted as the “leverage ratio” on (expected) consumption growth. The agent observes, contemporaneously, the realizations of g_t and d_t but never observes the realization of x or ϵ . It is assumed that the values of parameters $(\bar{g}, \bar{d}, \sigma_g, \sigma_d, \sigma_x, \psi, \rho)$ are known to the agent. The x_0 is believed to have a Gaussian distribution with mean \bar{x}_0 and variance σ_0^2 , fixing thereby the agent's prior belief μ_0 . We call this the *single- ρ model*; its defining property being that the agent knows the value of the persistence factor ρ .

Given a current node $\{(C_\tau, D_\tau)\}_{\tau=0}^t$, the immediate successor node is completely identified by the pair of growth realizations (g_{t+1}, d_{t+1}) . Given x_t , the probability distribution over the immediate successor nodes is the product of two (conditionally independent, given x_t) Normal distributions, $g_{t+1} \sim N(\bar{g} + \rho x_t, \sigma_g^2 + \sigma_x^2)$ and $d_{t+1} \sim N(\bar{d} + \psi \rho x_t, \sigma_d^2 + \sigma_x^2)$. This product distribution is the typical first order distribution, the object $\pi_\theta(\cdot | s^t)$ in the abstract KMM formulation, with the variable x_t playing the role of the unobserved parameter “ θ ”. Knowing x_t pins down the mean of the distribution over the successor nodes; this mean parameter is all that is needed to fix the distribution. The agent is uncertain about the mean parameter and has a (second order) belief μ_t over this parameter, the current x_t . The belief μ_t describes, exhaustively, her ambiguity about

the probability distribution on the successor nodes. The agent updates her second order belief using Bayes rule conditional on the history of realizations of g_t and d_t given the Gaussian prior $g_0 \sim N(\bar{x}_0, \sigma_0^2)$. Conditional (second order) beliefs are also Gaussian with mean \hat{x}_t and a (steady state) variance \hat{P} , i.e., $x_t \sim N(\hat{x}_t, \hat{P})$, where \hat{P} is defined in the sequel. Hence, the evolution of the *second order* belief may be summarized by a single parameter, its conditional mean \hat{x}_t , the filtered value of x at time t . The filtered value is updated, given the signal extraction structure, naturally via a Kalman filter as follows:

$$\hat{x}_{t+1} = \rho \hat{x}_t + K v_{t+1}. \quad (13)$$

The coefficient K is the Kalman gain, defined as follows:

$$K = \rho \hat{P} [1 \ \psi] F^{-1} \quad \text{where} \quad F = \begin{bmatrix} \hat{P} + \sigma_g^2 & \psi \hat{P} \\ \psi \hat{P} & \psi^2 \hat{P} + \sigma_d^2 \end{bmatrix}$$

The surprise or innovation to growth is given by

$$v_{t+1} = \begin{bmatrix} g_{t+1} - \bar{g} - \rho \hat{x}_t \\ d_{t+1} - \bar{d} - \psi \rho \hat{x}_t \end{bmatrix}.$$

Finally, the steady state variance, \hat{P} , is defined as the solution to

$$\hat{P} = \rho^2 \hat{P} - \rho^2 \hat{P}^2 [1 \ \psi] F^{-1} [1 \ \psi]' + \sigma_x^2. \quad (14)$$

3.1.2 Computing the rates of return

Given this specification of beliefs, the continuation value of holding a Lucas tree at time t is completely determined by the consumption and the parameter value describing the second order belief at t , i.e., the pair $(C_t; \hat{x}_t)$. The *direct* value function, adapted to the given specification, is:

$$V(C_t; \hat{x}_t) = u(C_t) + \beta \phi^{-1} (E_{\hat{x}_t} \phi (E_{x_t} V(C_t \exp(g_{t+1}); \hat{x}_{t+1}))) \quad (15)$$

where the operator $E_{\hat{x}_t}$ takes expectations over x_t with respect to the measure $N(\hat{x}_t, \hat{P})$ and E_{x_t} takes expectations over g_{t+1} and d_{t+1} , with respect to the bivariate normal,

$$N \left(\begin{bmatrix} \bar{g} + \rho x_t \\ \bar{d} + \psi \rho x_t \end{bmatrix}, \begin{bmatrix} \sigma_g^2 + \sigma_x^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_d^2 + \sigma_x^2 \end{bmatrix} \right), \quad (16)$$

and \hat{x}_{t+1} is related to \hat{x}_t as in eq. (13). The adapted Euler equations are:

$$\beta R_t^f E_{\hat{x}_t} \xi_t(x_t | C_t, \hat{x}_t; \alpha) \left[E_{x_t} \left(\left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \right) \right] = 1 \quad (17)$$

$$\beta E_{\hat{x}_t} \xi_t(x_t | C_t, \hat{x}_t; \alpha) \left[E_{x_t} \left(R_{t+1} \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \right) \right] = 1 \quad (18)$$

with the distortion function⁴ given as,

$$\xi_t(x_t | C_t, \hat{x}_t; \alpha) \equiv \frac{\exp(-\alpha(E_{x_t}(V(C_{t+1}; \hat{x}_{t+1}))))}{E_{\hat{x}_t}[\exp(-\alpha(E_{x_t}(V(C_{t+1}; \hat{x}_{t+1}))))]}. \quad (19)$$

The first step toward solving the model is to compute the direct value function. To that end, we assume that the direct value function can be approximated by

$$V(C_t; \hat{x}_t) \simeq \Phi_\nu(X_t) = \exp\left(\sum_{i_c, i_x \in \mathcal{I}} \theta_{i_c, i_x}^\nu H_{i_c}(\varphi_c(C_t)) H_{i_x}(\varphi_x(\hat{x}_t))\right) \quad (20)$$

where $X_t \equiv (C_t, \hat{x}_t)$ denotes the vector of “state variables” of the single- ρ model. The set of indices, $\mathcal{I} = \{i_z = 1, \dots, n_z; z \in \{C, x\} | i_c + i_x \leq \max(n_c, n_x)\}$, was chosen to ensure that we consider a complete basis of polynomials. The function $H_l(\cdot)$ is a Hermite polynomial of order l and $\varphi_z(\cdot)$, a strictly increasing function that maps \mathbb{R} into \mathbb{R} , is used to map Hermitian nodes into values for the vector of state variables. The vector of parameters θ^ν is then determined by a minimum weighted residuals method, using a Gauss Hermitian quadrature to approximate integrals involved in the computation of the expectations. Once a solution to the value function is obtained, we are in a position to compute an approximate solution for the rates of returns. The risk-free and risky rate, $R_t^f(C_t, \hat{x}_t; \alpha, \gamma)$ and $R_t(C_t, \hat{x}_t; \alpha, \gamma)$, are computed numerically by solving eqs. (17) and (18), after substituting the value function (in the expression for $\xi_t(x_t | C_t, \hat{x}_t; \alpha)$) by its approximate solution. Full details of the computation method may be found in Appendix D, which also gives details on accuracy checks, showing that the numerical solution is highly accurate. Separate from the numerical solution to rates of return we also obtain analytical approximations, discussed in Section 3.1.6.

3.1.3 Data and parameter values

The time-series parameters of the model (except for the persistence parameter ρ and the leverage-ratio parameter ψ) were estimated using maximum likelihood on annual U.S. data from 1930 to 1977 (see appendix C for details). By 1977 the parameter estimates had stabilized and the remaining years in the data set, 1978-2007, were used in the evaluation of the models. Hence, we have some justification in assuming that the agent behaves as if she knows the parameter values of the model from 1977 onwards. We set $\rho = 0.85$ in the most part (for the single- ρ case), which is the annualized equivalent of the value used in Bansal and Yaron (2004) and supported by the estimate obtained in Bansal, Gallant, and Tauchen (2007). The other value we apply in the single- ρ model, $\rho = 0.9$, is used to check for robustness and is approximately the annualized equivalent to the calibration used by Hansen and Sargent (2010). The dividend leverage parameter, ψ , was set to 3 as in Bansal and Yaron (2004), although Constantinides and Ghosh (2010) estimated it to be slightly lower, close to the value we use for robustness checks ($\psi = 2.5$).

⁴Henceforth, we shall write ξ_t as a function of direct continuation value $V(C_{t+1}; \hat{x}_{t+1})$ instead of the indirect value, $J(W_{t+1}, \mu_{t+1})$. In a single agent economy consumption equals the endowment and is thus, exogenously determined, and so it is possible to solve for the (continuation) value at any node on the event tree without solving for the equilibrium prices first.

Equity returns are computed using the CRSP value-weighted index. Dividend growth is imputed using the difference in the returns on the value-weighted index with and without dividends multiplied by the market value. The risk-free rate was taken from Ken French's data library. Consumption is defined as the sum of services and non-durable consumption and was taken from BEA Table 1.1. Population was taken from BEA Table 2.2. Both per-capita consumption growth and dividend growth were converted to real terms using the average CPI for the year taken from the BLS. Annual data was available from 1930 until 2007, a total of 78 observations.

Turning to preference parameters, in all cases the ambiguity aversion parameter α was calibrated to produce a real risk-free rate of 1.85%, averaged over $t = 1978, \dots, 2007$, which is the average observed rate in that period. No other moments were used in the choice of α . The relative risk aversion parameter γ was allowed to range between 1 (log utility) and 3, regarded as plausible in macroeconomic models (Ljungqvist and Sargent, 2004, pg. 426).⁵ The discount factor β was set to 0.95 (Weil, 1989). To check for robustness we varied a number of the key non-estimated parameters, including $\rho = 0.9$, $\beta = .97$ and $\psi = 2.5$.

3.1.4 Results

We use annual data on real per capita consumption C_t and estimates of \hat{x}_t corresponding to the filtration imposed by the observed history of growth in real consumption and the real dividend to obtain a time series of *conditional* moments rates of return using our numerical solution technique. That is, a time series of the conditional first and second moments of the random variables $r_t \equiv R_t - 1$ and $r_t - r_t^f$, predicted by the model along the sample path, conditional on the history at each time t . Table 1 reports the average of these conditional moments, averaged over the period 1978-2007. The point of looking at conditional moments is that we want to see predictions of rates of return conditional on (our measures of) the ambiguity that obtained along the actual sample path. The results show that the first moments of the risky rate and equity premium are well predicted by the model, but the second moments are smaller than what is observed in the data.

3.1.5 The mechanism of ambiguity aversion: endogenous pessimism and doubt

The intuition behind the mechanism of ambiguity and ambiguity aversion driving the results can be understood through the distortion function, $\xi_t(x_t | C_t, \hat{x}_t; \alpha)$. Given the posterior $N(\hat{x}_t, \hat{P})$ on x_t the effect of ξ_t is "as if" there is a new distorted posterior, $\tilde{\mu}_t \equiv \xi_t(x_t) \otimes N(\hat{x}_t, \hat{P})$, with density given by

$$\tilde{f}(x_t) = \xi_t(x_t | C_t, \hat{x}_t; \alpha) \frac{1}{\sqrt{2\pi\hat{P}}} \exp\left(-\frac{(x_t - \hat{x}_t)^2}{2\hat{P}}\right). \quad (21)$$

In the case of ambiguity aversion, i.e., $\alpha > 0$, it is evident from eq. (19) that $\tilde{\mu}_t$ puts relatively greater probability mass (compared to μ_t) on x_t 's that generate probability distributions associated with lower expected continu-

⁵As shown in KMM2005, a smooth ambiguity preference A is more ambiguity averse than another such preference B if, they share the same risk attitude, and ambiguity attitude parameter ϕ_A is more concave than ϕ_B . If the two preferences do not share the same risk attitude it is not necessarily true that a more concave ϕ means more ambiguity aversion. Hence α is meaningfully calibrated *given* a value of γ ; not independent of γ .

	γ	α	$E(r)$	$E(r - r^f)$	$\sigma(r^f)$	$\sigma(r)$	$\sigma(r - r^f)$
Data			9.80	7.95	2.08	13.5	13.9
Single- ρ Model Predictions							
	1.00	41.2	10.7	8.86	0.84	6.31	6.23
	1.50	24.9	10.1	8.30	1.37	7.29	7.14
	2.00	15.6	10.3	8.47	1.94	8.43	8.21
	2.50	9.27	10.7	8.93	2.50	9.66	9.36
Robustness Checks							
$\rho = 0.90$	2.00	10.6	10.0	8.19	2.17	8.70	8.40
$\psi = 2.5$	2.00	14.6	9.60	7.75	2.02	7.96	7.70
$\beta = 0.97$	2.00	10.2	7.69	5.84	1.88	8.24	8.02
Bayesian	2.00	0	9.12	0.038	1.93	8.37	8.14

Table 1: **Results (single- ρ)**: Average of the predicted *conditional* moments of rates of return (on dividend claim) in the single- ρ model over the period 1978–2007. The top panel varies the risk aversion coefficient and only reports the values computed when the coefficient of ambiguity aversion was calibrated to produce a risk-free rate of 1.85%. The bottom panel contains a series of robustness checks where the parameter in the left-most column was changed from the basic specification ($\rho = 0.85$, $\psi = 3$, $\beta = 0.95$). The final line contains the Bayesian case, that is $\alpha = 0$.

ation values, $E_{x_t}(V(C_{t+1}; \hat{x}_{t+1}))$. The distorted posterior gives rise to a conditional one-step-ahead distribution on growth which we call the *twisted (predictive) distribution*

$$g_{t+1} \sim \xi_t(x_t) \otimes N(\hat{x}_t, \hat{P}) \otimes N(\rho x_t + \bar{g}, \sigma_x^2 + \sigma_g^2). \quad (22)$$

When $\xi_t(x_t) = 1$ the formula (22) describes the belief of a Savage-Bayes rational (or, equivalently, ambiguity neutral) agent, an useful benchmark. Such an agent is uncertain about x_t and with belief about growth described by a mixture distribution, a mixture of Normals, with the weights of the mixture given by another Normal. We may think of this distribution as a “best estimate” distribution. The twisted distribution, on the other hand, describes the “as if” belief of an ambiguity sensitive agent; she uses this distribution, instead of the best estimate distribution, to evaluate the *equilibrium* portfolio. An ambiguity averse agent is wary of the uncertainty about the growth distribution and suspicious how good an estimate the posterior is. To ensure a more robust choice, the agent evaluates a prospect by testing it against a distribution which is somewhat less favorable to the prospect than the Bayesian posterior. The “as if” belief is the belief used to make the robustness check.⁶

Another useful benchmark is the belief of an agent with rational expectations, narrowly defined. This distribution is $N(\rho \hat{x}_t + \bar{g}, \sigma_x^2 + \sigma_g^2)$. It arises from a posterior that is degenerate on \hat{x}_t , displaying full/firm belief

⁶Different portfolios will be evaluated against, in general, different “as if” beliefs, since as the portfolio considered varies the continuation values vary too, thereby affecting the distortion. The twisted distribution associated with the Euler equation is the “as if” belief used to evaluate the equilibrium portfolio.

