



**HAL**  
open science

## Ambiguity and the historical equity premium

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

► **To cite this version:**

Fabrice Collard, Sujoy Mukerji, Kevin Sheppard, Jean-Marc Tallon. Ambiguity and the historical equity premium. 2011. halshs-00594096v2

**HAL Id: halshs-00594096**

**<https://shs.hal.science/halshs-00594096v2>**

Submitted on 4 Oct 2012 (v2), last revised 3 May 2016 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Documents de Travail du Centre d'Économie de la Sorbonne

C  
E  
S  
W  
o  
r  
k  
i  
n  
g  
P  
a  
p  
e  
r  
s



## **Ambiguity and the historical equity premium**

Fabrice COLLARD, Sujoy MUKERJI,  
Kevin SHEPPARD, Jean-Marc TALLON

**2011.32**

*Version révisée*



# Ambiguity and the historical equity premium\*

Fabrice Collard

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

Sujoy Mukerji<sup>†</sup>

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

Kevin Sheppard

Department of Economics and Oxford-Man Institute  
University of Oxford, Oxford OX2 6ED

Jean-Marc Tallon

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

Revised August 2012

## Abstract

This paper assesses the quantitative impact of ambiguity on the historically observed financial asset returns and prices. The single agent, in a dynamic exchange economy, treats uncertainty about the conditional mean of the probability distribution on consumption and dividends in the next period as ambiguous, an ambiguity that is *endogenously* dynamic, e.g. , increasing during recessions. We calibrate ambiguity aversion to match *only* the first moment of the risk-free rate in data and, importantly, the (conditional) ambiguity to match the uncertainty conditional on the actual history of macroeconomic growth outcomes. The model matches observed asset return dynamics very substantially.

**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, Hening Liu, 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).

<sup>†</sup>Contact author: [sujoy.mukerji@economics.ox.ac.uk](mailto:sujoy.mukerji@economics.ox.ac.uk)

# 1 Introduction

This paper seeks to assess the *quantitative* impact of ambiguity on financial asset returns and prices, in particular, their dynamic paths as observed in history. Ambiguity refers to uncertainty about the "true" probability distribution governing future consumption and dividend outcomes. The decision maker's ambiguity attitude determines how and to what extent such uncertainty affects her choices (e.g., whether she is averse to such uncertainty and if so, the level of aversion). Our goal is two fold: to connect the (macroeconomic) uncertainty *as it obtained on the path of history* to the movements in asset returns and prices along that path and to assess, quantitatively, the role of ambiguity sensitivity in *that* connection. To serve these goals we incorporate two components in our analysis. One, the (dynamic) ambiguity variable we construct as the key explanatory input is explicable in terms of the conditional uncertainty obtained at information sets along the path of observations of historical macroeconomic growth rates, as opposed to counterfactual, simulated sample paths. Two, our model of agent's preferences departs from standard expected utility by simply allowing for ambiguity sensitivity; take that away, you get back to expected utility. These two components, together with the demonstration that they alone are sufficient to explain asset return dynamics very substantially, distinguish the contribution in this paper.

Ambiguity and its possible relevance to economics were discussed intuitively by Knight (1921) and Ellsberg (1961). Decision theoretic formulations by Schmeidler (1989) and Gilboa and Schmeidler (1989) presented a first set of tools to incorporate the idea into formal economic analysis. Introspection 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)). Agents are typically posited as ambiguity averse, 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.<sup>1</sup> An important reason why we think ambiguity is pervasive in economic decision making 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

---

<sup>1</sup>See Dow and Werlang (1992), Epstein and Wang (1994), Mukerji and Tallon (2001), Caballero and Krishnamurthy (2008), Chen, Ju, and Miao (2009), Gollier (2011), Boyle, Garlappi, Uppal, and Wang (2010), Hansen and Sargent (2010), Maccheroni, Marinacci, and Ruffino (2010) and Uhlig (2010), *inter alia*.

same model, all of which are plausible on the basis of historical data. If the models make distinct (probabilistic) 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 dynamic, evolving, as the agent learns from history. Second, the agent's preferences are sensitive to this ambiguity. Let's take these two ingredients in turn.

We assume the agent believes the economy evolves according to (a modified version of) the hidden state model analyzed in Bansal and Yaron (2004). In that 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 dividend) growth distribution from the trend and thus, in this sense, the transient business cycle component. The latent state is not directly observable<sup>2</sup> and is assumed to evolve according to an  $AR(1)$  process. Inference can be made on its current value although it can never be fully pinned down through inference based on observation of growth outcomes. The growth distribution is assumed to be Gaussian with a given (time invariant) volatility parameter; the mean completely characterizes the distribution but is subject to innovation in every period, and never completely known, since it is partly determined by the latent state variable. The ambiguity in the agent's belief is her conditional uncertainty about the mean of the probability distribution on dividend and consumption growth next period.<sup>3</sup> 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 uncertainty (at date  $t$ ) about the probability distribution over future growth outcomes given the best statistical (rational, Bayesian) inference from history of observations on growth outcomes up until time  $t$ .

---

<sup>2</sup>If it were known to the agent, then the mean rate of growth of both consumption and dividends would be known.

<sup>3</sup>While ambiguity about higher order moments is plausible, we focus on mean uncertainty for parsimony and for the simplicity of the connection to data. Even this minimal ambiguity is enough to ensure time-varying heteroskedasticity in beliefs about growth rates and generate counter cyclical conditional volatility of returns.

The updated belief, the Bayes posterior, on the date  $t$  hidden state, is a key ingredient of the agent's decision model, the smooth ambiguity model (Klibanoff, Marinacci, and Mukerji (2005, 2009); henceforth KMM2005, KMM2009). In this model, the agent evaluates a contingent consumption plan in the following way: for each possible realization of the latent state variable, compute the expected utility of the contingent plan with respect to the corresponding (first order) probability distribution on growth. Then, aggregate a transformation of these expected utilities with respect to the second order prior, i.e., the updated belief over the latent state. The transformation of the expected utilities captures the agent's ambiguity attitude; in particular, if the transformation is concave then the agent is ambiguity averse while if it is affine then the agent is ambiguity neutral and simply maximizes a subjective expected utility. In the case of the specification of the transformation used in this paper, given a consumption plan its evaluation is an expected utility calculated using an "as if" second order probability. The "as if" distribution is derived by redistributing probability weights in the original towards those first order distributions for which the expected (continuation) utility of the consumption plan is lower, the extent of distortion increasing with ambiguity aversion. Since it adjusts the evaluating probability measure in a way that depends on the plan being evaluated, this evaluation procedure is different from expected utility but this is precisely what builds in the robustness concern into the evaluation.

The agent chooses a portfolio of two assets every period, one risk free another risky. Ambiguity aversion is calibrated to match only the first moment of the risk free rate in data. The magnitudes of ambiguity aversion so invoked, we argue, are plausible by showing that they imply levels of ambiguity premium that are similar in magnitude to levels of risk premium implied by levels of risk aversion parameters widely regarded as plausible. Constant relative risk aversion is restricted to lie between 1 and 3. This yields equilibrium rates of return and asset prices implied by the (amended) Lucas-tree model at each information set following a history of macroeconomic growth outcomes. While we provide some analytical approximations to help fix intuition, our results are obtained by numerical solution of the model which involves two key computational challenges: the dimension of the state space is large – four state variables – and the integral involves 4 nested integrals.

In our main model, dubbed the "two- $\rho$  model", the value of the persistence parameter ( $\rho$ ) of the  $AR(1)$  process driving latent states is not completely known to the agent: she believes  $\rho$  may take one of two values, moderate or high, and updates her beliefs about

this by Bayes rule following observation of growth outcomes. This is motivated, in part, by the difficulty in determining, through observation, whether the true growth process is highly persistent with the persistent component having a small variance or, moderately persistent with greater variance of the persistent component. Like the location of the hidden state the value of  $\rho$  is not observable directly; but unlike the latent state, in principle, it is (eventually) learnable from growth observations. However, in practice, it has proved difficult to do so, and even after several decades of data, its estimates remain fragile. The uncertainty about  $\rho$  affects and is in turn affected by the uncertainty about the latent variable. These seem good reasons to model an ambiguity sensitive agent as treating both uncertainties as ambiguous, in a consistent, unified fashion and that is what we do.

A second motivation for modeling the persistence as something not known from the outset, but learned over time, is that then the learning causes the uncertainty about the latent states, the ambiguity in our model, to vary endogenously over time in an intuitive way. 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 from time to time, depending on history. 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. On the other hand, in the aftermath of a significant shock alternative conjectures may disagree considerably about growth prospects, causing uncertainty about the conditional mean to increase temporarily.<sup>4</sup>

In the two- $\rho$  model, the posterior on the latent states is a weighted mixture of the posteriors obtained by assuming each possible value of  $\rho$ . Analogously, given an equilibrium consumption plan, the corresponding "as if" posterior (what matters for the ambiguity averse agent's evaluation) on latent states is a weighted average of the two "as if" posteriors corresponding to each value of  $\rho$ . As explained in Section 3.2.2 after a positive shock the two "as if" (component) posteriors sit closer together compared to the two Bayesian posteriors, while they move further apart following a negative shock. Consequently, the variance of the (mixed) "as if" posterior moves counter cyclically in a more pronounced fashion than the Bayes posterior. It is as if the ambiguity averse agent reacts *asymmetrically* to the uncertainties following positive and negative shocks; she under plays and

---

<sup>4</sup>Two 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, with some resonance in recent times.

under reacts to the former and exaggerates the latter, compared to an expected utility agent. 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 the dynamics of ambiguity on the movements of asset returns and prices, as they actually obtain in history.<sup>5</sup>

Our main results not only include statistics such as average *conditional* moments of the model implied returns and price-dividend ratio (Tables 3, 3.2.3 and 5) but also, going beyond the usual practice in the literature, plots of the model implied time series of returns and price-dividend ratio, all based on conditional uncertainty at information sets reconcilable with historical growth data (Figures 3 and 7). We compare the level, volatility and dynamics of the model implied rates of return and price-dividend ratio to their counterparts in U.S. data and show that the match is quite good. The implied (conditional) equity premium is high enough to match the data primarily because while ambiguity aversion does not affect the return on the risky asset (the dividend claim) significantly, it lowers the risk free rate substantially as robustness concerns drive up the equilibrium price of the risk free asset. These concerns show up in the model as *endogenously* generated doubt and pessimism, to use the language of Abel (2002). In a standard rational expectations framework, the agent behaves as if she "knows" the true growth distribution; she will evaluate consumption plans putting probability one on the filtered state. Here, the agent maintains a non-degenerate posterior over the (latent) states, and the distribution over the growth rates is a mixture distribution. Hence, the (endogenously evolving) doubt. The pessimism derives from ambiguity aversion, the concern for robustness, and is embodied by the distortion in the "as if", distorted, posterior. The volatility and dynamics of the model implied returns, prices and especially the equity premium are determined very largely by the variance of the (mixed) "as if" posterior which, as explained above, is endogenously countercyclical.