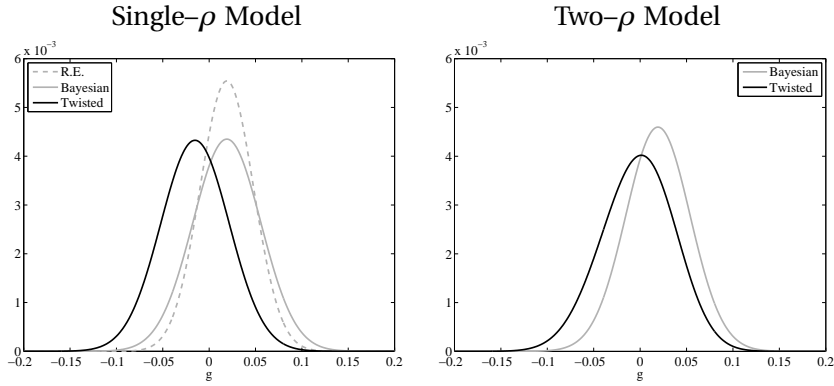


Figure 1: **Beliefs and “as-if” beliefs:** The left panel contains the agents “as-if” belief about the conditional distribution of consumption growth with no uncertainty about the latent state (R.E.), with uncertainty about the latent state but without ambiguity aversion (Bayesian) and with ambiguity aversion about the uncertainty of the latent state (Twisted) from the single- ρ model. The right panel shows the analogous construct in the two- ρ model (presented in Section 3.2) where the no uncertainty case is not possible by construction. In both panels, the distributions were computed setting $\gamma = 2$, and given that, α calibrated to produce a risk-free rate of 1.85%, and the level of consumption and the hidden state variables (single- ρ : \hat{x}_t , two- ρ : $\hat{x}_{l,t}$, $\hat{x}_{h,t}$, η_t) set to their averages over the period 1978–2007.

about the latent state. The left panel in Figure 1 shows the average one-step-ahead distributions (on growth) corresponding to these three cases of beliefs in the single- ρ model. Compared to the rational expectations distribution, the twisted distribution (under ambiguity aversion) has a lower mean *and* a larger spread. Abel (2002) argues that one can account for the observed equity premium and the risk-free rate by invoking pessimism and doubt in an otherwise standard asset pricing (Lucas tree) model. Pessimism is deemed, by Abel, as a subjective distribution (on growth) that is first order stochastically dominated by the “objective” distribution; doubt, corresponds to a subjective distribution that is a mean preserving spread of the objective distribution. Evidently, an ambiguity averse agent’s conditional (“as if”) beliefs, in effect, incorporate *endogenously both* these elements, pessimism *and* doubt, which is the key to understanding the mechanism through which ambiguity aversion affects asset returns.

3.1.6 Analytical Approximation and comparative statics

The results compiled in Tables 1 through 2 and Figures 1 and 2 are based on *numerical* solutions. However, in the case of the single- ρ model we can also find an *analytical* approximate solution (see Appendix A for details of the derivation) which is useful in understanding the qualitative effects of the elements of the tuple $(C_t, \hat{x}_t; \alpha, \gamma)$ on the rates of return. The key assumption used to derive the analytical approximation is that the density of the distorted posterior, described in eq. (21), is well approximated by a Normal density, whose mean and variance

are denoted by \tilde{x}_t and $\widetilde{Var}_t(x_t)$, respectively.⁷

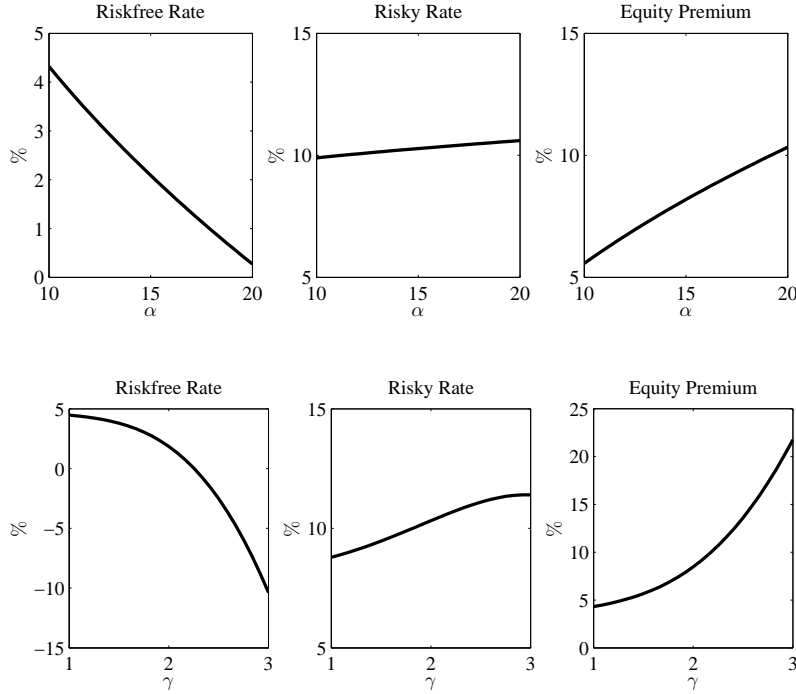


Figure 2: **Comparative statics (single- ρ)**:The top panel shows how the (conditional) risk-free rate, risky rate and equity premium change as the ambiguity parameter, α , varies with $\gamma = 2$. The bottom panel shows how the three rates change as the coefficient of risk aversion varies where the coefficient of ambiguity aversion was fixed at 15.6, the calibrated value which produces a risk-free rate of 1.85% with $\gamma = 2$. The average comparative statics are constructed by first computing the comparative statics for each year (based on numerical solutions, *not* the analytical approximation) using the filtered values of the latent state \hat{x}_t computed from the data and then averaging across $t = 1978, \dots 2007$.

The upper panel in Figure 2 depicts the comparative statics of ambiguity aversion on the rates of return which we discuss and explain, next, using the analytical approximation. The lower panel in Figure 2 shows the comparative statics of risk aversion (γ), for which the intuitions are just the same as they are in the standard case, as is evident from the approximation expressions. The risk-free rate is approximated as:

$$r_t^f = -\ln \beta + \gamma \bar{g} + \gamma \rho \tilde{x}_t - \frac{\gamma^2}{2} \left(\sigma_x^2 + \sigma_g^2 + \rho^2 \widetilde{Var}_t(x_t) \right). \quad (23)$$

An increase in ambiguity aversion, α , decreases \tilde{x}_t (see Figure 8), inducing a more pessimistic “as if” distribution (Figure 1) and making the agent behave as if she were expecting a lower endowment income in future (states); a larger ρ prolongs the expected effect of the shock to \tilde{x}_t . Buying more of the risk-free asset allows the agent to shift consumption from today to those future states. The strength of this desire to smooth consumption depends

⁷As may perhaps be intuited from the left hand panel in Figure 1 and seen more precisely from skewness and excess kurtosis numbers in Table 4 this is a particularly good approximation in the case of the single- ρ model. Indeed, as the table shows, in the case of the single- ρ model the variance is virtually unaffected by ambiguity aversion, with $\widetilde{Var}_t(x_t) \simeq \hat{P}$ (defined as in eq. (14)).

on the marginal utility of consumption in those states. If IES is low, as it is when γ is high, there will be increased emphasis on offsetting the greater future consumption. All this is encapsulated in the $\gamma\rho\tilde{x}_t$ term in (23) which shows an increase in ambiguity aversion implies a rise in demand for the risk-less asset. The agent desires a portfolio more robust to the uncertainty/ambiguity, precipitating a “flight to quality”, driving up its equilibrium price and lowering the risk-free rate. This is a key effect of ambiguity aversion, as has been widely emphasized in the literature, see e.g., Caballero and Krishnamurthy (2008), Hansen and Sargent (2010), Uhlig (2010). It is worth contrasting this with the effect of an increase in risk aversion. *Absent* ambiguity aversion, an increase in γ has, principally, two countervailing effects. The first effect shows up in the term $\gamma\bar{g}$. Here an increase in γ makes the agent want to smooth consumption between the present and future states more; since $\bar{g} > 0$, the agent expects to consume more in the future and thus the agent wants the risk-free asset *less*. The second effect appears in the term $-\frac{\gamma^2}{2}(\sigma_x^2 + \sigma_g^2)$ reflecting the agent’s desire to smooth risk *across* the future states. This need can be met (in part) by holding more of the risk-free asset. It turns out for $\gamma < 5$ the first effect dominates and in that range we find an increase in γ *increases* the risk-free rate. (Note, the comparative statics of risk aversion shown in Figure 2 correspond to $\alpha > 0$.) Finally, an increase in $\widetilde{Var}_t(x_t)$, which may be interpreted as an increase in ambiguity since it is an increase in the uncertainty of the second-order belief, will also decrease the risk-free rate. This is evident from the final term in (23).

The first moment of the (predicted) risky rate is approximated as:

$$E_t r_{t+1} = -\ln\beta + \gamma\bar{g} - \frac{\gamma^2}{2}\sigma_g^2 + \gamma\left(\frac{2\psi - \gamma}{2}\right)\sigma_x^2 + \frac{\psi^2\rho^2}{2}\hat{P} - \frac{(\rho(\psi - \gamma))^2}{2}\widetilde{Var}_t(x_t) + \rho\gamma\tilde{x}_t - \rho\psi\tilde{x}_t + \psi\rho\hat{x}_t \quad (24)$$

where $E_t \equiv E_{\hat{x}_t} E_{x_t}$ and the operators $E_{\hat{x}_t}$ and E_{x_t} take expectations with respect to the measure $N(\hat{x}_t, \hat{P})$ and the bivariate normal shown in (16), respectively. To see why these moments represent model predictions, suppose the model were correct. That is, the asset prices at a time t obtain per the Euler eqs. (17) and (18). Then (24) describes the *conditional* expectation of a Savage-Bayes rational observer/analyst who observes these prices and uses all available information (i.e., (12), (13) and the publicly observed history through t) to predict dividend at $t + 1$. The expression (24) can be seen to imply that the (first moment of) risky rate will rise with ambiguity aversion in the relevant range of parameter values. An increase in α has two countervailing effects. The first effect, shows up in the term $\rho\gamma\tilde{x}_t$, which was also present in the expression for the risk-free rate. The intuition here is analogous; an increase in α causes an increase in the demand for the risky asset to allow the agent to transfer income to the future. The second effect is evident from the term $-\rho\psi\tilde{x}_t$. As α increases \tilde{x}_t decreases, hence decreasing the (“as if”) expected future dividend payoff from the asset causing the the agent to want to pay less for the asset. Taking out the common factor, ρ , the strength of the first effect depends on γ (as explained earlier) while the second effect is exacerbated by leverage, ψ . The net effect depends on the difference, $(\psi - \gamma)$. With log utility (i.e., $\gamma = 1$) and $\psi = 1$, for example, the two effects cancel out, though with $\gamma \in [1, 2.5]$ and $\psi = 3$, as we have here, the second effect dominates and equilibrium risky rate varies positively with ambiguity aversion.

The analytical approximation for equity premium is:

$$E_t r_{t+1} - r_t^f = \rho \psi (\hat{x}_t - \tilde{x}_t) + \psi \gamma \sigma_x^2 + \frac{\psi^2 \rho^2}{2} \text{Var}_t(x_t) + \frac{\rho^2 \psi (2\gamma - \psi)}{2} \overline{\text{Var}_t(x_t)}. \quad (25)$$

The first term shows that the premium increases with ambiguity aversion (since it increases the difference $(\hat{x}_t - \tilde{x}_t)$) and this is accentuated by persistence and leverage. A doubt factor also comes into play since the premium is increasing in the variances of the latent variable, both the actual and the “as if”. In the rational expectations case the second-order belief is degenerate, i.e., $\text{Var}_t(x_t) = 0$ (which is why, in Figure 1, Twisted is riskier than R. E.). Hence, uncertainty about the latent state causes the equity premium to go up compared to the rational expectations case even if the agent were Bayesian, though only very slightly (Table 1).

	Single- ρ		Two- ρ	
γ	α	Corr	α	Corr
2.00	15.6	0.01	18.9	-0.48
Robustness Checks				
$\rho, \rho_h = 0.90$	10.6	0.17	15.2	-0.47
$\rho_l = 0.25$	–	–	16.9	-0.48
$\psi = 2.5$	14.6	0.02	19.1	-0.48
$\beta = 0.97$	10.2	0.03	15.0	-0.50

Table 2: **Correlation between model predicted (conditional) equity premium and HP-filtered consumption.**

In each case, α was set such that the model generates an average risk-free rate of 1.85%. The bottom panel contains a series of robustness checks where the parameter in the left-most column was changed from the basic specification (Single- ρ model: $\rho = .85$, Two- ρ model: $\rho_h = 0.85$, $\rho_l = 0.3$, Both: $\psi = 3$, $\beta = 0.95$). Note that the correlations are computed on the sample 1985–2007.

Uncertainty about the latent state affects the results on second moments even more. Since the rates of return depend on the realization of the latent state, the (conditional) second moments of the risky rate and equity premium go up, very significantly, compared to the rational expectations case. However, as Table 1 shows, while the predicted second moment of the risk-free rate matches data very well, the single- ρ model fails to predict about 40% of the volatility of the risky rate (and that of the equity premium). Table 2 reports that the single- ρ model predicts virtually no correlation between equity premium and the business cycle (as indicated by H-P filtered consumption). We interpret this as a clear indication that the single- ρ specification is inadequate as a model of (ambiguous) beliefs. The crucial stylized fact about the time variation of *expected* returns is the counter-cyclicity of equity premium; as the right panel in Figure 3 shows, the market return has been significantly negatively correlated with the business cycle in all but a brief period starting in late 70’s when inflation was very high. That the predicted premium fails to vary with time in way observed in data is due to a peculiarity of the dynamics of beliefs in the single- ρ model: the uncertainty, the riskiness, embodied in the second-order distribution is unvarying across information sets in an important sense. Recall, the second

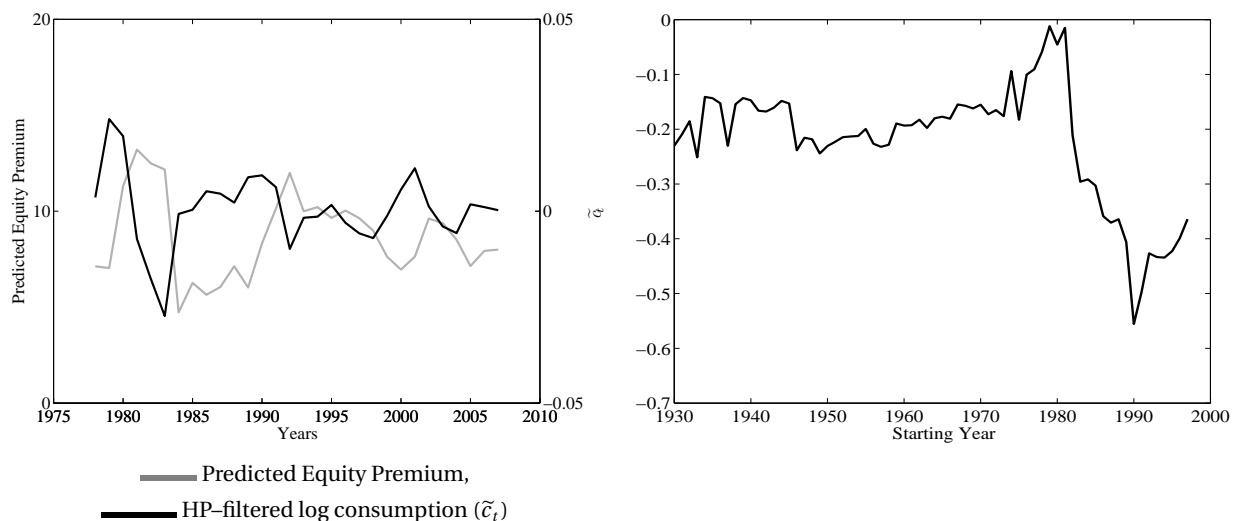


Figure 3: **Counter-cyclical equity premium:** The left panel plots the equity premium *predicted* by the two- ρ model (with $\gamma = 2$ and, given that, α calibrated to generate an average risk-free rate of 1.85%) and the transitory component of consumption growth extracted using the HP filter. The right panel contains the correlation between the “actual” equity premium (i.e., the estimated equity premium in the data) and transitory consumption growth for the sample t –2007, where t is the starting date as reported on the graph.

order distribution is given by $x_t \sim N(\hat{x}_t, \hat{P})$, hence all that changes from one information set to another is the position of the center of the distribution (all other moments remain fixed). This means this model of beliefs fails to capture intuitively evident aspects such as uncertainty/ambiguity about growth prospects varying over time, for example, increasing during recessions. In the next section we modify the single- ρ model to correct for this omission.