## 1.1 Related literature

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 representa-

---

<sup>5</sup>That 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.



tive agent frameworks. Three papers closest to ours are Bansal and Yaron (2004), Hansen and Sargent (2010) and Ju and Miao (2012).

Bansal and Yaron (2004) (henceforth BY) pioneered the use of the (basic) model of beliefs we apply to show how long run risk (LRR) and aversion to such risk (while allowing a Kreps and Porteus (1978)/Epstein and Zin (1989)/Weil (1989) like separation of elasticity of intertemporal substitution (EIS) 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, instead of assuming that the agent believes with probability one that the persistence parameter has a high value. We also assume that the volatility of innovations to consumption is constant (as in BY's CASE I model). They show that including (an exogenous) stochastic volatility to the innovation in consumption growth, the defining difference in BY's CASE II model, is essential to adequately match the second moment of returns. We show, (1) and (2) are sufficient to yield a model of beliefs where the uncertainty and ambiguity vary *endogenously* over time and enough to match return volatility. In this way, we believe, the analysis in the paper demonstrates a broader scope of application of the LRR framework. Furthermore, they focus on drawing out model implications for unconditional returns whereas here the focus is on dynamics of conditional returns.

Hansen and Sargent (2010) study the effect of model uncertainty and robustness concerns on the price of macroeconomic risk. The single agent believes the economy evolves according to a BY CASE I model but is uncertain whether the persistence is moderate or high, like the agent in our two- $\rho$  model. However, their agent processes belief by applying *two* "risk-sensitivity operators", unlike in our model. The first operator, which may be interpreted in terms of an enhanced risk aversion obtained via a Kreps and Porteus (1978) style preference for earlier resolution of risk, applies to the evaluation (of the consumption plan) *conditional* on each of the two values of  $\rho$ . The other operator may be interpreted as a KMM2005/KMM2009 style smooth ambiguity aversion transformation where the agent's second order uncertainty is a two point (Bernoulli) belief, where each point in the support is the conditional evaluation given a  $\rho$ . Hence, while uncertainty about the two possible evaluations is treated as ambiguity, the uncertainty conditional on a value of

$\rho$  is not processed as ambiguity, unlike in our model.<sup>6</sup> Following a negative shock, preference for early resolution combined with high persistence ensures that the evaluation conditional on high persistence is very low; this pessimism implies that the price of risk associated with the high persistence is greater. The lower evaluation causes the smooth ambiguity aversion operator to distort the Bernoulli belief to increase the (“as if”) posterior on the high persistence model. In other words, a negative shock generates an increase in price of risk, i.e., move countercyclically. In contrast, in our model, the countercyclicality is generated by the changes in the riskiness in the agent’s (“as if”) second order belief, a mixture distribution of two component non-degenerate (“as if”) posteriors on latent states. These component “as if” posteriors do not have a role in Hansen and Sargent’s model since there conditional on a value of  $\rho$  the uncertainty is not processed as ambiguity. Their paper gives results on what they call, “uncertainty price.” It does not derive results and (statistics on) model implied rates of return and price-dividend ratio with evident counterparts in data, so we do not know how much better their model would match the data compared to ours. Note though, even if such results and statistics were derived they would give a far more limited account of the impact of ambiguity than our results since their model does not treat the uncertainty about latent states as ambiguous and includes effects of departures from expected utility quite separate from (non-neutrality to) ambiguity.

Ju and Miao (2012) use a modified smooth ambiguity framework to assess the effect of ambiguity on dynamics of asset prices. The model of beliefs there is different in that the hidden/latent state variable driving the (mean) growth rate in the economy may take only two possible values (while it may take a continuum of values in our model). On the other hand, the preference model is richer in that it incorporates an Epstein and Zin (1989) type EIS parameter in addition to ambiguity aversion. With the EIS parameter set at 1.5, they simulate different sample paths (of growth outcomes) and generate model implied returns and prices on these paths. They produce statistics on *unconditional* moments of returns and prices, by averaging across the counter-factual paths, which match the corresponding statistics in data very well. They also report, using graphs, model implied conditional returns and prices along the observed, historical sample path; here, their model is evidently less successful. As panel B in their Figure 3 shows, throughout the post-war

---

<sup>6</sup>As we have argued, the location of the latent state variable, being subject to an innovation every period, is hard to learn. The innovation element, in particular, makes a compelling case to treat this uncertainty like the uncertainty about  $\rho$ , as ambiguity. That we treat both uncertainties as ambiguous, in a consistent, unified fashion implies our model has one less parameter, one less degree of freedom, than the case where one has two different risk-sensitivity operators for the two uncertainties.

period the (second-order) belief has been almost completely stuck (almost Dirac) on the same latent (high-growth) state. As they explain with their Figure 2, movements in returns/prices take place in their model when beliefs move around between the two states. Hence, a model of just two latent states seems too coarse to bring out the implications of ambiguity when restricted to observed history. Relatedly, in an earlier working paper version they had showed, without the separate EIS effect added in their model implied (unconditional) equity premium matched data in terms of level but not volatility. By allowing for a richer specification of belief our model complements their findings by showing, using a preference model which does not depart from expected utility except in regard to ambiguity attitude, that it is possible to explain observed movements in returns and prices with conditional beliefs based on observed history.<sup>7</sup>

Veronesi (1999) constructs and theoretically analyzes a dynamic, rational expectations, expected utility representative agent model of asset pricing where beliefs are based on two hidden states (each specifying a mean growth rate) and shows that it implies time-varying expected returns and prices. However, it is a theoretical exercise and does not show what actual values and magnitudes are implied along information paths based on observed history. Gollier (2011) shows analytically, using a (static) smooth ambiguity model, that an increase in ambiguity aversion may not, in general, increase the equity premium, thereby making a good case for empirical investigation of the question. 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 models with habits in consumption which can match the level, variation and countercyclicality 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

---

<sup>7</sup>Recently, Strzalecki (2012) has argued that it is theoretically possible that recursive ambiguity frameworks have some preference for early resolution inseparably mixed in with ambiguity aversion. Compared to the model in the present paper what is different about the preferences in Ju and Miao (2012) and Hansen and Sargent (2010) is that they include separate components explicitly adding preference for early resolution above and beyond what may be already mixed in with ambiguity aversion.

faces time-varying uncertainty and it is variation in that uncertainty, rather than variation 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 assuming a fairly general form of beliefs. 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 and prices. 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

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

We follow KMM2009, which develops a dynamic, recursive version of the smooth ambiguity model in KMM2005. 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  $t = 1$ ; and, so on. 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  $\Theta$ , 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  $\pi_\theta$  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  $\Theta$  is denoted by  $\mu$ . 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,  $\beta$  is a discount factor and  $\phi$  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 on  $\Theta$  at  $s^t$ . In particular, a concave  $\phi$  characterizes ambiguity aversion, which is defined to be an aversion to mean preserving spreads in the distribution over expected utility values. Intuitively, ambiguity averse agents prefer acts whose evaluation is more robust to the possible variation in probabilities. This preference model does not, in general, impose reduction between  $\mu$  and the  $\pi_\theta$ 's in the support of  $\mu$ ; reduction only occurs when  $\phi$  is linear, a situation identified with ambiguity neutrality and wherein the preferences are observationally equivalent to that of a Bayesian expected utility maximizer with a subjective prior  $\mu$  (over parameters).

## 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  $\mu_t$  denote the (second order) belief on parameters in  $\Theta$  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 (as opposed to the direct value function in eq. (1)). An equilibrium of this economy is given by  $\{(P_\tau, P_\tau^f, e_\tau, b_\tau, C_\tau)\}_{\tau=1}^\infty$  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

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

The expressions are thus similar to those in an economy where the agent is an expected utility maximizer, but for the terms  $\Upsilon_t$  and  $\xi_t$ . Both  $\Upsilon_t$  and  $\xi_t$  depend on the ambiguity attitude,  $\phi$ , and beliefs. The function  $\xi_t$  is a Radon–Nikodym derivative, effecting a *node specific* change of measure, or “distortion”, on the posterior  $\mu_t$ . 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  $\alpha$  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:

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

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

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  $z_t = \ln\left(\frac{P_t}{D_t}\right)$ ,  $g_{t+1} = \ln\left(\frac{C_{t+1}}{C_t}\right)$ ,  $d_{t+1} = \ln\left(\frac{D_{t+1}}{D_t}\right)$ , 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,  $\xi_t$ . The distortion, in the case of ambiguity aversion, increases the (posterior) weight on one-period ahead probability distributions  $\pi_\theta$  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

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 believed 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, more realistic dynamic picture 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- $\rho$ 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 BY.

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

where  $(\varepsilon_{g,t+1}, \varepsilon_{d,t+1}, \varepsilon_{x,t+1})' \sim N(0, I)$ . The long-run growth rate of consumption and dividend are shown by  $\bar{g}$  and  $\bar{d}$ , respectively. The shock  $x$  is 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  $\rho$ . The factor  $\psi$  accounts for the empirically observed greater 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 BY, 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  $\varepsilon$ . 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  $\sigma_0^2$ , fixing thereby the agent’s prior belief  $\mu_0$ . We call this the *single- $\rho$  model*; its defining property being that the agent knows the value of  $\rho$ .

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 “ $\theta$ ”. 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  $\mu_t$  over this parameter, the current  $x_t$ . The belief  $\mu_t$  describes, exhaustively, her ambiguity about the probability distribution on the successor nodes. The agent updates  $\mu_t$  using Bayes rule conditional on the history of realizations of  $g_t$  and  $d_t$  given the Gaussian prior  $g_0 \sim N(\bar{x}_0, \sigma_0^2)$ . Updated beliefs are also Gaussian with mean  $\hat{x}_{t+1}$  and a (steady state) variance  $\Omega$ , i.e.,  $x_{t+1} \sim N(\hat{x}_{t+1}, \Omega)$ , where  $\Omega$  is defined in eq. (14). Hence, the evolution of  $\mu_t$  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, via a Kalman filter as follows:

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



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

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

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,  $\Omega$ , is defined as the solution to

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

### 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  $N(\hat{x}_t, \Omega)$  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 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 function<sup>8</sup> given as,

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

<sup>8</sup>Henceforth, we shall write  $\xi_t$  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.