3.2 Uncertain persistence: the two- ρ model

Now we present and discuss our main model, what we think best captures the ambiguity in beliefs and its effect on levels and movements in asset returns.

3.2.1 Beliefs with time-varying ambiguity

Thus far the agent has been endowed with complete knowledge of the model and parameters and so has only needed to make inference on the latent state. Now we extend the single- ρ model to allow for uncertainty about the persistence of growth shocks. This extension of the model reflects the difficulty in determining, on the basis of observations of growth realizations, whether the true growth process is a very persistent process where the persistent component has a small variance or a moderately persistent process where the variance of the persistent component is larger (Shephard and Harvey, 1990). We assume the agent *believes* that the stochastic evolution of the economy follows a persistent latent state process given by a Bansal and Yaron CASE I type specification with either a low persistence (ρ^l) or a high persistence (ρ^h) with probability η_t and $1 - \eta_t$, respectively.

The actual evolution of growth rates is given by *one* of these two processes:

$$\begin{aligned}
& \textbf{Low Persistence } (\rho = \rho_l, Pr = \eta_t) \\
x_{l,t+1} &= \rho_l x_{l,t} + \sigma_{x_l} \varepsilon_{x_l,t+1} \\
d_{l,t+1} &= \bar{d} + \psi x_{l,t+1} + \sigma_{d_l} \varepsilon_{d_l,t+1} = \bar{d} + \psi (\rho_l x_{l,t} + \sigma_{x_l} \varepsilon_{x_l,t+1}) + \sigma_{d_l} \varepsilon_{d_l,t+1} \\
g_{l,t+1} &= \bar{g} + x_{l,t+1} + \sigma_{g_l} \varepsilon_{g_l,t+1} = \bar{g} + \rho_l x_{l,t} + \sigma_{x_l} \varepsilon_{x_l,t+1} + \sigma_{g_l} \varepsilon_{g_l,t+1} \\
& \textbf{High Persistence } (\rho = \rho_h, Pr = 1 - \eta_t) \\
x_{h,t+1} &= \rho_h x_{h,t} + \sigma_{x_h} \varepsilon_{x_h,t+1} \\
d_{h,t+1} &= \bar{d} + \psi x_{h,t+1} + \sigma_{d_h} \varepsilon_{d_h,t+1} = \bar{d} + \psi (\rho_h x_{h,t} + \sigma_{x_h} \varepsilon_{x_h,t+1}) + \sigma_{d_h} \varepsilon_{d_h,t+1} \\
g_{h,t+1} &= \bar{g} + x_{h,t+1} + \sigma_{g_h} \varepsilon_{g_h,t+1} = \bar{g} + \rho_h x_{h,t} + \sigma_{x_h} \varepsilon_{x_h,t+1} + \sigma_{g_h} \varepsilon_{g_h,t+1}
\end{aligned} \tag{26}$$

We call this the *two- ρ model*. The value function here depends on four state variables: the current consumption, the filtered state variables from each model (low and high persistence), and η_t , the posterior probability that the low persistence model is the "true" data generating process (DGP), and takes the form:

$$V(C_t; \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) = u(C_t) + \beta \phi^{-1}(\mathcal{V}_{t+1}),$$

where

$$\begin{aligned}
\mathcal{V}_{t+1} \equiv & \eta_t E_{\hat{x}_{l,t}} \left[\phi \left(E_{x_{l,t}} \left[V \left(C_t \exp(g_{l,t+1}), \hat{x}_{h,t+1}^{(l)}, \hat{x}_{l,t+1}^{(l)}, \eta_{t+1}^{(l)} \right) \right] \right) \right] \\
& + (1 - \eta_t) E_{\hat{x}_{h,t}} \left[\phi \left(E_{x_{h,t}} \left[V \left(C_t \exp(g_{h,t+1}), \hat{x}_{h,t+1}^{(h)}, \hat{x}_{l,t+1}^{(h)}, \eta_{t+1}^{(h)} \right) \right] \right) \right].
\end{aligned}$$

The filtered variable $\hat{x}_{j,t+1}^{(i)}$, $i = l, h, j = l, h$ is the agent's update to her belief next period if the growth outcome next period were interpreted by the Kalman filter that assumes $\rho = \rho_j$, when the data is actually generated by the model with persistence parameter $\rho = \rho_i$. For example, $\hat{x}_{h,t+1}^{(l)}$ is the value of the filtered latent state variable at $t + 1$ if the data were filtered using $\rho = \rho_h$ when in fact the data is generated by the low persistence model. Analogously, $\eta_{t+1}^{(l)}$ ($\eta_{t+1}^{(h)}$) is the Bayes update to the agent's posterior probability that the low persistence model is the correct model when the low (respectively, high) persistence model is the DGP. See Appendix B for further details, including the derivation of rates of return.

The value of ρ_h was chosen to be the same as in the single-rho model, which is 0.85 in the usual case (0.90 is used as a robustness check). The value of the parameter in the less persistent model, ρ_l , was chosen to produce a posterior probability of 50% against the model with $\rho_h = 0.85$ in 1977, the beginning of the model evaluation period. This procedure selected $\rho_l = 0.30$, though values between 0.25 and 0.35 have virtually identical posteriors (and implications for rates of returns). For these values of ρ_h and ρ_l , the posterior probability η_t stays in the interval [0.4, 0.6] throughout the period 1978-2007, demonstrating how difficult it is to separate the two models on the basis of growth data.⁸ The average value of the persistence parameter (0.57, assuming $\eta_t = 50\%$) is close to the maximum likelihood estimator. Beeler and Campbell (2009) and Constantinides and Ghosh (2010) argue

⁸Choosing ρ_l to be very small (≈ 0) so that the low persistence model is virtually i.i.d. produces posterior estimates of η_t near zero, i.e., the two- ρ model behaves almost as the single- ρ model.

that the value of the persistence parameter, when estimated on the basis of the time series properties of the growth data (i.e., without considering model specific pricing implications) is not as high as in the Bansal and Yaron calibration. Constantinides and Ghosh (2010) provide a GMM estimate (based on the years 1931-2006) of $\rho = 0.32$ (see their Table 4). As just noted, the $\rho = 0.30$ model has just as much support as the $\rho = 0.85$ model in the data, and hence it seems only appropriate that the agent's beliefs are agnostic as to which is the correct model to use for forecasting. This is one sense in which the two- ρ model, with the parameter values we adopt, is empirically more compelling than the assumption of a dogmatic belief in some value of ρ . All parameter estimates are presented in Appendix C.

There is another way in which the two- ρ model improves, empirically, on the single- ρ model: by introducing endogenously varying uncertainty of *beliefs*. Uncertainty about persistence leads to time-varying mixing of the two models (of persistence) through η_t , a belief that varies over time as the agent learns from successive growth shocks. This produces a posterior *predictive* distribution for consumption growth which is heteroskedastic even though in each model, when considered independently of the mixture, it is homoskedastic. The heteroskedasticity is influenced by two components – the spread in the filtered state from each model, and the mixing probability. The beliefs about the latent state, conditional on the low and high persistence models, are $N(\hat{x}_{l,t}, \hat{P}_l)$ and $N(\hat{x}_{h,t}, \hat{P}_h)$, respectively, and the variance of the mixture distribution of the latent state is,

$$\eta_t \hat{P}_l + (1 - \eta_t) \hat{P}_h + \eta_t (1 - \eta_t) (\hat{x}_{h,t} - \hat{x}_{l,t})^2. \quad (27)$$

Hence, when the two models have similar likelihoods – i.e., η_t is close to 50% – but quite different beliefs about the filtered state, the uncertainty in the forecast of future consumption and dividend increases. It is as if the agent has two forecasting models, and when the history is such that both models explain that history just as well (i.e., η_t is close to 0.5) and yet their core forecasts markedly disagree (i.e., $(\hat{x}_{h,t} - \hat{x}_{l,t})$ is large) the uncertainty about the mean of the growth distribution rises. In essence, learning about the true persistence model induces heteroskedasticity since from time-to-time the models disagree, credibly, about near future growth prospects, making the prospects appear more uncertain than usual. The divergence of beliefs has been strongest in the larger downturns, which also happened to be the larger shocks, and so historically, the time-variation of uncertainty has been counter-cyclical. Thus, the two- ρ model of beliefs embodies a theory of why and how ambiguity about growth prospects may vary over time.⁹

⁹The case for introducing time-varying volatility of macroeconomic variables has been argued strongly in the recent literature, e.g. Fernandez-Villaverde and Rubio-Ramírez (2010). Much of this literature, including Bansal and Yaron (2004) (see their main, CASE II, model) models this by positing an exogenously specified stochastic volatility. Beeler and Campbell (2009) and Constantinides and Ghosh (2010) argue that the assumption of highly persistent stochastic volatility of innovations to consumption (key factor underpinning the exogenous specification of stochastic volatility) is not well supported empirically. In contrast, the time-varying heteroskedasticity generated endogenously in the two- ρ model is a *forecast* uncertainty, of beliefs, empirically driven by the history of growth outcomes and consistent with a stationary volatility of consumption shocks.

	γ	α	$E(r)$	$E(r - r^f)$	$\sigma(r^f)$	$\sigma(r)$	$\sigma(r - r^f)$
Data			9.80	7.95	2.08	13.5	13.9
Two- ρ Model Predictions							
	1.00	43.2	10.3	8.47	1.21	11.3	11.3
	1.50	28.9	10.2	8.34	2.06	11.9	12.0
	2.00	18.9	10.5	8.69	2.91	12.6	12.8
	2.50	11.5	11.1	9.25	3.72	13.4	13.6
Robustness Checks							
$\rho_h = 0.90$	2.00	15.2	10.1	8.32	3.78	12.3	12.7
$\rho_l = 0.25$	2.00	16.9	10.5	8.74	2.71	12.9	12.9
$\rho_l = 0.25, \psi = 2.50$	2.00	19.1	9.97	8.11	2.97	12.9	13.0
$\beta = 0.97$	2.00	15.0	7.94	6.08	2.92	12.2	12.4
Bayesian	2.00	0.00	9.59	0.46	0.93	13.0	13.0

Table 3: **Results (two- ρ)**: Average of the predicted *conditional* moments of rates of return (on dividend claim) in the two- ρ model. The top panel varies the risk aversion coefficient and only reports the values computed where the coefficient of ambiguity aversion was calibrated to produce a risk-free rate of 1.85%. The bottom panel contains a series of robustness checks where the parameter in the left-most column was changed from the basic specification ($\rho_h = 0.85, \rho_l = 0.3, \psi = 3, \beta = 0.95$). The Bayesian case corresponds to $\alpha = 0$.

3.2.2 Results

The two- ρ model was solved using the same numerical solution techniques as in the single- ρ model. Table 3 contains predicted moments computed using parameter estimates until 1977, and consumption data between 1978 and 2007. The level of ambiguity aversion was calibrated so that the risk-free rate was 1.85% (without targeting any other features of the data). Compared with the single- ρ model (Table 1), we see little change in the first moments but a substantial increase (50%) in the second moments. The bottom panel contains experiments where the persistence parameters and the discount rate were varied (separately). These show that the increase in the variance of the risk-free and risky rates is not sensitive to the parametrization. The final line contains the Bayesian solution which corresponds to the ambiguity parameter, α , being set to 0. A Bayesian agent also sees the increase in the volatility of the risky rate and equity premium in the two- ρ specification, but the equity premium (0.46%) is tiny and average risk-free rate (9.13%) is far too high compared to the data. Figure 4 graphs the comparative statics for the two- ρ model. These are all qualitatively similar to the comparative statics presented for the single- ρ model.

What is most dramatically different from the single- ρ model is the time variation of expected returns as may be seen in Figure 3 and Table 2. The left panel in the figure contains the predicted equity premium from the two- ρ model and the transitory component of consumption extracted using the HP filter (where the filter parameter λ is set to 6.25 following the recommendation by Ravn and Uhlig (2002)), demonstrates to the extent movements in the equity premium predicted by the two- ρ model reflect movements in the business cycle. The

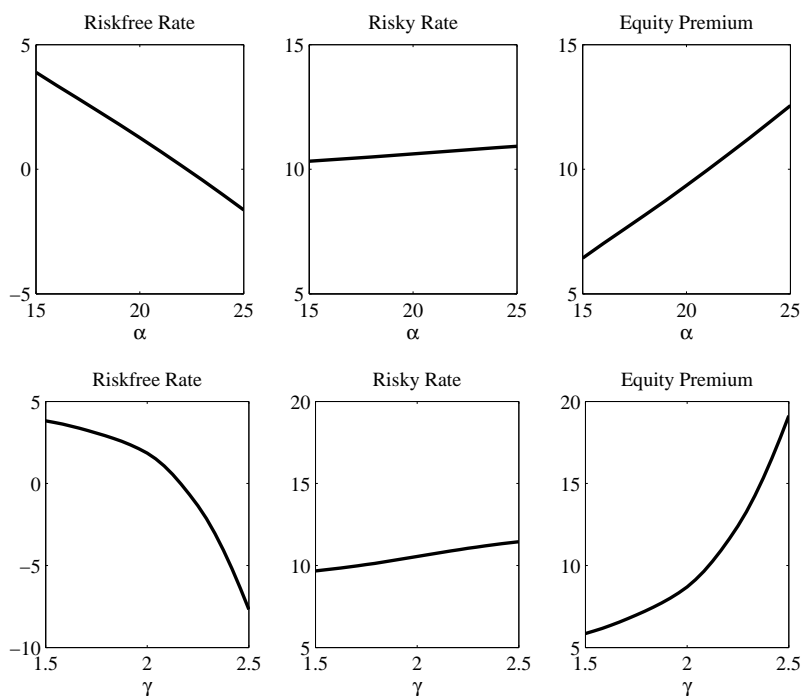


Figure 4: **Comparative statics (two- ρ)**:The top panel shows how the (conditional) risk-free rate, risky rate and equity premium change as the ambiguity parameter, α , varies when $\gamma = 2$. The bottom panel shows how the three rates change as the coefficient of risk aversion varies where the coefficient of ambiguity aversion was fixed at 15.6, the calibrated value which produces a risk-free rate of 1.85% when $\gamma = 2$. These are average comparative statics, constructed by first computing the comparative statics for each year using the filtered values of $\hat{x}_{l,t}$ and $\hat{x}_{h,t}$ and then averaging over the period 1978–2007.

table shows the stark contrast between the correlations of the predicted premium with the business cycle in the two models. The right panel in Figure 3 contains the expanding windows estimator of the correlation between HP filtered consumption and the excess return on the market. This correlation is uniformly negative, and, aside from a small period near 1980, substantially so. This correlation is especially negative in the past three decades, the period over which we evaluate the model. In the next section we discuss what explains the novelty in these results, in particular, the time-variation of expected returns.

3.2.3 The asymmetric effect of ambiguity aversion over the business cycle

Ambiguity aversion exacerbates the time-variation of the uncertainty in that the distorted posterior and the twisted predictive distributions are more skewed than the pure mixture used by a Bayesian. The mixture results in a distribution with excess kurtosis relative to a normal, and the change of measure transforms the small increase in kurtosis into substantial negative skewness. Table 4 in the Appendix A shows the magnitudes of these moments, averaged across the model evaluation period. The averages, however, do not reveal the more intriguing dynamic story.

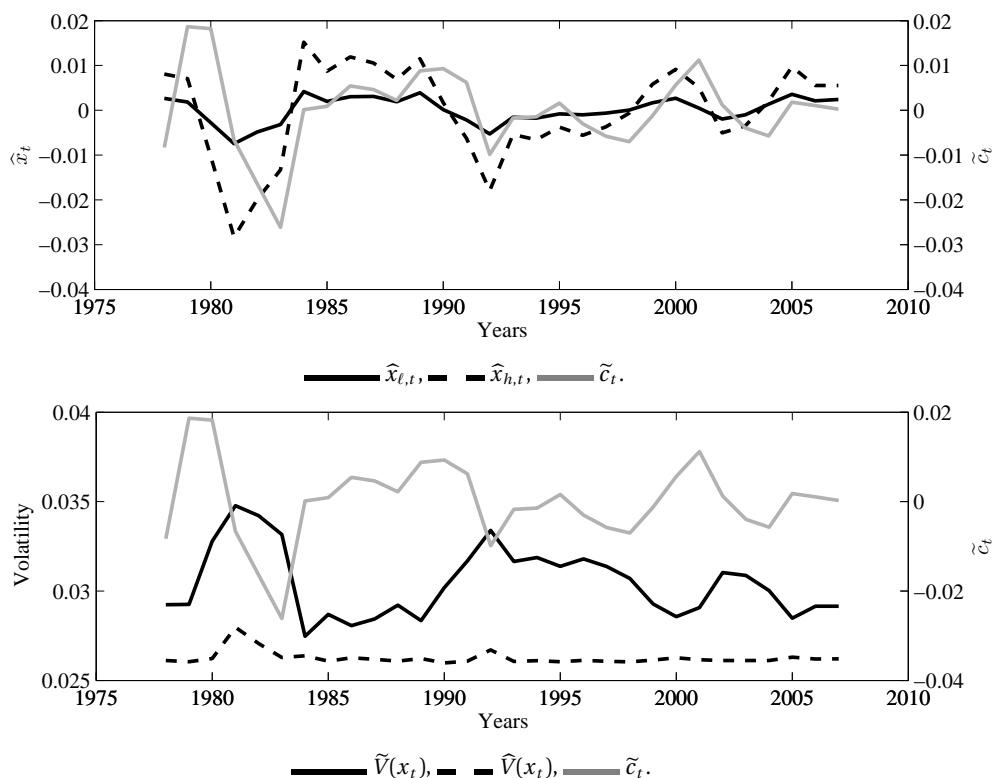


Figure 5: **Time-varying ambiguity.** The top panel shows the predicted latent variables for each Kalman filter in the two- ρ model ($\hat{x}_{l,t}$, $\hat{x}_{h,t}$, solid and dashed black lines). In the bottom panel the dashed line graphs the conditional variance of the latent state variable ($Var_t(x_t)$) as perceived by an ambiguity-neutral agent, and the solid line the conditional variance ($\tilde{V}(x_t)$) of the distorted posterior of an ambiguity averse agent. In both panels the gray line contains the the HP-filtered consumption growth. The distorted posterior corresponds to $\gamma = 2$ and α calibrated so that the risk-free rate was 1.85%.

Figure 5 contains two panels. The top panel shows how $\hat{x}_{h,t}$ and $\hat{x}_{l,t}$ have moved with time and business cycle (proxied by HP filtered log consumption) over the period 1978–2007. As was noted, it is movements in these state variables, especially their disagreement, which is the source of variation in uncertainty. The bottom panel depicts time-series of two ramifications of this uncertainty: the variance of the posterior (eq. 27) and the variance of the *distorted* posterior. The latter, evidently, greatly amplifies movements in the former, especially at downturns: instances of greater volatility in the distorted posterior arise when the distance between the two latent states is large *and* $\hat{x}_{h,t}$ falls below $\hat{x}_{l,t}$, as it does following a strong negative shock. Hence, the “as if” belief of the ambiguity averse agent exaggerates the volatility in the Bayesian belief in a way that makes it more pronouncedly counter-cyclical. Indeed, while $Var_t(x_t)$ has a correlation of -0.32 with HP-filtered consumption, the correlation is -0.47 with $\tilde{V}(x_t)$.

Figure 5 shows that in both 1982 and 1992 the distance between the two latent states is high *and* $\hat{x}_{h,t} < \hat{x}_{l,t}$, while in 1984 and 2005 $\hat{x}_{h,t} > \hat{x}_{l,t}$. In the years 1984 and 1992 the divergence between $\hat{x}_{h,t}$ and $\hat{x}_{l,t}$ was

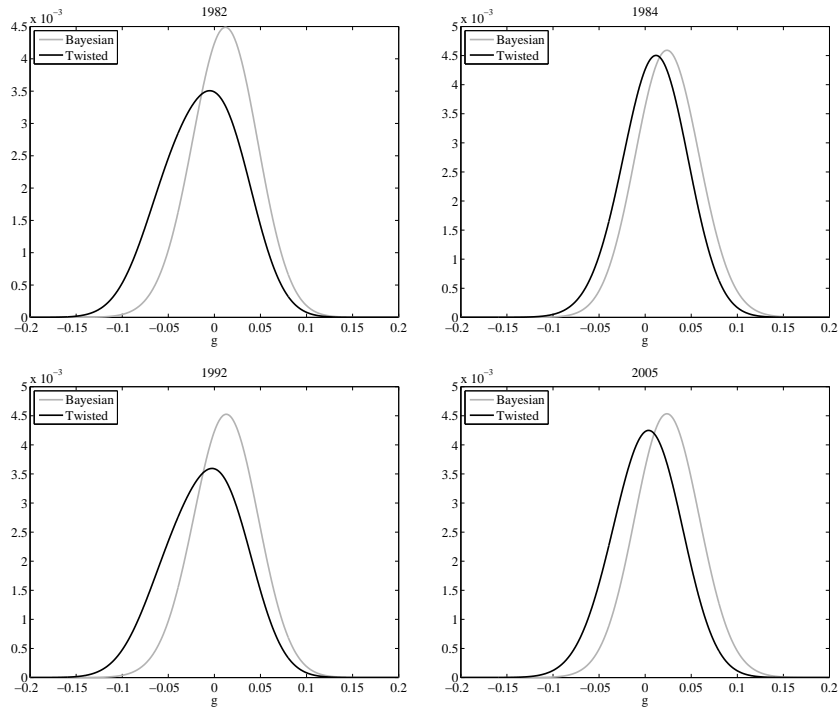


Figure 6: **Time-varying distortion:** These four panels contain plot beliefs about the latent state without ambiguity aversion (Bayesian) and with ambiguity aversion in the two- ρ model. The two years pictured on the left were bad years in the sense that the beliefs from the two models disagreed and the high persistence state was below the low persistence state. The two right years were good years in the sense that the state in the model with high persistence was above the latent state in the low persistence model.

similar, but it is only in 1992 that $\hat{x}_{h,t} < \hat{x}_{l,t}$. Though facing approximately the same forecast uncertainty in the two years, the ambiguity averse agent is more apprehensive about the uncertainty in the latter instance when the filtered state from the high persistence model is lower. Between the two instances, it is only in the latter case that the worse of the two forecasts come from the high persistence model, and hence if true would be expected to prevail far more in the future (affecting continuation value more severely). This leads to a larger twist over probabilities in the left tail leading to a prominent left skew and excess kurtosis, as can be seen when we compare the Bayesian and twisted predictive distributions in 1984 and 1992 in Figure 6. When the filtered beliefs from the model with $\rho = \rho_l$ are below those from the $\rho = \rho_h$ model, the worse case scenario (largely) coincides with the low persistence model being true. Such a truth is not that bad to face up to since the current adverse shock is expected to be short lived. Bear in mind the “as if” belief is a prospective probability distribution the agent wants to check her choice of portfolio is robust against. So in this case, it is an uncertainty she feels she has less of a need to guard against and the “as if” belief will not involve a substantive left tail. We can see this was the case in 1984 and 2005 where the twisted predictive distributions are simple shifts of the original posterior – as in the single ρ model – and do not exhibit a significant left skew. This asymmetry of the distorted

posterior in the order of filtered beliefs is the driving factor behind the counter-cyclical behavior of the equity premium. This may be further confirmed by comparing the graphs of the predicted equity premium in Figure 3 with the graphs in the two panels in Figure 5. The comparison shows counter-cyclical behavior is clearly associated with the crossings of the filtered state graphs corresponding to the high- and low-persistence models.

To summarize, pessimism and doubt are still what drive results in the two-rho case, like in the single-rho case. However, there are important differences. For one, even though the pessimism is still what largely fixes the level of the equity premium it is relatively muted in this case, since the agent's belief about the average persistence is more moderate, making the corresponding twisted predictive not as pessimistic. Two, the nature of doubt is quite different, and it plays a far more important role. Doubt is no longer characterized simply by the second moment; skewness and kurtosis come into play. Furthermore, it is dynamic, varying endogenously, thereby causing the equity premium to vary, in particular, over the business cycle.

4 Interpreting the magnitude of the ambiguity aversion

In Section 3 we predicted the effect of ambiguity on rates of return by calibrating the value of the ambiguity aversion coefficient α such that the predicted risk-free rate matched the sample average in the data between 1978 and 2007. The results of that exercise demonstrate that the calibration is consistent with observed aggregate market behavior across the two moments of *both* rates of return and the equity premium. But are the calibrated values in the realm of reasonable behavior at an individual level? Here we discuss some ways of understanding whether the levels of ambiguity aversion invoked imply individual behavior that may be regarded as plausible and consistent with behavior observed in experiments.

Consider, as a thought experiment, an individual faced with an uncertain consumption prospect similar to that faced by the agent in our two versions of the Lucas economy, the single- ρ or the two- ρ model. That is, a consumption prospect $C \exp(g)$, where $g \sim F(g; x)$ and $x \sim F(x)$. In the case of the single- ρ model, we have

$$F(g; x) \equiv N(\rho x + \bar{g}, \sigma_x^2 + \sigma_g^2), \quad F(x) \equiv N(\hat{x}, \hat{P}), \quad (28)$$

and in the case of the two- ρ we have,

$$F(g; x_i) \equiv N(\rho x_i + \bar{g}, \sigma_{x_i}^2 + \sigma_{g_i}^2), \quad i = h, l, \quad F(x, \eta) \equiv \begin{cases} N(\hat{x}_l, \hat{P}_l) & \text{with probability } \eta \\ N(\hat{x}_h, \hat{P}_h) & \text{with probability } 1 - \eta \end{cases}. \quad (29)$$

If the individual were an expected utility maximizer, the *risk premium* she would be willing to pay for such a prospect is given by,

$$R(u) \equiv u^{-1} \left(\iint u(C \exp(g)) dF(g; x) dF(x) \right) - \left(\iint C \exp(g) dF(g; x) dF(x) \right). \quad (30)$$

When $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$, we denote the risk premium $R(\gamma)$. Normalizing the premium by expressing it as a percentage of the expected value of consumption,

$$\frac{r(\gamma)}{100} = \frac{R(\gamma)}{\iint C \exp(g) dF(g; x) dF(x)} \quad (31)$$

allows the premium to be interpreted in units of consumption expected in the following period. The risk premium is the price, that the agent is willing to pay to swap the risky consumption for risk-less consumption, i.e., the price for removing the risk.

Next suppose the agent has smooth ambiguity preferences, and consider what she is willing to pay to swap the ambiguous and risky growth with another which removes the ambiguity but not the risk. That is, the alternative consumption prospect offers with certainty that the consumption growth is generated by the *reduced* distribution $\int F(g;x) dF(x)$. We call this the *ambiguity premium* – the price the individual is prepared to pay to remove the ambiguity:

$$A(\phi; u) \equiv u^{-1} \left(\phi^{-1} \left(\int \phi \left(\int u(C \exp(g)) dF(g;x) \right) dF(x) \right) \right) - u^{-1} \left(\iint u(C \exp(g)) dF(g;x) dF(x) \right). \quad (32)$$

The first term on the r.h.s. of 32 is the certainty equivalent of the prospect while the second term is the certainty equivalent of the alternative prospect, a lottery generated by the *reduced* distribution $\int F(g;x) dF(x)$. Notice, under ambiguity neutrality the two terms would be identical. When $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$, $\phi(x) = -\exp(-\alpha x)/\alpha$, we denote the ambiguity premium $A(\alpha, \gamma)$ and normalize by expected consumption, to get,

$$\frac{a(\alpha; \gamma)}{100} = \frac{A(\alpha; \gamma)}{\iint C \exp(g) dF(g;x) dF(x)}. \quad (33)$$

It is useful to recall that the magnitude of α cannot be meaningfully interpreted independent of the value of γ . First, consider the relative magnitude of the ambiguity premium and the risk premium in the thought experiments corresponding to the single and two- ρ models, using the consumption prospects in eqs. (28) and (29), with the tuple $(\alpha; \gamma)$ set to (15.6; 2) and (18.9; 2), the calibrated values of α given $\gamma = 2$ in the respective models. The corresponding (normalized) ambiguity premia, $a(\alpha; \gamma)$, are respectively, 0.28 and 0.32 percent, and the risk premium $r(\gamma)$ is 0.1 percent.¹⁰ For the *same* prospects, $6 \leq \gamma \leq 7$ would generate a risk premium of about 0.3%. That is, the ambiguity premium for these prospects is about the same as the risk premium that an expected utility maximizer with $\gamma \in [6, 7]$ will be willing to pay. It has been argued that a $\gamma \leq 10$ is plausible (see, e.g., Mehra and Prescott (1985), Kocherlakota (1996)). The ambiguity premium and even the total uncertainty premium (i.e., the risk premium plus the ambiguity premium) corresponding to the values of $(\alpha; \gamma)$ invoked in our estimates is less than the risk premium generated by levels of relative risk aversion ($\gamma \approx 10$) deemed plausible. Figure 7 contains plots of the ambiguity premium and risk premium for different levels of α and γ .

The expected growth of consumption (given the consumption prospect being considered) is around 2%. An ambiguity premium of about 0.3%, which is obtained when α is calibrated with $\gamma = 2$ in the model, implies

¹⁰For computing the (risk and ambiguity) premia in the thought experiment corresponding to the single- ρ model we consider the consumption prospect obtained by setting the values the tuple $(C_t; \hat{x}_t)$ equal to the corresponding average values observed in the data. Analogously, for the two- ρ model, the values in $(C_t; \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t)$ are set equal to the sample average values.

that the individual is willing to give up around $1/7$ of expected consumption *growth* to avoid the ambiguity. The magnitude of this trade-off gives a simple behavioral implication of the magnitude of ambiguity aversion we use in the model, and is a second way to think about its plausibility.

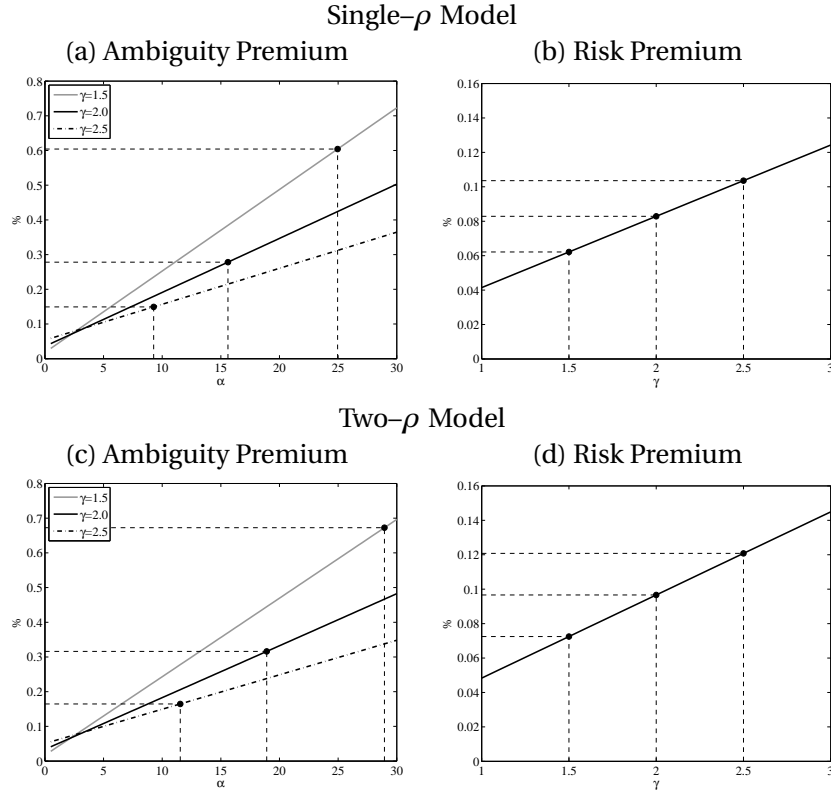


Figure 7: **Risk and Ambiguity Premia:** The left panels contains the (normalized) ambiguity premium function, $a(\alpha; \gamma)$, for the thought experiment described in section 4 for three different levels of risk aversion, $\gamma = 1.5, 2$ and 2.5 . The right panel shows the (normalized) risk premium, $r(\gamma)$, in the same thought experiment. The marked dots on the graphs indicate values when α is calibrated, given relevant value of γ , to obtain $r^f = 1.85\%$.