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- $\rho$  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  $H_\iota(\cdot)$  is a Hermite polynomial of order  $\iota$  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  $\theta^\nu$  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 found, we may compute an approximate solution for the rates of returns. The risk-free and risky rates,  $R_t^f(C_t, \hat{x}_t; \alpha, \gamma)$  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  $\rho$  and the leverage-ratio parameter  $\psi$ ) 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, 1978-2011, 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. For our baseline calibration, we set  $\rho = 0.85$  (for the single- $\rho$  case), which is the annualized equivalent of the value used in BY and supported by the estimate obtained in Bansal, Gallant, and Tauchen (2007). The other value we apply in the single- $\rho$  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,  $\psi$ , was set to 3 as in BY, although Constantinides and Ghosh (2010) estimated it to be slightly lower, close to the value we use for robustness checks ( $\psi = 2.5$ ).

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 2011, a total of 82 observations.

Turning to preference parameters, in all cases the ambiguity aversion parameter  $\alpha$  was calibrated to produce a real risk-free rate of 1.5%, averaged over  $t = 1978, \dots, 2011$ , which is the average observed rate in that period. No other moments were used in the choice of  $\alpha$ . The relative risk aversion parameter  $\gamma$  was allowed to range between 1 (log utility) and 3, regarded as plausible in macroeconomic models (Ljungqvist and Sargent, 2004, pg. 426); the "baseline" calibration set  $\gamma = 2.5$ .<sup>9</sup> The discount factor  $\beta$  was set to 0.975, which corresponds to the discount rate used in BY. To check for robustness we varied a number of the key non-estimated parameters, including  $\rho = 0.9$ ,  $\beta \in \{.965, .97, .98\}$  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 of real consumption and of real dividends to obtain a time series of *average conditional moments* of the rates of return using our numerical solution technique. Our model produces 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 along the sample path, conditional on the observed history at each time  $t$ , with  $t = 1$  corresponding to 1978. This is different from what is commonly presented where unconditional moments are instead given. This is important in our model for two reasons. First, our agent learns and so averages would make use of beliefs that could not be reconciled with historical growth. Second, when consumption growth has a highly persistent component, unconditional moments may be heavily influenced by economic conditions not experienced by the agent. The top panel of Table 1 reports the average of these conditional moments over the period 1978-2011. The bottom panel of Table 1

---

<sup>9</sup>If the two smooth ambiguity preferences do not share the same risk attitude it is not necessarily true that a more concave  $\phi$  means more ambiguity aversion. Hence  $\alpha$  is meaningfully calibrated *given* a value of  $\gamma$ ; not independent of  $\gamma$ .

<b>Returns and Volatility</b>								
	$\gamma$	$\alpha$	$E(r)$	$E(r - r^f)$	$\sigma(r^f)$	$\sigma(r)$	$\sigma(r - r^f)$	
Data			8.08	6.68	2.20	16.5	16.1	
Baseline	2.5	7.24	5.87	4.37	2.42	8.80	8.45	
Bayesian	2.5	$\approx 0$	5.45	-1.52	2.47	8.78	8.43	

<b>Price-Dividend Ratio</b>								
	$\gamma$	$\alpha$	$E(P/D)$	$\sigma(P/D)$	$E(p - d)$	$\sigma(p - d)$	AC1	AC2
Data			45.5	20.0	3.72	0.45	0.80	0.76
Baseline	2.5	7.24	25.4	3.30	3.23	0.13	0.37	0.38
Bayesian	2.5	$\approx 0$	22.6	3.65	3.10	0.17	0.60	0.59

Table 1: **Results (single- $\rho$ )**: Average of the predicted *conditional* moments of rates of return (on dividend claim) in the single- $\rho$  model over the period 1978–2011. The data based moments are averages of returns ( $r$ ) or excess returns ( $r - r^f$ ), or sample standard deviations of the same quantities. The corresponding model-implied moments are averages of the conditional expected market return and the conditional equity premium. The lower panel shows data based and model implied average and std. dev. of the price-dividend ratio in the single- $\rho$  model over the period 1978–2011. AC1 and AC2 denote the first and second order autocorrelation of  $p - d$ , the log price-dividend ratio.

reports moments of the price-dividend series ( $P/D$ ) of the market claims and of the log price-dividend series ( $p - d$ ) implied by the model predicted market returns (eqs. 9 – 11). We compute the price-dividend ratio applying the relationship

$$R_{t+1} = \frac{\exp(p_{t+1} - d_{t+1}) + 1}{\exp(p_t - d_t)} \exp(d_{t+1}) \quad (21)$$

where  $d_t$  is taken from the historical data,  $R_{t+1}$  and  $p_{t+1}$  are computed from the model, and the recursion is started from the actual price-dividend ratio in 1977 ( $t = 0$ ).

### 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  $N(\hat{x}_t, \Omega)$  on  $x_t$  the effect of  $\xi_t$  is “as if” there is a new distorted posterior,  $\tilde{\mu}_t \equiv \xi_t(x_t) \otimes N(\hat{x}_t, \Omega)$ , with density given by

$$\tilde{f}(x_t) = \xi_t(x_t | C_t, \hat{x}_t; \alpha) \frac{1}{\sqrt{2\pi\Omega}} \exp\left(-\frac{(x_t - \hat{x}_t)^2}{2\Omega}\right). \quad (22)$$

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  $\mu_t$ ) on  $x_t$ 's that generate probability distributions associated with lower expected continuation values,  $E_{x_t}(V(C_{t+1}; \hat{x}_{t+1}))$ . The dis-

torted posterior gives rise to a conditional one-step-ahead distribution on growth which we call the *twisted (predictive) distribution*

$$g_{t+1} \sim \xi_t(x_t) \otimes N(\hat{x}_t, \Omega) \otimes N(\rho x_t + \bar{g}, \sigma_x^2 + \sigma_g^2). \quad (23)$$

When  $\xi_t(x_t) = 1$  the formula (23) describes the belief of a Savage-Bayes rational (or, equivalently, ambiguity neutral) agent, a useful benchmark. Such an agent, whom we dub “Bayesian,” is uncertain about  $x_t$  with beliefs about growth which are described by 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.<sup>10</sup>

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 about the latent state. Figure 1 shows the average one-step-ahead distributions (on growth) corresponding to these three cases of beliefs in the single- $\rho$  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, while the Bayesian agent only incorporates the doubt. This is the key to understanding the mechanism through which ambiguity aversion affects asset returns and prices.

---

<sup>10</sup>Different 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.

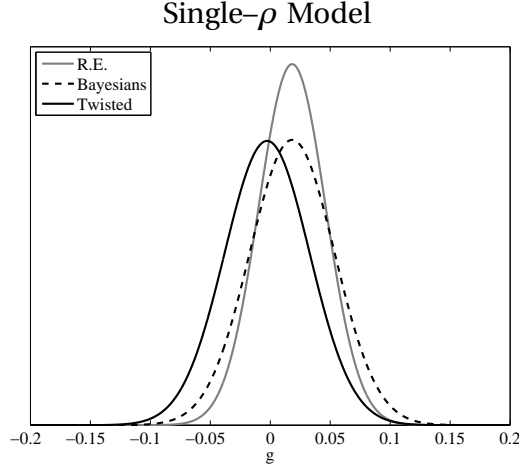


Figure 1: **Beliefs and “as-if” beliefs:** The agent’s “as-if” belief about the conditional distribution of consumption growth with no uncertainty about the latent state (R.E.), with uncertainty about the latent state but without ambiguity aversion (Bayesian) and with ambiguity aversion about the uncertainty of the latent state (Twisted) from the single- $\rho$  model. The distributions were computed using the baseline specification and the level of consumption and the hidden state variables set to their averages over the period 1978–2011.

### 3.1.6 Explaining the results: analytical approximation and comparative statics

The results compiled in Table 1 and Figures 1 and 2 are based on *numerical* solutions. However, in the case of the single- $\rho$  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. (22), is well approximated by a Normal density, whose mean and variance are denoted by  $\tilde{x}_t$  and  $\widetilde{Var}_t(x_t)$ , respectively.<sup>11</sup>

The left panel in Figure 2 depicts the comparative statics of ambiguity aversion and risk aversion on the rates of return. The risk-free rate is approximated as:

$$r_t^f = -\ln \beta + \gamma \bar{g} + \gamma \rho \tilde{x}_t - \frac{\gamma^2}{2} \left( \sigma_x^2 + \sigma_g^2 + \rho^2 \widetilde{Var}_t(x_t) \right). \quad (24)$$

*Absent* ambiguity aversion, an increase in  $\gamma$  has, principally, two countervailing effects. The first effect shows up in the term  $\gamma \bar{g}$ . Here an increase in  $\gamma$  makes the agent want to smooth consumption between the present and future states more; since  $\bar{g} > 0$ , the

<sup>11</sup>As may perhaps be intuited from Figure 1 and seen more precisely from skewness and excess kurtosis numbers in Table 7 (in Appendix A) this is a particularly good approximation in the case of the single- $\rho$  model. Indeed, as the table shows, in the case of the single- $\rho$  model the variance is virtually unaffected by ambiguity aversion, and so  $\widetilde{Var}_t(x_t) \approx \Omega$  (defined as in eq. (14)).

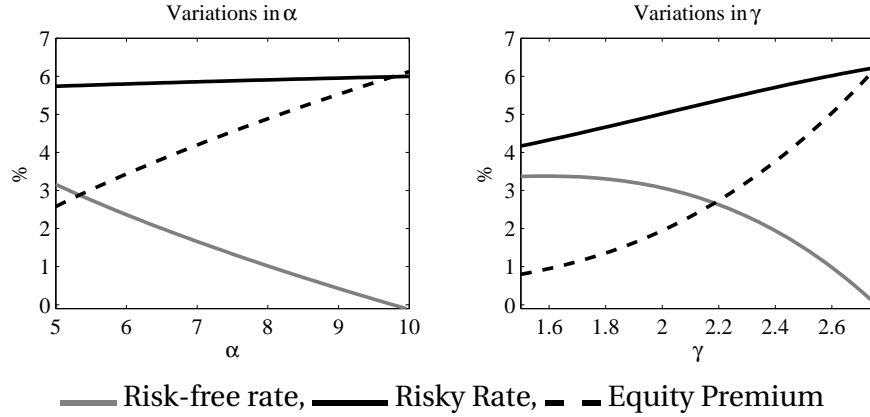


Figure 2: **Comparative statics (single- $\rho$ )**: In the left panel,  $\alpha$  varies with  $\gamma$  fixed at 2.5. In the right panel,  $\alpha$  was fixed at 7.24 and  $\gamma$  varies. The average comparative statics are constructed by first computing the comparative statics for each year using the filtered values  $\hat{x}_t$  and then averaging across  $t = 1978, \dots, 2011$ .