Thirdly, as a consistency check, we compare the ambiguity premia obtained in our thought experiment with estimates of ambiguity premia observed in lab experiments. Camerer and Weber (1992, p.334) report that an ambiguity premium of about 20% of the *net* expected value (i.e., of the pure “ticket” value of the uncertain prospect, net of the subject’s initial income or wealth) of the prospect has been found repeatedly in the experimental literature.¹¹ In our thought experiment it is the *growth* in consumption, not the full final consumption, that corresponds to the pure “ticket” value of prizes in the lab experiments. As noted, the ambiguity premia in the thought experiment are about $1/7$ of expected growth in consumption, i.e., roughly 15%, and hence slightly

¹¹Typically, the lab experiments implemented a variation of the famous Ellsberg two-urn example, where one urn has a known mixture of two colors of balls while the other urn an unknown mixture and the subject is offered bets on a draw from each urn. No attempt is made to estimate an ambiguity aversion parameter in these experiments. We are simply comparing the ambiguity premium estimated in these lab experiments with the ambiguity premium we obtain when we invoke the calibrated values of $(\alpha; \gamma)$ in our thought experiment.

lower than the finding in the lab experiments. In this sense, we may conclude that the calibrated values of (α, γ) generate individual behavior (i.e., ambiguity premia in our thought experiment) broadly consistent with individual behavior observed in lab experiments.¹²

5 Concluding remarks

We have constructed a quantitative assessment of the effect of ambiguity (about macroeconomic risk) on asset returns. The ambiguity we focused on arose from the conditional uncertainty about the transitory component of consumption and dividend growth. This uncertainty was quantified by using a variation of Bansal and Yaron's Long Run Risk model where, due to the working of the business cycle, the conditional mean growth rate is subject to temporary shocks with uncertain persistence. This uncertainty in persistence leads to time-varying ambiguity for the agent. Ambiguity, so quantified, is combined through a simple modification of the Lucas tree model with, (1) a degree of ambiguity aversion that is, arguably, plausible and consistent with experimental observation and, (2) a degree of risk aversion that would be considered moderate by most. We find the ambiguity conditional on the actual, observed, history predicts rates of return that match the salient characteristics of the data on rates of return, their level, volatility *and* movement – in particular the counter-cyclicity of the market premium. What drives results is that ambiguity averse agents are more attracted to the risk-free asset, relative to the risky asset, because it enables them to maintain a portfolio that is more robust to ambiguity, which is found to increase endogenously in downturns. On the basis of all that we find, it seems fair to conclude that ambiguity and robustness concerns do matter, in quantitatively significant ways, for asset pricing.

The work suggests some directions for future research, especially in finance and macroeconomics. For instance, the model could be used to investigate predictions for term structure of (real) interest rates and related issues. More broadly, it would be of interest to incorporate the model of quantitative ambiguity about macroeconomic risk developed here into standard DSGE frameworks to investigate the effect of such uncertainty, varying endogenously over the business cycle, on macroeconomic interaction. This method of incorporating time varying ambiguity, is parsimonious, empirically grounded, neither dependent on exogenous, difficult to calibrate specifications of stochastic volatility, nor on ad hoc survey data. It works using simple analytic linear filtering and is, above all, intuitive, relying on a transparent interplay between uncertainty about the true extent of a shock and uncertainty about its longevity.

¹²Epstein (2010) suggests two Ellsberg (1961)-style thought experiments and argues that they pose difficulties for the smooth ambiguity model. In particular, on the basis of his analysis of the first thought experiment, he claims that efforts to calibrate an individual's ϕ in a context of interest (e.g., financial markets), by examining the behavior of that individual in another environment (e.g., real or hypothetical Ellsberg experiments), have no justification. Klibanoff, Marinacci, and Mukerji (2011) revisit these thought experiments and show that Epstein's conclusions arise because his analysis does not use a state space complete enough to allow the formal incorporation of the key information defining the experiments. The interested reader is referred to the two papers for details.

Appendix

A An analytical approximation for rates of return in the single- ρ model

This section develops an analytical approximation to the equilibrium rates of return in the single- ρ model. The crucial assumption on which the following second order approximation analysis depends is that $E_{\tilde{\mu}_t}$ operates with respect to some normal distribution $N(\tilde{x}_t, \hat{P})$. As the numbers (reporting skewness and excess kurtosis) in Table 4 generated using the accurate numerical approximation demonstrate, Normality is a fairly accurate description in the case of the single- ρ model.

Single- ρ Model				
	x_t		$g_{c,t}$	
	E	σ	E	σ
Rat. Exp.	--	--	0.019	0.028
Bayesian	0.000	0.026	0.019	0.036
Twisted	-0.040	0.027	-0.015	0.036
	sk	κ	sk	κ
Rat. Exp.	--	--	0.000	0.000
Bayesian	0.000	0.000	0.000	0.000
Twisted	0.002	0.000	0.0007	0.000

Two- ρ Model				
	x_t		$g_{c,t}$	
	E	σ	E	σ
Bayesian	0.000	0.026	0.019	0.034
Twisted	-0.031	0.030	-0.002	0.039
	sk	κ	sk	κ
Bayesian	-0.000	0.008	0.000	0.028
Twisted	-0.063	-0.181	-0.119	-0.068

Table 4: **Conditional moments of distributions.** In each case, $\gamma = 2$ and α was set such that the model generates an average risk-free rate of 1.85%. C_t , $\tilde{x}_{\ell,t}$, $\tilde{x}_{h,t}$ and η_t are set equal to their mean in the data. sk and κ denote skewness and excess Kurtosis (relative to a Gaussian distribution), respectively. The latent state variable is known to a rational expectations agent and so the conditional distribution is degenerate.

Assumption 1 *The density $\tilde{f}(x_t)$ of the distorted posterior, $\tilde{\mu}_t$, defined in (21) satisfies $\tilde{f}(x_t) \propto N(\tilde{x}_t, \hat{P})$.*

This is equivalent to assuming that eq. (21) is exactly a normal density with the same variance as the Bayesian posterior \hat{P} but with a different mean (\tilde{x}_t instead of \hat{x}_t). Let $E_t \equiv E_{\hat{x}_t} E_{x_t}$; $\tilde{E}_t \equiv E_{\tilde{\mu}_t} E_{x_t} \equiv E_{\tilde{x}_t} E_{x_t}$. It is useful to recall, if x_t is normally distributed, then for any $k \in \mathbb{R}$,

$$E_t [\exp(kx_t)] = \exp \left(kE_t x_t + \frac{k^2}{2} Var_t(x_t) \right)$$

Also, $\widetilde{Var}_t(x_t) \equiv Var_{\tilde{\mu}_t}(x_t)$ and $Var_t(x_t) = Var_{\mu_t}(x_t) = \widehat{P}$ and all ε terms have expectation zero under both \tilde{E}_t and E_t since the terms have expectation zero conditional on x_t .

The first Euler equation relating to the risk-free asset may be rewritten as follows:

$$\begin{aligned} 1 &= \beta R_t^f \tilde{E}_t \left[\exp \left(-\gamma \bar{g} - \gamma \rho x_t - \gamma \sigma_x \varepsilon_{x,t+1} - \gamma \sigma_g \varepsilon_{g,t+1} \right) \right] \\ &= \beta R_t^f \exp \left(-\gamma \bar{g} - \gamma \rho \tilde{x}_t + \frac{\gamma^2}{2} \left(\sigma_x^2 + \sigma_g^2 \right) + \frac{\gamma^2 \rho^2}{2} \widetilde{Var}_t(x_t) \right). \end{aligned}$$

Taking logs and rearranging terms we obtain an approximate solution for the risk-free rate of return:

$$r_t^f = -\ln \beta + \gamma \bar{g} + \gamma \rho \tilde{x}_t - \frac{\gamma^2}{2} \left(\sigma_x^2 + \sigma_g^2 + \rho^2 \widetilde{Var}_t(x_t) \right). \quad (34)$$

The second Euler equation relating to the risky asset may then be written as:

$$\tilde{E}_t \exp \left[\ln \beta + \ln \left(\frac{p_{t+1} + D_{t+1}}{p_t} \right) - \gamma \ln \left(\frac{C_{t+1}}{C_t} \right) \right] = 1 \quad (35)$$

We adopt the following approximation (to the risky rate of return), proposed in Campbell and Shiller (1988).

Assumption 2 *The risky rate of return is approximated as*

$$r_t \equiv \ln \left(\frac{p_{t+1} + D_{t+1}}{p_t} \right) \simeq \kappa_0 + \kappa_1 z_{t+1} - z_t + d_{t+1} \quad (36)$$

where $z_t = \ln \left(\frac{p_t}{D_t} \right)$ and κ_0 and κ_1 are approximating constants.

Next, we conjecture that the log price-dividend ratio is given by

$$z_t = A_0 + A_1 \tilde{x}_t. \quad (37)$$

Our final assumption is that the mean of the distorted conditional distribution is an affine function of the mean of the (contemporaneous) undistorted, Bayesian conditional distribution. Figure 8 shows that linearity is a good approximation by plotting \tilde{x}_t against \hat{x}_t (as obtained in numerical solutions)

Assumption 3 $\tilde{x}_t = \delta_0 + \delta_1 \hat{x}_t$ for $t = 1, 2, \dots$, where δ_0, δ_1 are approximating constants.

Claim 1 *Assumption 3 implies that the time t conditional expectation of \tilde{x}_{t+1} satisfies: (i) $E_t \tilde{x}_{t+1} = \tilde{E}_t \tilde{x}_{t+1} = \delta_0 (1 - \rho) + \rho \tilde{x}_t$, and (ii) $E_t \tilde{x}_{t+1}^2 = \tilde{E}_t \tilde{x}_{t+1}^2$.*

The proof of these claims are straight forward and so omitted.

Hence, we obtain a second order approximation of the second Euler equation as follows:

$$1 = \tilde{E}_t \exp \left[\log(\beta) + \kappa_0 + \kappa_1 z_{t+1} - z_t + d_{t+1} - \gamma g_{t+1} \right]$$

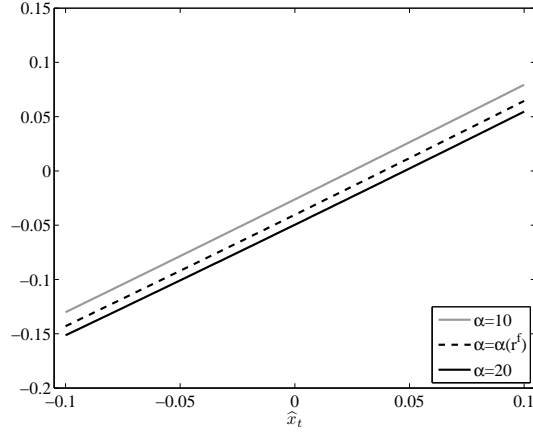


Figure 8: $\tilde{x}_t = E_{\tilde{\mu}_t}(x_t)$ plotted against \hat{x}_t . The level of consumption is set to the average value between 1978 and 2007. In each case, $\gamma = 2.00$.

Plugging the guess for z_t and using the processes of growth rates, we obtain

$$\begin{aligned}
1 &= \tilde{E}_t \exp \left[\ln \beta + \bar{g}_d - \gamma \bar{g} + \kappa_0 + (\kappa_1 - 1)A_0 + \kappa_1 A_1 \tilde{x}_{t+1} - A_1 \tilde{x}_t + (\psi - \gamma)x_{t+1} + \sigma_d \varepsilon_{d,t+1} - \gamma \sigma_g \varepsilon_{g,t+1} \right] \\
&= \tilde{E}_t \exp \left[\ln \beta + \bar{g}_d - \gamma \bar{g} + \kappa_0 + (\kappa_1 - 1)A_0 + \kappa_1 A_1 \tilde{x}_{t+1} - A_1 \tilde{x}_t + (\psi - \gamma)\rho x_t + (\psi - \gamma)\sigma_x \varepsilon_{x,t+1} \right. \\
&\quad \left. + \sigma_d \varepsilon_{d,t+1} - \gamma \sigma_g \varepsilon_{g,t+1} \right] \\
&= \tilde{E}_t \exp \left[\ln \beta + \bar{g}_d - \gamma \bar{g} + \kappa_0 + (\kappa_1 - 1)A_0 + \kappa_1 A_1 \tilde{x}_{t+1} - A_1 \tilde{x}_t + (\psi - \gamma)\rho x_t + (\psi - \gamma)\sigma_x \varepsilon_{x,t+1} \right. \\
&\quad \left. + \sigma_d \varepsilon_{d,t+1} - \gamma \sigma_g \varepsilon_{g,t+1} \right]
\end{aligned}$$

Using Assumptions 2 and 3 (and resulting Claim 1) and results on log-normal distributions

$$\begin{aligned}
1 &= \exp \left[\ln \beta + \bar{g}_d - \gamma \bar{g} + \kappa_0 + (\kappa_1 - 1)A_0 + \kappa_1 A_1 \delta_0 (1 - \rho) + ((\kappa_1 \rho - 1)A_1 + (\psi - \gamma)\rho) \tilde{x}_t \right. \\
&\quad \left. + (\psi - \gamma)^2 \frac{\sigma_x^2}{2} + \frac{\sigma_d^2}{2} + \gamma^2 \frac{\sigma_g^2}{2} + (\kappa_1 A_1)^2 \frac{\widetilde{Var}_t(\tilde{x}_{t+1})}{2} + (\psi - \gamma)^2 \rho^2 \frac{\widetilde{Var}_t(x_t)}{2} \right]
\end{aligned}$$

Taking logs, we finally obtain

$$\begin{aligned}
0 &= \ln \beta + \bar{g}_d - \gamma \bar{g} + \kappa_0 + (\kappa_1 - 1)A_0 + \kappa_1 A_1 \alpha (1 - \rho) + \kappa_1 A_1 \delta_0 (1 - \rho) + ((\kappa_1 \rho - 1)A_1 + (\psi - \gamma)\rho) \tilde{x}_t \\
&\quad + (\psi - \gamma)^2 \frac{\sigma_x^2}{2} + \frac{\sigma_d^2}{2} + \gamma^2 \frac{\sigma_g^2}{2} + (\kappa_1 A_1)^2 \frac{\widetilde{Var}_t(\tilde{x}_{t+1})}{2} + (\psi - \gamma)^2 \rho^2 \frac{\widetilde{Var}_t(x_t)}{2}
\end{aligned}$$

We solve for A_0 and A_1 using the method of undetermined coefficients to obtain,

$$A_0 = \frac{1}{1 - \kappa_1} \left(\ln \beta + \bar{g}_d - \gamma \bar{g} + \kappa_0 + \kappa_1 A_1 \delta_0 (1 - \rho) + \right.$$

$$(\psi - \gamma)^2 \frac{\sigma_x^2}{2} + \frac{\sigma_d^2}{2} + \gamma^2 \frac{\sigma_g^2}{2} + (\kappa_1 A_1)^2 \frac{\widehat{Var}_t(\tilde{x}_{t+1})}{2} + (\psi - \gamma)^2 \rho^2 \frac{\widehat{Var}_t(x_t)}{2}$$

$$A_1 = \frac{\rho(\psi - \gamma)}{1 - \kappa_1 \rho}$$

Hence, the expected risky rate (per Assumption 2), when expectations are taken w.r.t. μ_t is given by:

$$E_t R_{t+1} = E_t \left[\exp(\kappa_0 + \kappa_1 z_{t+1} - z_t + d_{t+1}) \right]$$

Plugging the guess for z_t and using the processes for the growth rates,

$$E_t R_{t+1} = E_t \left[\exp(\kappa_0 + A_0(\kappa_1 - 1) + \kappa_1 A_1 \tilde{x}_{t+1} - A_1 \tilde{x}_t + \bar{g}_d + \psi \rho x_t + \psi \sigma_x \varepsilon_{x,t+1} + \sigma_d \varepsilon_{d,t+1}) \right].$$

Then using Claim 1 and taking logs, we obtain

$$E_t r_{t+1} = \kappa_0 + A_0(\kappa_1 - 1) + \kappa_1 A_1 \delta_0 (1 - \rho) + (\kappa_1 \rho - 1) A_1 \tilde{x}_t + \psi \rho \hat{x}_t + \bar{g}_d$$

$$+ (\psi \rho)^2 \frac{\hat{P}}{2} + \psi^2 \frac{\sigma_x^2}{2} + \frac{\sigma_d^2}{2} + (\kappa_1 A_1)^2 \frac{Var_t(\tilde{x}_{t+1})}{2}.$$

Finally, using the solution for A_0 obtained above and Claim 1 (ii), we get the expression for the first moment of the risky rate:

$$E_t r_{t+1} = -\ln \beta + \gamma \bar{g} - \frac{\gamma^2}{2} \sigma_g^2 + \gamma \left(\frac{2\psi - \gamma}{2} \right) \sigma_x^2 + \frac{\psi^2 \rho^2}{2} \hat{P} - \frac{(\rho(\psi - \gamma))^2}{2} \widehat{Var}_t(x_t) - \rho(\psi - \gamma) \tilde{x}_t + \psi \rho \hat{x}_t. \quad (38)$$

Hence the overall effect of α on $E_t r_{t+1}$ is a combination of a positive effect via the $[-\rho(\psi - \gamma) \tilde{x}_t]$ term but negative effect via the term $[-\frac{(\rho(\psi - \gamma))^2}{2} \widehat{Var}_t(x_t)]$. However, Table 4 shows that ambiguity aversion has a far more significant effect on the mean of the distorted conditional distribution than on its variance in the case of the single- ρ model.

The equity premium is then computed as

$$E_t r_{t+1} - r_t^f = \rho \psi (\hat{x}_t - \tilde{x}_t) + \psi \gamma \sigma_x^2 + \frac{\psi^2 \rho^2}{2} \hat{P} + \frac{\rho^2 \psi (2\gamma - \psi)}{2} \widehat{Var}_t(x_t) \quad (39)$$

Clearly, the equity premium is increasing in the parameters ρ , γ , and ψ . Notice, α has a unequivocal positive influence for values of $\gamma > 1.5$ when we set $\psi = 3$.

We need values of the approximating constants, κ_0 and κ_1 , to compute the log price-dividend ratio. Beeler and Campbell (2009) obtain the constants as follows

$$\bar{z} = \frac{\sum z_t}{N}$$

$$\kappa_1 = \frac{\exp \bar{z}}{1 + \exp \bar{z}}$$

$$\kappa_0 = \ln(1 + \exp \bar{z}) - \kappa_1 \bar{z}.$$

which we compute from the same data used throughout. Table 5 compares this approximation to the solution from the numerical results presented in the main body of the paper. These broadly agree. The comparative statics in Figure 9 were derived using the analytical approximation, and are similar to those in the top panel of figure 2 which were computed using numerical methods.

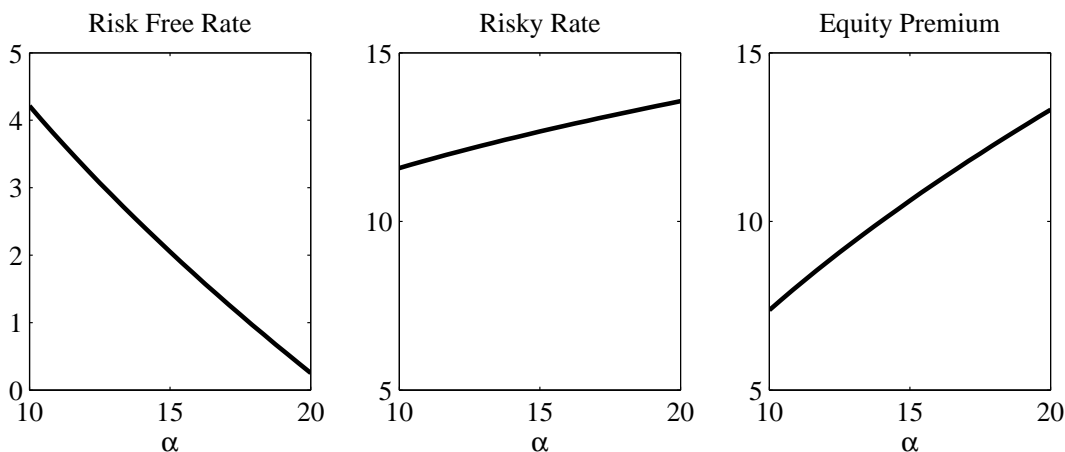


Figure 9: Comparative static, variations in α , computed using the analytical approximation. The level of consumption is set to the average value between 1978 and 2007. In each case, $\gamma = 2.00$.

		$E(r^f)$		$E(r)$		$E(r - r^f)$	
γ	α	Numerical	Analytical	Numerical	Analytical	Numerical	Analytical
2.00	15.6	1.85	1.81	10.3	12.7	8.47	10.9

Table 5: **Moments:** Numerical approximation versus 2nd order analytical approximation ($\rho = 0.85$)

B Details of the model where agents beliefs are given by the $2\text{-}\rho$ specification

B.1 Beliefs and the direct value function:

The agent believes that the stochastic evolution of the economy follows a persistent latent state process given by a Bansal and Yaron type specification with either a low persistence (ρ_l) or a high persistence (ρ_h), but does not know for sure which. That is, she believes either of the models described in equation (26) represent the true data generating process. Define $\hat{x}_{i,t} \equiv E[x_{i,t} | g_{i,1}, \dots, g_{i,t}, d_{i,1}, \dots, d_{i,t}]$, $i = l, h$, to denote the filtered x at time t conditional on the observed history of growth rates (of consumption and dividend), if the history were interpreted and beliefs updated using a Kalman filter which takes the model with $\rho = \rho_i$ as the data generating process. At any node on the growth path, at a time t , the agent's beliefs may be summarized by the tuple $(\hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t)$, where the first two elements show the beliefs about the latent state variable conditional on alternative assumptions about the true data generating process (low or high persistence, respectively) while the last element shows the posterior belief that the true data generating process is the low persistence model. We denote by $\hat{x}_{j,t+1}^{(i)}$, $i = l, h$, $j = l, h$, the agent's forecast for the (one period ahead) update to her belief about the filtered x if the growth outcome next period (along with the previous history) were interpreted using a Kalman filter which takes the model with $\rho = \rho_j$ as the data generating process, when the data is actually generated by

the i persistence model. The direct value function obtains as follows:¹³

$$\begin{aligned}
V(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) &= (1 - \beta) \frac{C_t^{1-\gamma}}{1-\gamma} \\
&\quad - \frac{\beta}{\alpha} \ln \left[\eta_t \left\{ \int_{-\infty}^{\infty} \exp \left(-\alpha \iiint_{-\infty}^{\infty} V \left(C_t \exp(g_{l,t+1}), \hat{x}_{l,t+1}^{(l)}(\vec{\epsilon}_{l,t+1}), \right. \right. \right. \right. \\
&\quad \left. \left. \left. \hat{x}_{h,t+1}^{(l)}(\vec{\epsilon}_{l,t+1}), \eta_{t+1}^{(l)}(\vec{\epsilon}_{l,t+1}) \right) dF(\vec{\epsilon}_{l,t+1}) \right) dF(x_{l,t}) \right\} \\
&\quad + (1 - \eta_t) \left\{ \int_{-\infty}^{\infty} \exp \left(-\alpha \iiint_{-\infty}^{\infty} V \left(C_t \exp(g_{h,t+1}), \hat{x}_{l,t+1}^{(h)}(\vec{\epsilon}_{h,t+1}), \right. \right. \right. \\
&\quad \left. \left. \left. \hat{x}_{h,t+1}^{(h)}(\vec{\epsilon}_{h,t+1}), \eta_{t+1}^{(h)}(\vec{\epsilon}_{h,t+1}) \right) dF(\vec{\epsilon}_{h,t+1}) \right) dF(x_{h,t}) \right\} \Big]
\end{aligned} \tag{40}$$

where $\vec{\epsilon}_{l,t+1} = [\epsilon_{x_{l,t+1}} \epsilon_{d_{l,t+1}} \epsilon_{g_{l,t+1}}]$ is a 3 by 1 vector of standard normal shocks (and so is $\vec{\epsilon}_{h,t+1}$) and η_t is the posterior probability at time t that the model with ρ_l is the data generating process. $F(\vec{\epsilon}_{l,t+1})$ and $F(\vec{\epsilon}_{h,t+1})$ are both trivariate independent standard normal distributions. The updates for $\hat{x}_{j,t+1}^{(i)}$ are obtained as follows:

$$\begin{aligned}
\hat{x}_{l,t+1}^{(l)}(\vec{\epsilon}_{l,t+1}) &= \rho_l \hat{x}_{l,t} + K_l v_{l,t+1}^{(l)} \\
\hat{x}_{h,t+1}^{(l)}(\vec{\epsilon}_{l,t+1}) &= \rho_h \hat{x}_{h,t} + K_h v_{h,t+1}^{(l)} \\
\hat{x}_{l,t+1}^{(h)}(\vec{\epsilon}_{h,t+1}) &= \rho_l \hat{x}_{l,t} + K_l v_{l,t+1}^{(h)} \\
\hat{x}_{h,t+1}^{(h)}(\vec{\epsilon}_{h,t+1}) &= \rho_h \hat{x}_{h,t} + K_h v_{h,t+1}^{(h)}
\end{aligned}$$

where $v_{j,t+1}^{(i)}$, $(i) = (l)$ or $(i) = (h)$ and $j = l, h$, denote the "surprises". For example, when the DGP is $(i) = (l)$ and the filter uses ρ_j , $j = h$, the surprise is defined

$$v_{h,t+1}^{(l)} = \begin{bmatrix} g_{l,t+1} - \bar{g} - \rho_h \hat{x}_{h,t} \\ d_{l,t+1} - \bar{d} - \psi \rho_h \hat{x}_{h,t} \end{bmatrix} = \begin{bmatrix} \bar{g} - \bar{g} + \rho_l x_{l,t} - \rho_h \hat{x}_{h,t} + \sigma_{x_l} \epsilon_{x_{l,t+1}} + \sigma_{g_l} \epsilon_{g_{l,t+1}} \\ \bar{d} - \bar{d} + \psi \rho_l x_{l,t} - \psi \rho_h \hat{x}_{h,t} + \psi \sigma_{x_l} \epsilon_{x_{l,t+1}} + \sigma_{d_l} \epsilon_{d_{l,t+1}} \end{bmatrix}.$$

The Kalman gain parameters, K_i , $i = l, h$, depending on whether low or high persistence model is assumed to be the true model, respectively, are

$$K_i = \rho_i \hat{P}_i [1 \ \psi] \hat{F}_i^{-1}, \quad \text{where} \quad \hat{F}_i = \begin{bmatrix} \hat{P}_i + \sigma_{g_i}^2 & \psi \hat{P}_i \\ \psi \hat{P}_i & \psi \hat{P}_i + \sigma_{d_i}^2 \end{bmatrix}$$

Finally, \hat{P}_i , $i = l, h$, is defined as the solution to

$$\hat{P}_i = \rho_i^2 \hat{P}_i - \rho_i^2 \hat{P}_i^2 [1 \ \psi] \hat{F}_i^{-1} [1 \ \psi]' + \sigma_{x_i}^2$$

¹³Note that the utility function is pre-multiplied by $1 - \beta$ in order to avoid the value function takes on very high values that would prevent numerical stability of the algorithm.

The Bayes update of η_t is obtained as follows :

$$\eta_{t+1}^{(l)}(\vec{\epsilon}_{l,t+1}) = \frac{\eta_t L(v_{l,t+1}^{(l)}, \hat{F}_l)}{\eta_t L(v_{l,t+1}^{(l)}, \hat{F}_l) + (1 - \eta_t) L(v_{h,t+1}^{(l)}, \hat{F}_h)}$$

$$\eta_{t+1}^{(h)}(\vec{\epsilon}_{h,t+1}) = \frac{\eta_t L(v_{l,t+1}^{(h)}, \hat{F}_l)}{\eta_t L(v_{l,t+1}^{(h)}, \hat{F}_l) + (1 - \eta_t) L(v_{h,t+1}^{(h)}, \hat{F}_h)}$$

where the likelihood is

$$L(v_{j,t+1}^{(i)}, \hat{F}_j) = \frac{1}{2\pi|\hat{F}_j|} \exp\left(-\frac{(v_{j,t+1}^{(i)})' \hat{F}_j^{-1} v_{j,t+1}^{(i)}}{2}\right) \quad \text{where } i = l, h \text{ and } j = l, h.$$

B.2 The rates of return

In the two- ρ model the risky rate of return is a function of four state variables, $C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t$, just like V and ξ_t .

In the sequel, it should be clear that variables in $t + 1$ are evaluated using the relevant stochastic components.