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 an increase in  $\gamma$  *increases* the risk-free rate. (Note, the comparative statics of risk aversion shown in Figure 2 correspond to  $\alpha \approx 7$ , i.e., in presence of ambiguity aversion.) This explains the (evidently) high risk free rate obtained in the Bayesian case. An increase in ambiguity aversion,  $\alpha$ , decreases  $\tilde{x}_t$  (see Figure 8), inducing a more pessimistic “as if” distribution (see Figure 8 in Appendix A) and making the agent behave as if she were expecting a lower endowment income in future (states); a larger  $\rho$  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. If EIS is low, as it is when  $\gamma > 1$ , there will be increased emphasis on offsetting the greater future consumption. All this is encapsulated in the  $\gamma\rho\tilde{x}_t$  term in (24) which shows an increase in ambiguity aversion implies a rise in demand for the risk-free 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, e.g., in Caballero and Krishnamurthy (2008).

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

$$E_t r_t = \text{Const}_1 + \rho (\gamma - \psi) \tilde{x}_t + \psi \rho \hat{x}_t - \frac{\rho^2}{2} [(\gamma - \psi)^2 \text{Const}_2] \widetilde{\text{Var}}_t(x_t) \quad (25)$$

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  $N(\hat{x}_t, \Omega)$  and the bivariate normal shown in (16), respectively.  $\text{Const}_1$  and  $\text{Const}_2$  collect terms which are both constant across time and not affected by ambiguity aversion. 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 (25) describes the *conditional* expectation of a Savage-Bayes rational observer/analyst who observes these prices and uses the same information as the agent to predict dividend at  $t + 1$ . The expression (25) 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  $\alpha$  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. The second effect is evident from the term  $-\rho\psi\tilde{x}_t$ : as  $\alpha$  increases  $\tilde{x}_t$  decreases, hence decreasing the (“as if”) expected future dividend payoff from the asset causing the agent to want to pay less for the asset. Taking out the common factor,  $\rho$ , the strength of the first effect depends on  $\gamma$  while the second effect is exacerbated by leverage,  $\psi$ . With  $\gamma \leq 3$  and  $\psi = 3$ , as we have here, the second effect dominates and equilibrium risky rate varies positively (but quite minimally) with ambiguity aversion. Since  $\widetilde{\text{Var}}_t(x_t)$  is essentially constant in the single- $\rho$  model ambiguity aversion does not affect risky rate through this route.

To get a first intuition for the results on returns volatilities, observe when  $\alpha \approx 0$  while  $\psi$  drops out as a coefficient of  $(\hat{x})^2$  in the expression for  $(Er)^2$  it is present in that form in  $Er^2$  and hence plays the more significant role in fixing the (average) volatility of the risky rate for the Bayesian and ambiguity averse case alike. The risk free rate is, of course, not affected by the volatility of the dividend claim, which explains its comparatively lower volatility.

Finally, the approximation for the equity premium may be written as

$$E_t r_t - r_t^f = \text{Const}_3 + \psi\rho(\hat{x}_t - \tilde{x}_t) + \frac{\rho^2}{2} [\gamma^2 - (\psi - \gamma)^2 \text{Const}_2] \widetilde{\text{Var}}_t(x_t). \quad (26)$$

where we have explicitly left the two terms which are affected by ambiguity aversion,  $(\hat{x}_t - \tilde{x}_t)$  and  $\widetilde{\text{Var}}_t(x_t)$ . The first term shows that the premium increases with ambiguity aversion (the difference  $(\hat{x}_t - \tilde{x}_t)$  increases when  $\alpha$  is increased) and the magnitude of this effect is accentuated by persistence and leverage. A doubt factor also comes into play since the premium is increasing in the (distorted) variance of the latent variable but only significantly when compared to the rational expectations case where the second-order belief is degenerate. Since conditionally the risk free rate is not random, the (conditional) volatility of equity premium is determined by that of the risky rate, discussed earlier. How-



	Data	Model	
		Single- $\rho$	Two- $\rho$
Excess Return	-0.10	0.56	-0.21
Equity Premium	-	0.07	-0.61
HP filtered $p - d$	0.03	0.06	0.04
$Var(x_t)$	-	-	-0.69
$\widetilde{Var}(x_t)$	-	0.43	-0.83

Table 2: **Countercyclicality of returns and pro-cyclicality of price-dividend ratio:** Correlation of data and model-implied excess returns, equity premium and price-dividend ratio with  $\hat{x}_t$ . (In the two- $\rho$  case, the correlation can be assumed to be with the filtered value coming from either the high- or the low-persistence model since the correlations are identical up to 2 decimal places.)

ever, as Table 1 shows, while the predicted second moment of the risk-free rate matches data very well, the single- $\rho$  model fails to predict about 50% of the volatility of the risky rate (and that of the equity premium).

The (log of) price dividend ratio,  $z$ , is approximated as shown in (27) below. The formulas for  $A_0, A_1$  are given by eqs. (40) and (39) in the Appendix. Notably, given our parameter values,  $A_1$  is positive and proportional to  $\rho(\psi - \gamma)$ .

$$z_{t+1} = A_0 + A_1 \tilde{x}_{t+1} \quad (27)$$

Table 2 shows the correlation between the filtered state,  $\hat{x}_t$ , and three series: the excess return, the equity premium and HP filtered  $p - d$  (where the filter parameter  $\lambda$  is set to 6.25 following the recommendation by Ravn and Uhlig (2002)). The excess return – the difference between the market return and the risk free rate – in the single- $\rho$  model is strongly pro-cyclical and differs substantially from the data, where it is weakly counter-cyclical. The equity premium, which is the conditional expected value of the excess returns, is also weakly pro-cyclical. Being an expectation, it is not directly observable in data, hence the missing entry; we only give the correlations for the equity premium implied/predicted by our model(s). The lack of cyclicality (and near constancy) is primarily driven by the constancy of the conditional variance of the “as-if” belief, as shown in the top panel of Figure 3. We interpret the lack of countercyclicality in the returns as a clear indication that the single- $\rho$  specification is inadequate as a model of (ambiguous) beliefs.

## 3.2 Uncertain persistence: the two- $\rho$ 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 and prices.

### 3.2.1 Beliefs with time-varying ambiguity

Now we extend the single- $\rho$  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 BY CASE I type specification with either a low persistence ( $\rho^l$ ) or a high persistence ( $\rho^h$ ) with probability  $\eta_t$  and  $1 - \eta_t$ , respectively. The two processes are:

$$\begin{aligned}
 & \textbf{Low Persistence } (\rho = \rho_l, Pr = \eta_t) \\
 x_{l,t+1} &= \rho_l x_{l,t} + \sigma_{x_l} \varepsilon_{x_l,t+1} \\
 d_{l,t+1} &= \bar{d} + \psi x_{l,t+1} + \sigma_{d_l} \varepsilon_{d_l,t+1} = \bar{d} + \psi (\rho_l x_{l,t} + \sigma_{x_l} \varepsilon_{x_l,t+1}) + \sigma_{d_l} \varepsilon_{d_l,t+1} \\
 g_{l,t+1} &= \bar{g} + x_{l,t+1} + \sigma_{g_l} \varepsilon_{g_l,t+1} = \bar{g} + \rho_l x_{l,t} + \sigma_{x_l} \varepsilon_{x_l,t+1} + \sigma_{g_l} \varepsilon_{g_l,t+1} \\
 & \textbf{High Persistence } (\rho = \rho_h, Pr = 1 - \eta_t) \\
 x_{h,t+1} &= \rho_h x_{h,t} + \sigma_{x_h} \varepsilon_{x_h,t+1} \\
 d_{h,t+1} &= \bar{d} + \psi x_{h,t+1} + \sigma_{d_h} \varepsilon_{d_h,t+1} = \bar{d} + \psi (\rho_h x_{h,t} + \sigma_{x_h} \varepsilon_{x_h,t+1}) + \sigma_{d_h} \varepsilon_{d_h,t+1} \\
 g_{h,t+1} &= \bar{g} + x_{h,t+1} + \sigma_{g_h} \varepsilon_{g_h,t+1} = \bar{g} + \rho_h x_{h,t} + \sigma_{x_h} \varepsilon_{x_h,t+1} + \sigma_{g_h} \varepsilon_{g_h,t+1}
 \end{aligned} \tag{28}$$

We call this the two- $\rho$  *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  $\eta_t$ , the posterior probability that the low persistence model is the “true” data generating process (DGP), and takes the form:

$$V(C_t; \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) = u(C_t) + \beta \phi^{-1}(\mathcal{V}_{t+1}), \tag{29}$$

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

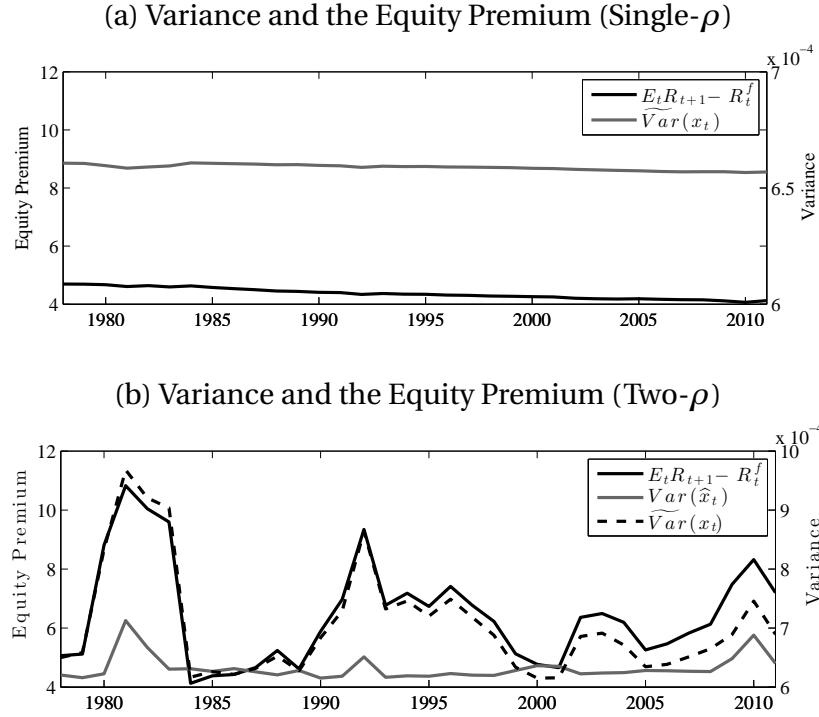


Figure 3: **Movements in variance and model implied equity premium:** The top panel shows the model-implied conditional equity premium and the conditional variance of the “as-if” posterior from the single- $\rho$  model. The bottom panel shows the conditional equity premium, as well as the variance of the posterior and “as-if” posterior from the two- $\rho$  model.

$\rho = \rho_j$ , 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 data generating process. See Appendix B for further details, including the formulas for rates of return.

The value of  $\rho_h$  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), based on the empirical arguments given in the LRR literature. The value of  $\rho_l$  was motivated by Beeler and Campbell (2009) and Constantinides and Ghosh (2010) who argue 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 BY calibration. Constantinides and Ghosh (2010) provide a GMM estimate (based on the years 1931-2006) of  $\rho = 0.32$  (see their Table 4). We set  $\rho_l = 0.30$ , although (we found) values

between 0.25 and 0.40 have virtually identical posteriors (and implications for rates of returns). Using this value for  $\rho_l$ , the posterior probability against the model with  $\rho_h = 0.85$  is approximately 50% in 1977, the beginning of the model evaluation period, and the posterior probability  $\eta_t$  is consistently in the interval  $[0.3, 0.7]$  throughout the period 1978-2011, demonstrating how difficult it is to separate the two models on the basis of growth data.<sup>12</sup> To summarize, one substantial strand of literature (following on from BY) argues there is strong empirical justification for assuming a high value of  $\rho$  in beliefs of representative agents in asset pricing models, while another points out that pure consumption growth data suggests a more moderate value, and it is generally agreed the estimates are quite fragile. As just noted, the  $\rho = 0.30$  model has about as much support as the  $\rho = 0.85$  model in the (growth) data, and hence it seems only appropriate that an agent who is uncertain about the latent state, an uncertainty whose evolution depends significantly on the belief about the value of persistence, treats beliefs about the correct  $\rho$  as ambiguous subject to updating. This is one sense (as also argued by Hansen and Sargent (2010)) in which the two- $\rho$  model, with the parameter values we adopt, is empirically more compelling than the assumption of a dogmatic belief in some value of  $\rho$ . All parameter estimates are presented in Appendix C.

There is another way in which the two- $\rho$  model improves, empirically, on the single- $\rho$  model: by introducing endogenously varying uncertainty of *beliefs*. Uncertainty about persistence leads to time-varying mixing of the two models (of persistence) through  $\eta_t$ , 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.

---

<sup>12</sup>Choosing  $\rho_l$  to be very small ( $\approx 0$ ) so that the low persistence model is virtually i.i.d. produces posterior estimates of  $\eta_t$  near zero, i.e., the two- $\rho$  model behaves almost as the single- $\rho$  model.

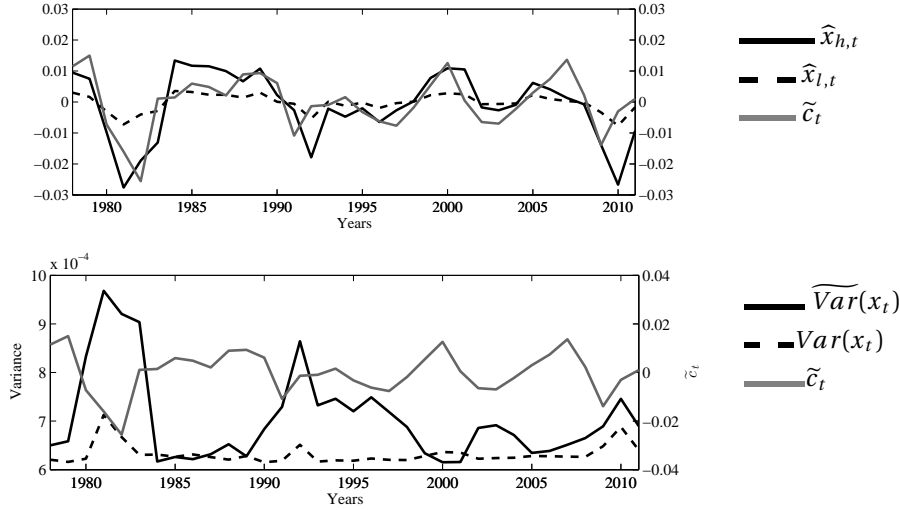


Figure 4: **Explaining time-varying ambiguity:**The upper panel shows the filtered latent variables assuming that the high ( $\hat{x}_{h,t}$ ) and low ( $\hat{x}_{l,t}$ ) persistence states were the DGP. In the lower panel, the dashed line graphs the conditional variance of the latent state variable ( $Var_t(x_t)$ ) and the solid line the “as-if” conditional variance ( $\widetilde{Var}_t(x_t)$ ). In both panels the gray line shows the HP-filtered consumption growth.

The beliefs about the latent state, conditional on the low and high persistence models being true, are  $N(\hat{x}_{l,t}, \Omega_l)$  and  $N(\hat{x}_{h,t}, \Omega_h)$ , respectively, and the variance of the mixture distribution of the latent state is,

$$\eta_t \Omega_l + (1 - \eta_t) \Omega_h + \eta_t (1 - \eta_t) (\hat{x}_{h,t} - \hat{x}_{l,t})^2. \quad (30)$$

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.,  $\eta_t$  is close to 0.5) and yet their core forecasts markedly disagree (i.e.,  $(\hat{x}_{h,t} - \hat{x}_{l,t})^2$  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. 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 countercyclical. Thus, the two- $\rho$  model of beliefs embodies a theory of why and how ambiguity about growth prospects may vary over time.<sup>13</sup>

<sup>13</sup>The 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 under-

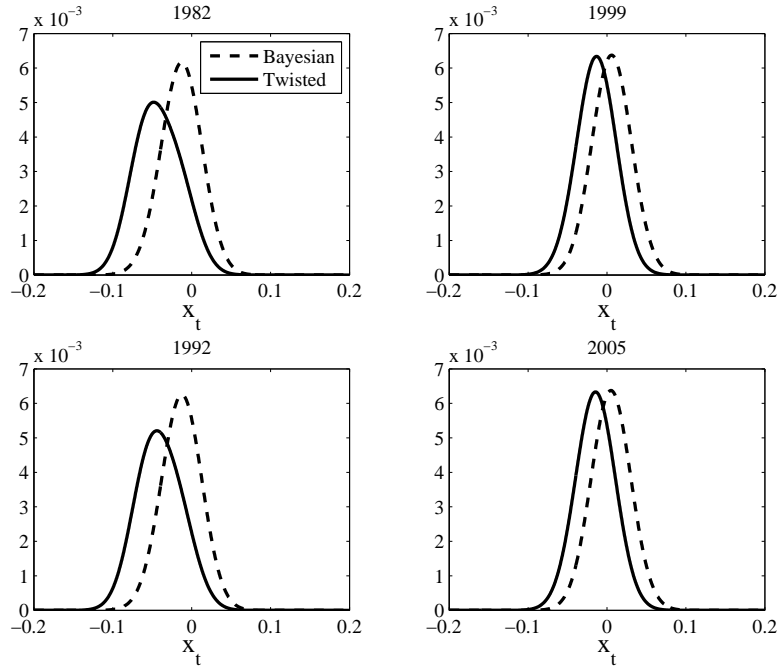


Figure 5: **Time-varying distortion:** These four panels contain plot beliefs about the latent state without ambiguity aversion (Bayesian) and with ambiguity aversion (both from the two- $\rho$  model). The left panels picture “bad” years where  $\hat{x}_{h,t} < \hat{x}_{l,t}$ , and the two right panels show “good” years.

### 3.2.2 The asymmetric effect of ambiguity aversion over the business cycle

Ambiguity aversion exacerbates the time-variation of the uncertainty. The Bayesian mixture is a distribution with excess kurtosis relative to a normal but the change of measure, the “twist”, transforms the small increase in kurtosis into substantial negative skewness, while increasing the variance. Table 7 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: the significant changes and movement in the variance of the as if (distorted) posterior along the time-path of information sets in the sample period. This variation was almost completely absent in the single- $\rho$  case, as is evident in Figure 3.

Figure 4 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–

---

pinning the exogenous specification of stochastic volatility) is not well supported empirically. In contrast, the time-varying heteroskedasticity generated endogenously in the two- $\rho$  model is a *forecast* uncertainty, of beliefs, empirically driven by the history of growth outcomes and consistent with a stationary volatility of consumption shocks.

2011. It is movements in these state variables, especially their disagreement, which is the source of variation in uncertainty. The lower panel depicts time-series of the variance of the posterior (eq. (30)) 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  $(\hat{x}_{h,t} - \hat{x}_{l,t})^2$  rises *and*  $\hat{x}_{h,t}$  falls below  $\hat{x}_{l,t}$ , as it does following a negative shock. To see why, consider the following. Loosely speaking, we can think of the overall distorted posterior as a weighted mixture of two component distorted posteriors,  $\xi_t^i \otimes N(\hat{x}_{i,t}, \Omega_i)$  for  $i = h, l$ , where  $\xi_t^i$  is as in eq. (19) with  $\hat{x}_{i,t+1}$  replacing  $\hat{x}_{t+1}$  on the right hand side of the equation. Let  $\tilde{x}_{i,t}$  denote the mean of a distorted component posterior. Due to the greater persistence, the aggregate uncertainty around  $\hat{x}_h$  – captured by  $\Omega_h$  – is larger than that around  $\hat{x}_l$ . As a result, the magnitude of distortion is greater in the high persistence model, i.e.,  $\hat{x}_h - \tilde{x}_h > \hat{x}_l - \tilde{x}_l$ . Which means that  $(\hat{x}_{h,t} - \hat{x}_{l,t})^2$  is smaller when  $\hat{x}_h > \hat{x}_l$  than when  $\hat{x}_h < \hat{x}_l$ , and the squared difference between the means of the distorted distribution is an important component in the variance of a mixed “as-if” posterior (see eq. 30). This asymmetry also makes the skewness of the mixed “as-if” posterior more negative since the two components are closer to each other in good times and further apart in bad times. Hence, the “as if” posterior of the ambiguity averse agent exaggerates the volatility in the Bayesian posterior in a way that makes it more pronouncedly countercyclical (Table 2).