Let $C_{i,t+1} = C_t \exp(g_{i,t+1})$, $i = l, h$. The risk rate, R_t , will satisfy:

$$\begin{aligned} & \beta \eta_t \int_{-\infty}^{\infty} \xi_t^{(l)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) \left(\iiint_{-\infty}^{\infty} R_t(C_{l,t+1}, \hat{x}_{l,t+1}^{(l)}, \hat{x}_{h,t+1}^{(l)}, \eta_{t+1}^{(l)}) \times \right. \\ & \quad \left. (u'(\exp(g_{l,t+1}))) dF(\vec{\epsilon}_{l,t+1}) \right) dF(x_{l,t}) \\ & + \beta (1 - \eta_t) \int_{-\infty}^{\infty} \xi_t^{(h)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) \left(\iiint_{-\infty}^{\infty} R_t(C_{h,t+1}, \hat{x}_{l,t+1}^{(h)}, \hat{x}_{h,t+1}^{(h)}, \eta_{t+1}^{(h)}) \times \right. \\ & \quad \left. (u'(\exp(g_{h,t+1}))) dF(\vec{\epsilon}_{h,t+1}) \right) dF(x_{h,t}) = 1 \end{aligned}$$

where,

$$\xi_t^{(l)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) = \frac{\phi' \left(\iiint_{-\infty}^{\infty} V(C_{l,t+1}, \hat{x}_{l,t+1}^{(l)}, \hat{x}_{h,t+1}^{(l)}, \eta_{t+1}^{(l)}) dF(\vec{\epsilon}_{l,t+1}) \right)}{\Psi}$$

and

$$\xi_t^{(h)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) = \frac{\phi' \left(\iiint_{-\infty}^{\infty} V(C_{h,t+1}, \hat{x}_{l,t+1}^{(h)}, \hat{x}_{h,t+1}^{(h)}, \eta_{t+1}^{(h)}) dF(\vec{\epsilon}_{h,t+1}) \right)}{\Psi}$$

with

$$\begin{aligned} \Psi &= \eta_t \int_{-\infty}^{\infty} \phi' \left(\iiint_{-\infty}^{\infty} V(C_{l,t+1}, \hat{x}_{l,t+1}^{(l)}, \hat{x}_{h,t+1}^{(l)}, \eta_{t+1}^{(l)}) dF(\vec{\epsilon}_{l,t+1}) \right) dF(x_{l,t}) \\ & + (1 - \eta_t) \int_{-\infty}^{\infty} \phi' \left(\iiint_{-\infty}^{\infty} V(C_{h,t+1}, \hat{x}_{l,t+1}^{(h)}, \hat{x}_{h,t+1}^{(h)}, \eta_{t+1}^{(h)}) dF(\vec{\epsilon}_{h,t+1}) \right) dF(x_{h,t}) \end{aligned}$$

Then, we have

$$E_t R_t = \eta_t \iiint_{-\infty}^{\infty} R_t(C_{l,t+1}, \hat{x}_{l,t+1}^{(l)}, \hat{x}_{h,t+1}^{(l)}, \eta_{t+1}^{(l)}) dF(\vec{\epsilon}_{l,t+1}) dF(x_{l,t}) \\ + (1 - \eta_t) \iiint_{-\infty}^{\infty} R_t(C_{h,t+1}, \hat{x}_{l,t+1}^{(h)}, \hat{x}_{h,t+1}^{(h)}, \eta_{t+1}^{(h)}) dF(\vec{\epsilon}_{h,t+1}) dF(x_{h,t})$$

and the risk-free rate is

$$R_t^f = \left[\beta \eta_t \int_{-\infty}^{\infty} \xi_t^{(l)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) \left(\iiint_{-\infty}^{\infty} (u'(\exp(g_{l,t+1}))) dF(\vec{\epsilon}_{l,t+1}) \right) dF(x_{l,t}) \right. \\ \left. + \beta (1 - \eta_t) \int_{-\infty}^{\infty} \xi_t^{(h)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) \left(\iiint_{-\infty}^{\infty} (u'(\exp(g_{h,t+1}))) dF(\vec{\epsilon}_{h,t+1}) \right) dF(x_{h,t}) \right]^{-1}$$

and so the equity premium is $E_t R_t^p = E_t R_t - R_t^f$. The variance of equity premium is computed as

$$\sigma^2(R_t^p) = E_t R_t^2 - (E_t R_t)^2$$

where

$$E_t R_t^2 = \eta_t \iiint_{-\infty}^{\infty} \left(R_t(C_{l,t+1}, \hat{x}_{l,t+1}^{(l)}, \hat{x}_{h,t+1}^{(l)}, \eta_{t+1}^{(l)}) \right)^2 dF(\vec{\epsilon}_{l,t+1}) dF(x_{l,t}) \\ + (1 - \eta_t) \iiint_{-\infty}^{\infty} \left(R_t(C_{h,t+1}, \hat{x}_{l,t+1}^{(h)}, \hat{x}_{h,t+1}^{(h)}, \eta_{t+1}^{(h)}) \right)^2 dF(\vec{\epsilon}_{h,t+1}) dF(x_{h,t})$$

C Data and estimation of parameters of the stochastic models

The long-run risk model was fit to annual data using maximum likelihood. Parameter estimates are shown in Table 6. All parameters, except ρ and ψ were estimated using data 1930–1977. The mean of consumption and dividends, \bar{g} and \bar{d} , respectively were set to their values in the period 1930 – 1977. The variances of the latent state process, consumption growth and dividend growth were estimated using the Kalman Filter. The dividend leverage, ψ , was set to either 3 or 2.5, which is slightly lower than values which maximize the likelihood.

D Details of the numerical solution procedure

D.1 Solution Method: 2- ρ model

This section describes the minimum weighted residuals method we use to obtain an approximate solution for the value function and the risky rate. We then explain how we assess the accuracy of the method.

Both the value function and the risky rate are approximated by a parametric function of the form

$$\Phi_y(X_t) = \exp \left(\sum_{i_c, i_h, i_\ell, i_\eta \in \mathcal{I}} \theta_{i_c, i_h, i_\ell, i_\eta}^y H_{i_c}(\varphi_c(C_t)) H_{i_h}(\varphi_h(\hat{x}_{h,t})) H_{i_\ell}(\varphi_\ell(\hat{x}_{\ell,t})) H_{i_\eta}(\varphi_\eta(\eta_t)) \right)$$

Parameter Estimates

Parameter	$\psi = 3$				$\psi = 2.5$	
	$\rho = .25$	$\rho = .3$	$\rho = .85$	$\rho = .9$	$\rho = .3$	$\rho = .85$
\bar{g}	1.93 (0.647)	1.93 (0.647)	1.93 (0.647)	1.93 (0.647)	1.93 (0.647)	1.93 (0.647)
\bar{d}	0.62 (1.837)	0.62 (1.837)	0.62 (1.837)	0.62 (1.837)	0.62 (1.837)	0.62 (1.837)
σ_g	2.09 (0.530)	2.05 (0.531)	1.48 (0.585)	1.32 (0.635)	2.03 (0.533)	1.35 (0.617)
σ_d	9.71 (0.518)	9.77 (0.517)	10.55 (0.513)	10.64 (0.513)	10.27 (0.515)	10.71 (0.512)
σ_x	2.53 (0.525)	2.50 (0.524)	2.46 (0.531)	2.57 (0.530)	2.57 (0.524)	2.54 (0.528)

Table 6: **Parameter estimates** (standard errors below in parentheses) using annual data and the long-run risk model, shown above, using data from 1930 until 1977.

where $X_t \equiv (C_t, \hat{x}_{h,t}, \hat{x}_{\ell,t}, \eta_t)$ denotes the vector of state variables of our two- ρ case¹⁴ and $y \in \{V, R\}$. The set of indices \mathcal{I} is defined by

$$\mathcal{I} = \{i_z = 1, \dots, n_z; z \in \{C, h, \ell, \eta\} | i_c + i_h + i_\ell + i_\eta \leq \max(n_c, n_h, n_\ell, n_\eta)\}$$

Implicit in the definition of this set is that we are considering a complete basis of polynomials.¹⁵ $H_\ell(\cdot)$ is a Hermite polynomial of order ℓ and $\varphi_z(\cdot)$ is a strictly increasing function that maps \mathbb{R} into \mathbb{R} . This function is used to map Hermitian nodes into values for the vector of state variables, $X_t \equiv (C_t, \hat{x}_{h,t}, \hat{x}_{\ell,t}, \eta_t)$,¹⁶ The parameters θ^y , $y \in \{V, R\}$, are then determined by a minimum weighted residuals method. More precisely, we define the residuals associated to both the direct Value function equation, $\mathcal{R}_V(\theta^V; X_t)$, and the Euler equations for risky assets (consumption claims and dividend claims), $\mathcal{R}_R(\theta^V; X_t)$, as

$$\mathcal{R}_V(\theta^V; X_t) \equiv \Phi_V(C_t, \hat{x}_t^h, \hat{x}_t^\ell, \eta_t) - (1 - \beta)u(C_t) - \frac{\beta}{\alpha} \log(\mathcal{V}_{t+1})$$

where

$$\begin{aligned} \mathcal{V}_{t+1} \equiv & \eta_t \int_{-\infty}^{\infty} \exp\left(-\alpha \iiint_{-\infty}^{\infty} \Phi_V\left(C_{t+1}^{(\ell)}, \hat{x}_{h,t+1}^{(\ell)}, \hat{x}_{\ell,t+1}^{(\ell)}, \eta_{t+1}^{(\ell)}\right) dF(\vec{\epsilon}_{\ell,t+1})\right) dF(x_{\ell,t}) + \\ & (1 - \eta_t) \int_{-\infty}^{\infty} \exp\left(-\alpha \iiint_{-\infty}^{\infty} \Phi_V\left(C_{t+1}^{(h)}, \hat{x}_{h,t+1}^{(h)}, \hat{x}_{\ell,t+1}^{(h)}, \eta_{t+1}^{(h)}\right) dF(\vec{\epsilon}_{h,t+1})\right) dF(x_{h,t}) \end{aligned}$$

and

$$\mathcal{R}_R(\theta^R, \theta^V; X_t) \equiv u'(C_t) - \beta \mathcal{E}_{t+1}$$

¹⁴In the single ρ case, the vector of state variables reduces to $X_t = (C_t, x_t)$ and the approximant takes the simpler form $\Phi_y(X_t) = \exp\left(\sum_{i_c, i_x \in \mathcal{I}} \theta_{i_c, i_x}^y H_{i_c}(\varphi_c(C_t)) H_{i_x}(\varphi_x(\hat{x}_t))\right)$.

¹⁵See Judd (1998), Chapter 7.

¹⁶We use this function in order to be able to narrow down the range of values taken by the state variables, such that the approximation performs better when evaluated on the data.

where

$$\begin{aligned} \mathcal{E}_{t+1} \equiv & \eta_t \int_{-\infty}^{\infty} \left(\xi_{\ell,t} \iiint_{-\infty}^{\infty} u' \left(C_{t+1}^{(\ell)} \right) \Phi_R \left(C_{t+1}^{(\ell)}, \widehat{x}_{h,t+1}^{(\ell)}, \widehat{x}_{\ell,t+1}^{(\ell)}, \eta_{t+1}^{(\ell)} \right) \underbrace{\frac{D_{t+1}^{(\ell)}}{D_t}}_{(i)} dF(\vec{\mathcal{E}}_{\ell,t+1}) \right) dF(x_{\ell,t}) \\ & + (1 - \eta_t) \int_{-\infty}^{\infty} \left(\xi_{h,t} \iiint_{-\infty}^{\infty} u' \left(C_{t+1}^{(h)} \right) \Phi_R \left(C_{t+1}^{(h)}, \widehat{x}_{h,t+1}^{(h)}, \widehat{x}_{\ell,t+1}^{(h)}, \eta_{t+1}^{(h)} \right) \underbrace{\frac{D_{t+1}^{(h)}}{D_t}}_{(ii)} dF(\vec{\mathcal{E}}_{h,t+1}) \right) dF(x_{h,t}) \end{aligned}$$

where $\vec{\mathcal{E}}_{v,t+1} = \{\varepsilon_{x_v,t+1}, \varepsilon_{d_v,t+1}, \varepsilon_{g_v,t+1}\}$, with $v \in \{h, \ell\}$ is a vector of standard normal shocks with distribution $F(\vec{\mathcal{E}}_{v,t+1})$. (i) and (ii) are only present in the dividend claim case. We also define

$$\xi_{v,t} \equiv \frac{\phi' \left(\iiint_{-\infty}^{\infty} \Phi_V \left(C_{t+1}^{(v)}, \widehat{x}_{h,t+1}^{(v)}, \widehat{x}_{\ell,t+1}^{(v)}, \eta_{t+1}^{(v)} \right) dF(\vec{\mathcal{E}}_{v,t+1}) \right)}{\Psi_t} \text{ for } v \in \{h, \ell\}$$

with

$$\begin{aligned} \Psi_t \equiv & \eta_t \int_{-\infty}^{\infty} \phi' \left(\iiint_{-\infty}^{\infty} \Phi_V \left(C_{t+1}^{(\ell)}, \widehat{x}_{h,t+1}^{(\ell)}, \widehat{x}_{\ell,t+1}^{(\ell)}, \eta_{t+1}^{(\ell)} \right) dF(\vec{\mathcal{E}}_{\ell,t+1}) \right) dF(x_{\ell,t}) \\ & + (1 - \eta_t) \int_{-\infty}^{\infty} \phi' \left(\iiint_{-\infty}^{\infty} \Phi_V \left(C_{t+1}^{(h)}, \widehat{x}_{h,t+1}^{(h)}, \widehat{x}_{\ell,t+1}^{(h)}, \eta_{t+1}^{(h)} \right) dF(\vec{\mathcal{E}}_{h,t+1}) \right) dF(x_{h,t}) \end{aligned}$$

In both cases, $C_{t+1}^{(v)}$, $\widehat{x}_{h,t+1}^{(v)}$, $\widehat{x}_{\ell,t+1}^{(v)}$, $\eta_{t+1}^{(v)}$, $v \in \{h, \ell\}$, are obtained using the dynamic equations described in Section 3.2. These expression are simplified in the single ρ model as the agent is certain about the persistence. This case amounts to setting $\eta_t = 0$ for all t in the preceding expressions and consider only one process for \widehat{x}_t .

The vector of parameters θ^V and θ^R are then determined by projecting the residuals on Hermite polynomials. This then defines a system of orthogonality conditions which is solved for θ^V and θ^R . More precisely, we solve¹⁷

$$\begin{aligned} \langle \mathcal{R}_V(\theta^V; X_t) | \mathcal{H}(X_t) \rangle &= \int \mathcal{R}_V(\theta^V; X_t) \mathcal{H}(X_t) \Omega(X_t) dX_t = 0 \\ \langle \mathcal{R}_R(\theta^R, \theta^V; X_t) | \mathcal{H}(X_t) \rangle &= \int \mathcal{R}_R(\theta^R, \theta^V; X_t) \mathcal{H}(X_t) \Omega(X_t) dX_t = 0 \end{aligned}$$

where

$$\mathcal{H}(X_t) \equiv H_{i_c}(\varphi_h(C_t)) H_{i_h}(\varphi_h(\widehat{x}_t^h)) H_j(\varphi_\ell(\widehat{x}_t^\ell)) H_k(\varphi_\eta(\eta_t)) \text{ with } i_c + i_h + i_\ell + i_\eta \leq \max(n_c, n_h, n_\ell, n_\eta)$$

and

$$\Omega(X_t) \equiv \omega(\varphi_h(C_t)) \omega(\varphi_h(x_t^h)) \omega(\varphi_\ell(x_t^\ell)) \omega(\varphi_\eta(\eta_t))$$

¹⁷It should be clear to the reader that the integral refers to a multidimensional integration problem, as we integrate over C , x^h , x^ℓ and η .

where $\omega(x) = \exp(-x^2)$ is the appropriate weighting function for Hermite polynomials. Note that since the knowledge of the risky interest rate is not needed to evaluate the direct value function in equilibrium, the system can be solved recursively. We therefore first solve the value function approximation problem, and use the result vector of parameters θ^V to solve for the risky rate problem.

Integrals are approximated using a monomial approach whenever we face a multidimensional integration problem (inner integrals in the computation of expectations and projections) and a Gauss Hermitian quadrature approach when dealing with unidimensional integrals (outer integrals in the computation of expectations).¹⁸

The algorithm imposes that several important choices be made for the algorithm parameters. The first one corresponds to the degree of polynomials we use for the approximation. The results for the 2- ρ model are obtained with polynomials of order

- $(n_c, n_{x_h}, n_{x_\ell}, n_\eta) = (5, 2, 2, 2)$ for the value function when $\rho_h = 0.85$, and $(n_c, n_{x_h}, n_{x_\ell}, n_\eta) = (4, 2, 2, 2)$ for the value function when $\rho_h = 0.90$
- $(n_c, n_{x_h}, n_{x_\ell}, n_\eta) = (3, 3, 3, 3)$ for the interest rate,
- $(n_c, n_{x_h}, n_{x_\ell}, n_\eta) = (2, 4, 4, 1)$ for the asset prices.