Figure 4 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 1999 and 2005  $\hat{x}_{h,t} > \hat{x}_{l,t}$ . In all four years the absolute difference between  $\hat{x}_h$  and  $\hat{x}_l$  was similar. However, it is only in 1982 and 1992 that  $\hat{x}_{h,t} < \hat{x}_{l,t}$ . Though facing approximately the same forecast uncertainty in all four years, the ambiguity averse agent is more apprehensive about the uncertainty in 1982 and 1992. This leads to a larger twist over probabilities in the left tail, yielding a higher variance, a prominent left skew and excess kurtosis; compare the Bayesian and twisted predictive distributions in Figure 5.

### 3.2.3 Results

The top panel of Table 3 reports the two- $\rho$  model implied conditional moments. The level of ambiguity aversion was again calibrated so that the risk-free rate was 1.5%. Compared with the single- $\rho$  model (Table 1), we see the model’s match of the first moments is now quite perfect and there is a substantial increase in the magnitude of the predicted second moments. The table also shows results of robustness checks where assumed val-

ues of persistence, leverage and discount rate were varied (separately). A Bayesian agent also sees the increase in the volatility of the risky rate and equity premium in the two- $\rho$  specification, but the equity premium is tiny.

Results on price-dividend ratio in the two- $\rho$  model, in the bottom panel, show an improvement over the single- $\rho$  model, similar to that seen in the case of returns. This adds to the findings in Table 2, which show that the weak pro-cyclicality of  $p - d$  in data is closely replicated in the model implied series in the two- $\rho$  case. A common observation concerning excess returns is that they tend to mean revert over long horizons. Applying a statistic used in the literature (see, e.g. Guvenen (2009)) that aggregates consecutive autocorrelation coefficients of excess returns from the U.S. data in our 1978-2011 sample, we find a strong pattern of mean reversion, shown in the second row in Table 5. The third row displays the model counterparts of this measure of mean reversion, which are consistent with the signs and rough magnitudes of these statistics in the data. Such mean reversion is a clear departure from the martingale hypothesis of returns and is sometimes linked to the predictability of returns. However, both in the data (in our 1978-2011 sample) and in the model implied time series, we found returns are only very weakly predicted by price-dividend ratio (though, as is evident in the final two columns in the bottom panel of Table 3, the persistence of  $p - d$  found in the data is very well predicted by the model). Figure 6 confirms the comparative statics of returns are all qualitatively very similar to those from the single- $\rho$  model.

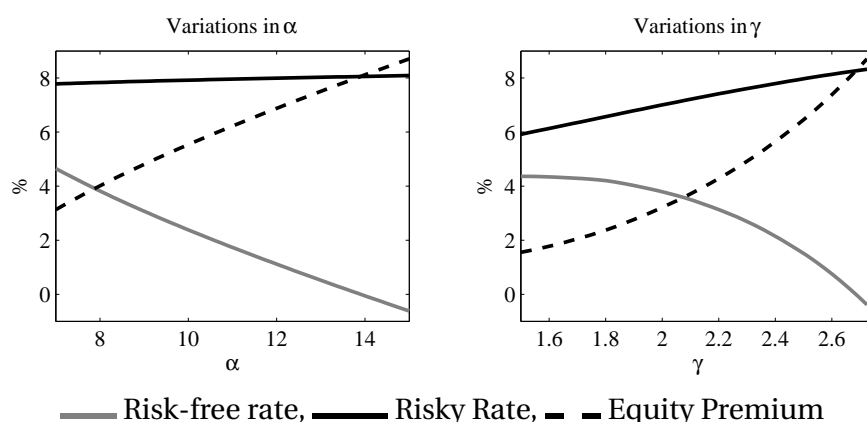


Figure 6: **Comparative statics (two- $\rho$ )**: In the left panel,  $\alpha$  varies with  $\gamma$  fixed at 2.5. In the right panel,  $\alpha$  was calibrated at 11.3 and  $\gamma$  varies. The average comparative statics are constructed by first computing the comparative statics for each year using the filtered values  $\hat{x}_t$  and then averaging across  $t = 1978, \dots, 2011$ .

What accounts for the significant improvement in the match with data compared to



Returns and Volatility								
	$\gamma$	$\alpha$	$E(r)$	$E(r - r^f)$	$\sigma(r^f)$	$\sigma(r)$	$\sigma(r - r^f)$	
Data			8.08	6.68	2.20	16.5	16.1	
	1.0	31.5	6.61	5.08	1.20	22.2	22.2	
	2.0	17.8	7.36	5.85	2.58	23.0	23.0	
	2.5	11.3	7.97	6.46	3.29	23.55	23.6	
	3.0	6.65	8.66	7.14	3.96	24.17	24.2	
Robustness Checks								
$\rho_h = 0.90$	2.5	7.30	7.88	6.36	3.83	23.5	23.6	
$\rho_l = 0.25$	2.5	11.1	7.98	6.48	3.05	23.7	23.7	
$\psi = 2.50$	2.5	11.3	7.58	6.07	3.15	23.6	23.5	
$\beta = 0.965$	2.5	13.0	9.15	7.62	3.44	23.8	23.8	
$\beta = 0.97$	2.5	12.2	8.56	7.05	3.36	23.7	23.7	
Bayesian	2.5	$\approx 0$	7.62	0.62	1.70	23.1	23.2	
Price-Dividend Ratio								
	$\gamma$	$\alpha$	$\mathbb{E}(P/D)$	$\sigma(P/D)$	$\mathbb{E}(p - d)$	$\sigma(p - d)$	AC1	AC2
Data			45.513	19.954	3.724	0.445	0.803	0.759
	1.0	31.5	29.3	4.34	3.37	0.15	0.51	0.48
	2.0	17.8	32.3	5.92	3.46	0.19	0.65	0.60
	2.5	11.3	44.0	14.5	3.73	0.34	0.85	0.78
	3.0	6.65	52.9	22.2	3.88	0.43	0.88	0.81
Robustness Checks								
$\rho_h = 0.90$	2.5	7.30	42.9	13.7	3.71	0.33	0.84	0.78
$\rho_l = 0.25$	2.5	11.1	44.3	14.8	3.74	0.35	0.85	0.78
$\psi = 2.5$	2.5	11.3	39.6	11.1	3.64	0.29	0.82	0.75
$\beta = 0.965$	2.5	13.0	59.9	28.1	3.98	0.49	0.89	0.82
$\beta = 0.97$	2.5	12.2	51.3	20.6	3.86	0.42	0.88	0.81
Bayesian	2.5	$\approx 0$	40.0	11.5	3.65	0.30	0.82	0.75
Bayesian, $\beta = .97$	2.5	$\approx 0$	46.3	16.5	3.77	0.37	0.86	0.79

Table 3: **Results (two- $\rho$ ):** The top panel contains the average of the predicted *conditional* moments of rates of return (on dividend claim) in the two- $\rho$  model for different values of  $\gamma$  and calibrated  $\alpha$ . Immediately below is a series of robustness checks where the parameter in the left-most column was changed from the basic specification ( $\rho_h = 0.85$ ,  $\rho_l = 0.3$ ,  $\psi = 3$ ,  $\beta = 0.975$ ). The bottom panel contains the average of the price/dividend ratio in the two- $\rho$  model over the period 1978–2011. AC1 and AC2 denote the first and second order autocorrelation of  $p - d$ .

the single- $\rho$  case? The single- $\rho$  approximations in eqs. (24), (25), (26) give important clues when taken together with some observations. The first observation is that in the two- $\rho$  model, the filtered latent state  $\hat{x}$  is less pessimistic, overall, since the exaggerated downward plunges in its value when assuming high persistence is moderated by the more modest swings obtained by assuming lower persistence, as may be seen in (top panel of) Figure 4. The second observation is the increase in the “as if” volatility,  $\widetilde{Var}_t(x_t)$ , which is evident in Figure 3 (discussed in subsection 3.2.2). Finally, for purposes of fixing intuition, it is helpful to note that the rates of risk free and risky returns in this model are close to being a weighted average of the corresponding rates from a high and a low persistence model (see Appendix B2). Because the (average)  $\hat{x}$  is less pessimistic, the two- $\rho$  case requires a comparatively greater calibrated value of  $\alpha$  to ensure that  $\tilde{x}$  is low enough to generate a risk free rate of 1.5%. However, to make the calibration,  $\tilde{x}$  is not required to be as low as it had to be in the single- $\rho$  case, since the higher  $\widetilde{Var}_t(x_t)$  helps to keep the risk free rate down. Ambiguity aversion does not significantly affect the risky rate due to the two countervailing effects mentioned in the discussion following eq. (25); notice the moments of the risky rate for the Bayesian case. The fact that the implied risky rate is greater (than in the single- $\rho$  case) is primarily due to the fact that  $\hat{x}_t$  is less pessimistic and correspondingly, the risky rate in the low persistence model is a lot larger. As is evident from the result in the Bayesian case, the increase in volatility of the risky rate has far more to do with the move from single to two- $\rho$  than ambiguity aversion per se. Indeed, putting together the discussion of volatility following eq. (25), and formula in (30) we see the variance of the risky rate in this model would be significantly determined by the term  $\psi^2(\hat{x}_{h,t} - \hat{x}_{l,t})^2$ . However, compared to  $\psi^2(\hat{x}_{h,t} - \hat{x}_{l,t})^2$ ,  $\psi^2(\tilde{x}_{h,t} - \tilde{x}_{l,t})^2$  will exaggerate the counter-cyclical variation, for reasons explained in the previous sub-section. So, even though ambiguity aversion may not affect the volatility of risky rate very much on average, it does shape its time-profile. The average level of the equity premium is largely determined by the term  $(\hat{x} - \tilde{x})$ , which is now greater since the calibrated value of  $\alpha$  has risen.<sup>14</sup>

What is most dramatically different from the single- $\rho$  model is the time variation of expected returns as may be seen in Figure 3. The bottom panel in the figure demonstrates that the dynamics of equity premium predicted by the two- $\rho$  model are strongly influenced by movements in  $\widetilde{Var}_t(x_t)$ . In the case of the single- $\rho$  model, shown in the top

<sup>14</sup>However, this discussion based on the approximation equations does not give the complete explanation since, as was noted, in the two- $\rho$  case moments higher than the second order come into play which are not taken into account in the approximations.

$\gamma$	$\alpha$	$C$	$D$
Data		0.74	0.39
1.00	31.5	0.55	0.08
2.00	17.8	0.87	0.54
2.50	11.4	0.93	0.65
3.00	6.7	0.95	0.72

Table 4: **Price-dividend ratio correlations (two- $\rho$ ):** Average correlation of the price/dividend ratio with consumption and dividends (in logarithms) in the two- $\rho$  model over the period 1978–2011.

Cumulative Autocorrelation					
Lag, in years	1	2	3	4	5
Data	-0.16	-0.30	-0.32	-0.79	-0.33
Model implied returns	-0.54	-0.35	-0.58	-0.76	-0.52

Table 5: **Mean reversion of returns:** Autocorrelation structure of excess returns in the data and as implied by the two- $\rho$  model (baseline specification). The cumulative autocorrelation is defined as  $\left(\sum_{i=1}^j Corrl((R_t - R_t^f), (R_{t-i} - R_{t-i}^f))\right)$ .

panel, there is virtually no time-variation, neither in the equity premium nor in  $\widetilde{Var}_t(x_t)$ .

The price-dividend ratio carries essentially the same information as the return on the dividend claim. As we noted in the case of risky return, here too the degree of ambiguity aversion has minimal effect (because of the two offsetting effects): it is driven very substantially by the filtered value of the hidden state, which causes it to be pro-cyclical. The lower pessimism of  $\hat{x}$  here too acts to lift the price-dividend ratio, while the larger  $\widetilde{Var}_t(x_t)$  lifts its second moment, while the AC1 and AC2 match reflects the persistence of beliefs about the filtered state.

While the time variation of the conditional equity premium shown in Figure 3 may be impressive, we do not have a comparable time series in data since equity premium is not directly observed. However, we do observe the realized risky rate, risk free rate, the realized excess return (the difference between the two) and the price-dividend ratio. Figure 7 plots these and the corresponding series implied by the model (each point shows the value of the variable forecast by the model at a date given the information set at *that* date). This sets out a stark, stiff test for the model. We are not aware of comparable graphs in the literature, possibly because most model predictions are not derived as

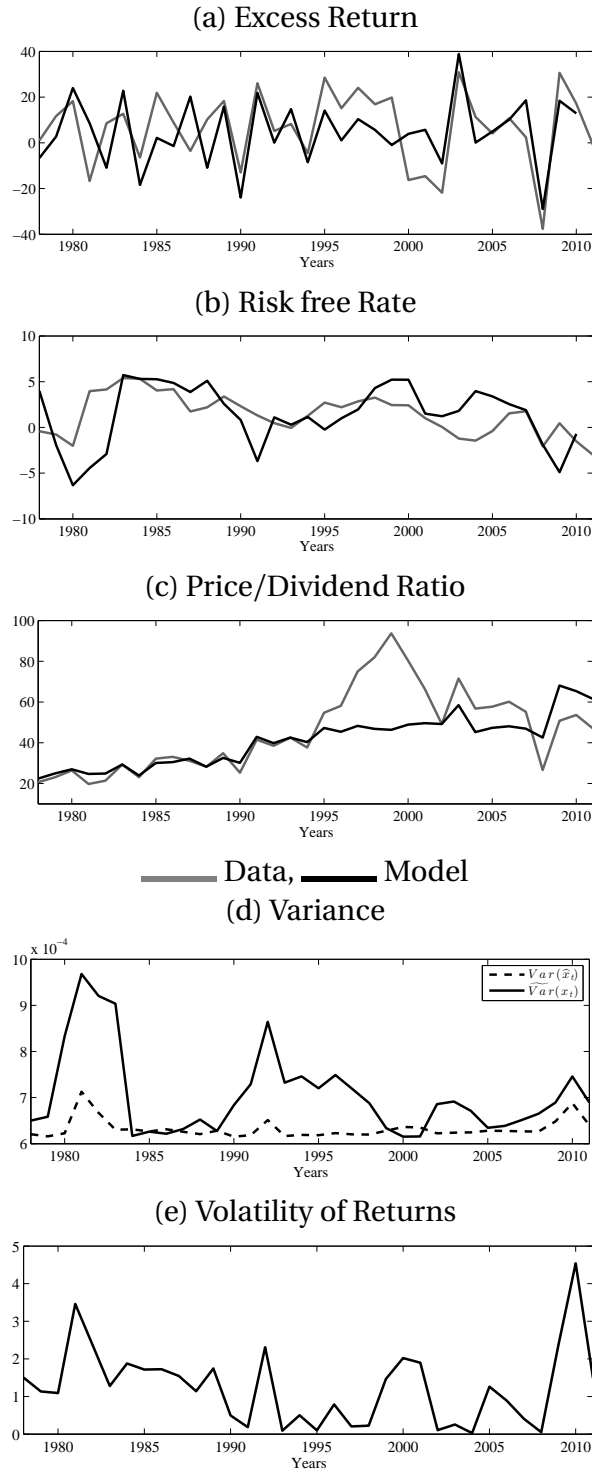


Figure 7: **Returns and Price-Dividend Ratio:** Panel (a) contains a plot of the model-implied excess return along with the actual excess return. Panel (b) shows the model-implied risk-free rate along with the actual real risk-free rate. Panel (c) contains the actual and model implied price-dividend ratios. Panel (d) plots the variance of the mixture and the “as-if” variance of the mixture. Panel (e) graphs the (two- $\rho$ ) model implied times series of  $\sqrt{E_t (R_{t+1} - ER_{t+1})^2}$ .

conditional expectations at information sets on the observed history.<sup>15</sup> Panel (c) of Figure 7 reproduces the  $\widetilde{Var}_t(x_t)$  graph shown earlier, just to see in a single glance the role this variable, embodying the as if (conditional) uncertainty of the ambiguity averse agent, plays in the model predictions. The predictions are evidently good, especially for returns but reasonably good too for the price-dividend ratio. Indeed, the correlation of the linearly detrended (in logs), HP-filtered (in logs) and “raw” (i.e., unfiltered) predicted ratio and the correspondingly treated price-dividend observed in data are 0.67, 0.77 and 0.83, respectively. However, we cannot match the period between 1995 and 2000 which corresponds to the dot-com bubble (see, e.g., Kraay and Ventura (2007)). This is only to be expected; ours is a simple Lucas tree exchange economy, with prices determined in general equilibrium entirely based on the stochastic evolution of *real* output. In this respect, it is significant that the actual price-dividend returns to the predicted path following the collapse of the bubble. The effects of the recent recession are captured too, though not as well as in the returns predictions. The performance of the model seems remarkable given the simplicity of input; annual, aggregate, output realizations. This a very coarse and discrete model of evolution of signals compared to what is evident in the real world.<sup>16</sup> These results show that observed *movements* in asset returns and prices can be substantively explained simply on the basis of aggregate macroeconomic risk, conditional on aggregate uncertainty grounded in actual historically observed public signals.

A fact of significant interest to financial economics is that stock return volatility varies quite a lot over time. A related, intriguing, stylized fact is the countercyclicity of the return volatility; that, it tends to be higher in recessions. Our model predicts this countercyclicity. Panel (e) of Figure 7 plots the model implied times series of the (square root of) conditional expectation of the deviation of the rate of return from its unconditional<sup>17</sup> mean, with the conditional expectation taken at information sets along the observed history. This prediction is testable; but, testing it would require higher frequency data (e.g., monthly, see Figure 1 in Veronesi (1999)). Veronesi’s model gives a qualitative prediction of conditional volatility of returns turning on the idea that investors tend to be more uncertain about the future growth rate of the economy during recessions. While the expla-

---

<sup>15</sup>Figures in section 7 in Hansen and Sargent (2010) show time series of conditional expectations of several variables in their theory but not rates of returns and price-dividend ratio and nor any comparisons with observed time series in data; Figure 3 in Ju and Miao (2012) shows the conditional returns but does not compare with data.

<sup>16</sup>The fact that we use annual data inevitably makes the time alignment across variables rather imperfect, which needs to be taken into account when reading the graphs.

<sup>17</sup>More precisely, the unconditional mean  $ER_{t+1} \equiv T^{-1} \sum_{t=1}^T R_t$ , where  $R_t$  is as implied by the model given the observed history growth outcomes up to  $t$ .

nation in our model is in similar spirit, it is different in that it applies ambiguity aversion and ambiguity about the persistence and extent of the (hidden) temporary shocks to fundamentals explicable on the basis of observed history to make a quantitative prediction (note the closeness of this plot to the plot of  $\widetilde{Var}_t(x_t)$ ).

## 4 Interpreting the magnitude of the ambiguity aversion

Here we discuss some ways of understanding whether the calibrated levels of ambiguity aversion that may be regarded as plausible in terms of implied individual (as opposed to market) behavior. Specifically, we consider two thought experiments in a model of an agent with preferences as in this paper. In each experiment the agent evaluates an uncertain consumption prospect and identify part of that evaluation as the risk premium and part as the ambiguity premium. We then compare the two premia across pairs of values of  $(\gamma, \alpha)$ . Years of conceptual familiarity, application in economic modeling, reams of empirical studies and casual introspection informs us what extent of risk aversion is plausible and typically acceptable in applied economic models. Given an uncertain prospect, we compare ambiguity premia with  $\alpha$  set in the range considered in our calibrations with risk premia obtained for values of  $\gamma$  thought to be plausible. If the ambiguity premia were of an order similar to the risk premia we think there is basis to argue that our calibrated values were plausible. To simulate the domain modeled in this paper, each thought experiment considers circumstances which involve uncertainty similar to that faced by the agent in our asset pricing model. In the first experiment the agent faces an uncertain consumption one period ahead but not in subsequent periods. In the second, the agent evaluates an entire Lucas tree, involving dynamic uncertainty, precisely like the agent in our model.