The second choice pertains to the number of nodes. We use 8 nodes in each dimension (4096 nodes). The transform functions $\varphi(\cdot)$ are assumed to be linear $\varphi_z(x) = \kappa_z x$ where $\kappa_z, z \in \{c, h, \ell, \eta\}$ is a constant chosen such that the focus of the approximation is put on values of state variables taken in the data. More precisely, we set $\kappa_c = 2.0817$, $\kappa_h = 40$, $\kappa_\ell = 350$ and $\kappa_\eta = 1$.

The number of nodes used in the unidimensional quadrature method used in the outer integral involved in the computation of expectations is set to 12. In the case of the multidimensional integrals, we use a degree 5 rule for an integrand on an unbounded range weighted by a standard normal.¹⁹ Finally, the stopping criterion is set to 1e-6.

Given these parameters, the algorithm associated to each problem works as follows

1. Choose two candidate vectors of parameters θ^V and θ^R

¹⁸See Judd (1998), chapter 7.

¹⁹More precisely, we approximate

$$\int_{\mathbb{R}^k} F(x) \exp\left(\sum_{i=1}^k x_i^2\right) dx \simeq a_0 F(0) + a_1 \sum_{i=1}^k (F(re_i) + F(-re_i)) + a_2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k (F(se_i + se_j) + F(se_i - se_j) + F(-se_i + se_j) + F(-se_i - se_j))$$

where e_i denotes the i^{th} column vector of the identity matrix of order k . $r = \sqrt{1 + \frac{k}{2}}$, $s = \frac{\sqrt{2r}}{2}$, $a_0 = \frac{2\pi^{\frac{k}{2}}}{k+2}$, $a_1 = \frac{4-k}{4(k+2)} a_0$ and $a_2 = \frac{a_0}{2(k+2)}$. See Judd (1998) for greater details.

2. Find the nodes, $r_{j_z}, j_z = 1, \dots, m_z$, at which the residuals are evaluated. These nodes corresponds to the roots of the different Hermite polynomials involved in the approximation, then compute the values of the state variables as

$$C_{j_c} = \varphi_c^{-1}(r_{j_c}), x_{j_h}^h = \varphi_h^{-1}(r_{j_h}), x_{j_\ell}^\ell = \varphi_\ell^{-1}(r_{j_\ell}), \eta_{j_\eta} = \varphi_\eta^{-1}(r_{j_\eta})$$

3. Evaluate the residuals $\mathcal{R}_V(\theta^V; X_t)$ and $\mathcal{R}_R(\theta^R, \theta^V; X_t)$ and compute the orthogonality conditions $\langle \mathcal{R}_V(\theta^V; X_t) | \mathcal{H}(X_t) \rangle$ and $\langle \mathcal{R}_R(\theta^R, \theta^V; X_t) | \mathcal{H}(X_t) \rangle$.
4. if the orthogonality conditions are satisfied, in the sense the residuals are lower than the stopping criterion ε , then the vector of parameters are given by θ^V and θ^R . Else update θ^V and θ^R using a Gauss Newton algorithm and go back to step 1.

D.2 Computation of Returns

Given an approximate solution for the value function and the risky return, and given a sequence $\{X_t\}_{t=t_1}^{t=t_2} = \{C_t, \hat{x}_{h,t}, \hat{x}_{\ell,t}, \eta_t\}_{t=t_1}^{t=t_N}$ of annual observations of aggregate per capita consumption, beliefs and prior probabilities in the time periods $t = t_1$ through $t = t_N$ we compute the conditional n–th order moment of the risky rate in period t as

$$E_t^n R_{t+1} = \iiint_{-\infty}^{\infty} \Phi(X_{t+1})^n dF(\vec{\varepsilon}_{t+1}) dF(x_t) \quad (41)$$

The model average n–th order moment is then computed as

$$ER^n = \frac{1}{t_2 - t_1} \left[\sum_{t=t_1}^{t=t_2} E_t^n R_{t+1} - (E_t^1 R_{t+1})^n \right] \quad (42)$$

Similarly, given a sequence $\{C_t, \hat{x}_{h,t}, \hat{x}_{\ell,t}, \eta_t\}_{t=t_1}^{t=t_N}$, the risk-free rate can be directly computed

$$R_t^f = \left[\beta \eta_t \int_{-\infty}^{\infty} \xi_t^{(l)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) \left(\iiint_{-\infty}^{\infty} (U'(\exp(g_{l,t+1}))) dF(\vec{\varepsilon}_{l,t+1}) \right) dF(x_{l,t}) \right. \\ \left. + \beta (1 - \eta_t) \int_{-\infty}^{\infty} \xi_t^{(h)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) \left(\iiint_{-\infty}^{\infty} (U'(\exp(g_{h,t+1}))) dF(\vec{\varepsilon}_{h,t+1}) \right) dF(x_{h,t}) \right]^{-1}$$

Just as in the preceding section, integrals are approximated using a monomial approach whenever we face a multidimensional integration problem (inner integrals in the computation of expectations and projections) and a Gauss Hermitian quadrature approach when dealing with unidimensional integrals (outer integrals in the computation of expectations). The n–order moments are then obtained in a similar fashion as for the risky rate.

The (conditional) equity premium at time t , is the random variable denoted $R_t^p \equiv E_t^1 R_{t+1} - R_t^f$. Therefore, the n–order order moments of the equity premium can be computed as in eq. (42).

D.3 Accuracy

Our measure of accuracy of the risky rate builds heavily on previous work by Judd (1992). Since we are mostly interested in the empirical properties of the model, we mainly evaluate the accuracy of the solution for the data. Accuracy is assessed by considering the following rearrangement of the Euler equation error (both in the case of the consumption claim based approach and the dividend claim based approach)

$$\mathcal{E}(X_t) = \frac{u'^{-1}(\beta \mathcal{E}_{t+1})}{C_t} - 1$$

This measure then gives us the error an agent would make by using the approximate solution for the risky rate as a rule of thumb for deciding investing one additional dollar as asset holding. This quantity is computed for each value of the state variables in the data. Then three measures, formerly proposed by Judd (1992) are considered

$$E_1 = \log_{10}(E(|\mathcal{E}(X_t)|)), \quad E_2 = \log_{10}(E(\mathcal{E}(X_t)^2)), \quad \text{and} \quad E_\infty = \log_{10}(\sup |\mathcal{E}(X_t)|)$$

The first measure corresponds to the average absolute error, the second one corresponds to the quadratic average of the error, while the last one reports the maximal error an agent would make using the rule of thumb. All measures are expressed in \log_{10} terms, which furnishes a natural way of interpreting the accuracy measure. For instance, a value of E_1 equal to -4 indicates that an agent who uses the approximated decision rule would make –on average– a mistake of 1 dollar for each 10000 dollars invested in the risky asset. These measures are evaluated outside the grid points that are used to compute the approximation. Since our ultimate goal is to assess the quantitative relevance of the model, we need to make sure that our approximation performs well for the data we use. Hence, the measures are evaluated using the data. Results for both models are reported in Table 7 and show that the approximation is accurate.

		Consumption Claims			Dividend Claims		
γ	α	E_1	E_2	E_∞	E_1	E_2	E_∞
Single- ρ							
1.50	24.9	-4.27	-6.83	-3.79	-4.27	-6.83	-3.79
2.00	15.6	-4.14	-6.59	-3.70	-4.44	-7.19	-4.00
2.50	9.27	-4.04	-6.41	-3.63	-4.74	-7.81	-4.33
Two- ρ							
1.50	28.9	-3.23	-4.89	-2.97	-3.23	-4.89	-2.97
2.00	18.9	-3.30	-5.00	-2.97	-3.60	-5.60	-3.27
2.50	11.5	-3.34	-5.07	-2.98	-4.04	-6.46	-3.68

Table 7: **Accuracy of the numerical solution** (single- ρ and two- ρ models). In each case, α was set such that the model generates an average risk-free rate of 1.85%.

For example, let us consider the single ρ case with $\gamma = 2$, an agent who uses the approximate solution based on consumption claims would make, on average, a 1 dollar mistake for every 14,000 dollars invested in the

assets, while the maximal error would be of the same order. The table also indicates that even better levels of accuracy are also obtained for the dividend claim approach (1\$ for every 28,000\$). Good performances are valid for the two values of persistence (ρ) we consider. In the two rho case, the performances of the approximation slightly deteriorate. This accuracy loss is essentially due to the structure of the problem. In the single ρ case, the model is almost log-linear, such that our approximation performs remarkably well. In the two ρ case, the quasi log-linearity is lost as we have to compose probabilities of each model. Increasing the degree of the polynomials yields some (marginal) improvements but (i) leave the results almost unchanged and (ii) comes at a substantial computational cost. We therefore kept the degrees of the polynomials as they are. The accuracy properties of the approximate solution are very similar for the parametrization we consider in the robustness check exercise.²⁰

²⁰Accuracy is actually improved by increasing persistence, lowering the leverage and the discount factor.

References

- ABEL, A. (1999): "Risk premia and term premia in general equilibrium," *Journal of Monetary Economics*, 43(1), 3–33.
- (2002): "An exploration of the effects of pessimism and doubt on asset returns," *Journal of Economic Dynamics and Control*, 26(7-8), 1075–1092.
- BANSAL, R., A. GALLANT, AND G. TAUCHEN (2007): "Rational pessimism, rational exuberance, and asset pricing models," *Review of Economic Studies*, 74(4), 1005–1033.
- BANSAL, R., AND A. YARON (2004): "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.
- BARRO, R. (2006): "Rare disasters and asset markets in the twentieth century*," *The Quarterly Journal of Economics*, 121(3), 823–866.
- BEELER, J., AND J. CAMPBELL (2009): "The long-run risks model and aggregate asset prices: an empirical assessment," Discussion paper, NBER.
- BOYLE, P., L. GARLAPPI, R. UPPAL, AND T. WANG (2010): "Keynes Meets Markowitz: The Trade-off Between Familiarity and Diversification," Discussion paper, London Business School.
- CABALLERO, R., AND A. KRISHNAMURTHY (2008): "Collective risk management in a flight to quality episode," *The Journal of Finance*, 63(5), 2195–2230.
- CAMERER, C. E., AND M. WEBER (1992): "Recent Developments in Modelling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty*, 5, 325–70.
- CAMPBELL, J., AND R. SHILLER (1988): "The dividend-price ratio and expectations of future dividends and discount factors," *Review of Financial Studies*, 1(3), 195–228.
- CAMPBELL, J. Y. (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104, 298–345.
- CAMPBELL, JOHN, Y., AND J. COCHRANE (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205.
- CECCHETTI, S., P. S. LAM, AND C. M. NELSON (2000): "Asset Pricing with Distorted Beliefs: Are Equity Returns Too Good to Be True?," *The American Economic Review*, 90, 787–805.
- CHEN, H., N. JU, AND J. MIAO (2009): "Dynamic asset allocation with ambiguous return predictability," Discussion paper, Massachusetts Institute of Technology.
- CONSTANTINIDES, G. (1990): "Habit formation: A resolution of the equity premium puzzle," *The Journal of Political Economy*, 98(3), 519–543.
- CONSTANTINIDES, G., AND A. GHOSH (2010): "Asset pricing tests with long run risks in consumption growth," Discussion paper, Chicago Booth GSB.
- DOW, J., AND S. D. C. WERLANG (1992): "Uncertainty aversion, risk aversion, and the optimal choice of portfolio," *Econometrica: Journal of the Econometric Society*, 60(1), 197–204.
- ELLSBERG, D. (1961): "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643–669.
- EPSTEIN, L., AND M. SCHNEIDER (2008): "Ambiguity, information quality, and asset pricing," *The Journal of Finance*, 63(1), 197–228.
- EPSTEIN, L., AND T. WANG (1994): "Intertemporal asset pricing under Knightian uncertainty," *Econometrica: Journal of the Econometric Society*, 62(2), 283–322.
- EPSTEIN, L. G. (2010): "A Paradox for the "Smooth Ambiguity" Model of Preference," *Econometrica*, 78(6), 2085–2099.
- EPSTEIN, L. G., AND S. E. ZIN (1991): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99(2), 263–86.

- FERNANDEZ-VILLAYERDE, J., AND J. F. RUBIO-RAMÍREZ (2010): "Macroeconomics and Volatility: Data, Models, and Estimation," Discussion paper, University of Pennsylvania.
- GILBOA, I., AND D. SCHMEIDLER (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18(2), 141–153.
- GIORDANI, P., AND P. SODERLIND (2006): "Is there evidence of pessimism and doubt in subjective distributions? Implications for the equity premium puzzle," *Journal of Economic Dynamics and Control*, 30(6), 1027–1043.
- GOLLIER, C. (2009): "Portfolio Choices and Asset Prices: The Comparative Statics of Ambiguity Aversion," IDEI Working Papers 357, Institut d'Économie Industrielle (IDEI), Toulouse.
- HANSEN, L., AND T. SARGENT (2010): "Fragile beliefs and the price of uncertainty," *Quantitative Economics*, 1(1), 129–162.
- HANSEN, L. P. (2007): "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97(2), 1–30.
- JOUINI, E., AND C. NAPP (2006): "Heterogeneous beliefs and asset pricing in discrete time: An analysis of pessimism and doubt," *Journal of Economic Dynamics and Control*, 30, 1233–1260.
- JU, N., AND J. MIAO (2010): "Ambiguity, Learning, and Asset Returns," Working paper series, Boston University.
- JUDD, K. (1992): "Projection methods for solving aggregate growth models* 1," *Journal of Economic Theory*, 58(2), 410–452.
- JUDD, K. L. (1998): *Numerical Methods in Economics*. MIT Press, Cambridge, MA.
- KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2005): "A smooth model of decision making under ambiguity," *Econometrica*, 73(6), 1849–1892.
- KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2009): "Recursive smooth ambiguity preferences," *Journal of Economic Theory*, 144(3), 930–976.
- KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2011): "On the Smooth Ambiguity Model: A Reply," Discussion paper, Department of Economics, University of Oxford, <http://www.economics.ox.ac.uk/research/WP/pdf/paper524.pdf>.
- KNIGHT, F. H. (1921): *Risk, Uncertainty, and Profit*. Houghton Mifflin Company, Boston, MA.
- KOCHERLAKOTA, N. (1996): "The equity premium: It's still a puzzle," *Journal of Economic Literature*, 34(1), 42–71.
- KREPS, D. M., AND E. L. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46(1), 185–200.
- LJUNGQVIST, L., AND T. SARGENT (2004): *Recursive macroeconomic theory*. The MIT Press.
- MACCHERONI, F., M. MARINACCI, AND D. RUFFINO (2010): "Alpha as Ambiguity: Robust Mean-Variance Portfolio Analysis," *Working Papers*.
- MEHRA, R., AND E. C. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15(2), 145–161.
- MUKERJI, S., AND J. M. TALLON (2001): "Ambiguity aversion and incompleteness of financial markets," *Review of Economic Studies*, 68(4), 883–904.
- RAVN, M., AND H. UHLIG (2002): "On adjusting the Hodrick-Prescott filter for the frequency of observations," *Review of Economics and Statistics*, 84(2), 371–376.
- SCHMEIDLER, D. (1989): "Subjective Probability and Expected Utility without Additivity," *Econometrica*, 57(3), 571–87.
- SHEPARD, N., AND A. HARVEY (1990): "On the probability of estimating a deterministic component in the local level model," *Journal of Time Series Analysis*, 11(4), 339–347.
- UHLIG, H. (2010): "A model of a systemic bank run," *Journal of Monetary Economics*, 57(1), 78–96.
- WEIL, P. (1989): "The equity premium puzzle and the risk-free rate puzzle," *Journal of Monetary Economics*, 24(3), 401–421.
- WEITZMAN, M. (2007): "Subjective expectations and asset-return puzzles," *The American Economic Review*, 97(4), 1102–1130.