In the first thought experiment, an agent at time  $t$  faces an uncertain consumption prospect with a one-off risk:  $C_{t+1} = \bar{C} \exp(g)$ , where  $g \sim F(g|x)$  and  $x \sim F(x)$ , with  $C_{t+n} = \bar{C}$ ,  $n = 2, 3, \dots$ . For instance, in the case of the single- $\rho$  model, we have

$$F(g|x) \equiv N(\rho x + \bar{g}, \sigma_x^2 + \sigma_g^2), \quad F(x) \equiv N(\hat{x}, \Omega), \quad (31)$$

and in the case of the two- $\rho$  we have,

$$F(g|x_i) \equiv N(\rho x_i + \bar{g}, \sigma_{x_i}^2 + \sigma_{g_i}^2), \quad i = h, l, \quad F(x) \equiv \begin{cases} N(\hat{x}_l, \Omega_l) & \text{with probability } \eta \\ N(\hat{x}_h, \Omega_h) & \text{with probability } 1 - \eta \end{cases}. \quad (32)$$

Following eqs. (15), (29), utility from this prospect with one-off uncertainty is:

$$V^A \equiv \beta \left( \phi^{-1} \left( \int \phi \left( \int u(C_{t+1}) dF(g|x) \right) dF(x) \right) + \frac{\beta \bar{u}}{1-\beta} \right) \quad (33)$$

where,  $u(\bar{C}) \equiv \bar{u}$ . The consumption (certainty) equivalent is then given by  $\mathcal{C}^A = u^{-1} \left( \frac{V^A}{\beta} \right)$ . Analogously, letting  $\phi$  be the identity function in eq. (33), we define  $V^B$ , the value to a Bayesian agent and define the corresponding certainty equivalent,  $\mathcal{C}^B$ , by substituting  $V^B$  in place of  $V^A$ . Finally, letting  $\phi$  and  $u$  be identities in eq. (33) we define  $V^C$ , the value to a risk and ambiguity neutral agent and thus define  $\mathcal{C}^C$ , the discounted expected sum. The risk premium is  $\mathcal{R} = \mathcal{C}^C - \mathcal{C}^B$ , and the ambiguity premium is  $\mathcal{A} = \mathcal{C}^B - \mathcal{C}^A$ .

For the second thought experiment consider the full Lucas tree, the uncertain prospect actually considered by our agents and evaluated per eqs. (15) and (29), for the single- $\rho$  and two- $\rho$  models, respectively. The consumption certainty equivalents and risk and ambiguity premium are then defined in directly analogous fashion.<sup>18</sup>

Table 6 reports the relative premium for each of the two prospects, the ratio  $\mathcal{A}/\mathcal{R}$ , corresponding to beliefs in the two- $\rho$  model, averaged over the set of  $\hat{x}$  obtained on the actual sample path. The  $\gamma = 2, 2.5, 3$  and  $\alpha$  accordingly calibrated to obtain a risk free rate of 1.5, per our usual practice. So, as the  $\gamma$  increases, the ratio falls since the calibrated  $\alpha$  decreases. For each prospect, we see the ambiguity premium is about the same, mostly less, than the risk premium, when using levels of relative risk aversion that are widely regarded as plausible. This gives us a good idea of what the magnitude of  $\alpha$  implies behaviorally in the context of our model. Cubitt, van de Kuilen, and Mukerji (2012) report measurements with Ellsberg style experiments assuming smooth ambiguity preferences, and estimate the average magnitude of ambiguity premium to be similar to that of risk premium (for the same uncertain prospect).<sup>19</sup>

Table 6 also shows the relative premia for different values of the discount factor,  $\beta$ . In the case of the prospect in the first experiment, we see the relative premia remain similar

---

<sup>18</sup>In the case of this Lucas tree prospect, the value maybe equivalently computed without involving the  $\phi$  directly by applying a distorted posterior, with precisely the same distortion that was obtained in our model(s). However, the same distortion would not work when we evaluate the one-off stochastic prospect, since the as if belief that applies depends on the act being evaluated; the distortion depends on the consumption plan being considered, in particular, the associated continuation utility (see, e.g., eq. (19)). And, for the one-off prospect the continuation value is not the same as that for the full tree.

<sup>19</sup>Epstein (2010) suggests two Ellsberg (1961)-style thought experiments and argues that they pose difficulties for the smooth ambiguity model. In particular, he claims that efforts to calibrate an individual's  $\phi$  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 (2012) 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.

$\gamma$	$\alpha$	One-off increase in $C_t$	
		$\mathcal{A}/\mathcal{R}$	Full tree $\mathcal{A}/\mathcal{R}$
$\beta = 0.975$			
2.0	17.8	0.85	–
2.5	11.3	0.28	–
3.0	6.65	0.09	–
$\beta = 0.965$			
2.0	20.5	0.99	0.69
2.5	13.0	0.32	0.66
3.0	7.6	0.10	0.62
$\beta = 0.96$	13.9	0.35	0.87
$\beta = 0.95$	15.4	0.38	1.18

Table 6: **Risk and ambiguity premia in thought experiments:** Values of  $\mathcal{A}/\mathcal{R}$  in the two experiments with two- $\rho$  beliefs, averaged over the sample path. For  $\beta = 0.95, 0.96$  the  $\gamma = 2.5$ .

as the discount factor is varied. The same is true for the Lucas tree prospect, though for this prospect we are unable to report the value for  $\beta = 0.975$ . The reason for this is that the discounted expected value of the tree (i.e., valuation by an agent who is *both* risk and ambiguity neutral) does not exist (though it does for risk averse agents with  $\gamma \geq 0.16$ ) and hence, neither does  $C^C$ , so it is not possible to compute  $\mathcal{R}$ . The discount rate applied as the baseline parametrization in our models is  $\beta = 0.975$ , since this is standard practice in the finance literature, e.g., Bansal and Yaron (2004) and Hansen and Sargent (2010).<sup>20</sup> However, as noted in the robustness checks in Table 3, the results we obtain with  $\beta = 0.965$  are very similar.

## 5 Concluding remarks

We have found conditional (macroeconomic) uncertainty can explain levels, volatility and dynamics of asset returns and prices very substantially. Our model applied three links to establish this connection: Uncertainty and learning (in the sense of having an evolving, non-degenerate belief) about persistent hidden states describing temporary shocks to fundamentals; uncertainty and learning about the level of persistence; treating both these uncertainties as ambiguous and incorporating a plausible level of aversion to am-

<sup>20</sup>Note, in these papers, with expected value preferences, the models are identical to our single- $\rho$  and two- $\rho$  cases, respectively, and the same remark with regard to non-existence of value applies.



biguity. The first two elements, compatible with a Bayesian agent (but not with rational expectations), were enough to explain quite substantially the average volatility of returns and prices, and also the level of risky rate. Ambiguity aversion was important in explaining the levels of risk free rate and equity premium, and for shaping the time profile, the dynamics, of all the variables, especially the equity premium, on the basis of an endogenously varying conditional uncertainty. What was perhaps more striking was the simplicity of the framework and minimality of the departure from expected utility that was sufficient to capture so many aspects of returns data. That suggests the approach in this paper may be fruitfully applied to other domains of macro-finance research where effects of endogenously time-varying uncertainty are of interest.

## 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., AND M. WEBER (1992): "Recent Developments in Modelling Preferences: Uncertainty and Ambiguity," *Journal of Risk and Uncertainty*, 5, 325–70.
- CAMPBELL, J. (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104, 298–345.
- CAMPBELL, J., AND J. COCHRANE (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107(2), 205.
- 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.
- 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," Working paper 16618, Chicago Booth GSB and NBER.
- CUBITT, R., G. VAN DE KUILEN, AND S. MUKERJI (2012): "Sensitivity towards Ambiguity: A Qualitative Test and a Measurement," Discussion paper, University of Oxford.
- 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. (2010): "A Paradox for the "Smooth Ambiguity" Model of Preference," *Econometrica*, 78(6), 2085–2099.
- 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., AND S. ZIN (1989): "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework," *Econometrica: Journal of the Econometric Society*, 57, 937–969.
- FERNANDEZ-VILLAYERDE, J., AND J. RUBIO-RAMÍREZ (2010): "Macroeconomics and Volatility: Data, Models, and Estimation," Working Paper 16618, University of Pennsylvania and NBER.
- 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. (2011): "Portfolio choices and asset prices: The comparative statics of ambiguity aversion," *The Review of Economic Studies*, 78(4), 1329–1344.
- GUVENEN, F. (2009): "A Parsimonious Macroeconomic Model for Asset Pricing," *Econometrica*, 77(6), 1711–1750.
- HANSEN, L. (2007): "Beliefs, Doubts and Learning: Valuing Macroeconomic Risk," *American Economic Review*, 97(2), 1–30.
- HANSEN, L., AND T. SARGENT (2010): "Fragile beliefs and the price of uncertainty," *Quantitative Economics*, 1(1), 129–162.
- 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 (2012): "Ambiguity, Learning, and Asset Returns," *Econometrica*, 80(2), 559–591.
- JUDD, K. (1992): "Projection methods for solving aggregate growth models\* 1," *Journal of Economic Theory*, 58(2), 410–452.
- (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.

- (2009): “Recursive smooth ambiguity preferences,” *Journal of Economic Theory*, 144(3), 930–976.
- (2012): “On the Smooth Ambiguity Model: A Reply,” *Econometrica*, 80(3), 1303–1321.
- KNIGHT, F. (1921): *Risk, Uncertainty, and Profit*. Houghton Mifflin Company, Boston, MA.
- KRAAY, A., AND J. VENTURA (2007): “The Dot-Com Bubble, the Bush Deficits, and the U.S. Current Account,” in *G7 Current Account Imbalances: Sustainability and Adjustment*, pp. 457–496. National Bureau of Economic Research Inc.
- KREPS, D., AND E. 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,” Discussion paper.
- 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.
- SHEPHARD, 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.
- STRZALECKI, T. (2012): “Temporal Resolution of Uncertainty and Recursive Models of Ambiguity Aversion,” Discussion paper, Harvard University Working Paper.
- UHLIG, H. (2010): “A model of a systemic bank run,” *Journal of Monetary Economics*, 57(1), 78–96.
- VERONESI, P. (1999): “Stock market overreactions to bad news in good times: a rational expectations equilibrium model,” *Review of Financial Studies*, 12(5), 975–1007.
- 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.

# Appendix

## A An analytical approximation for rates of return in the single- $\rho$ model

This section develops an analytical approximation to the equilibrium rates of return in the single- $\rho$  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  $N(\tilde{x}_t, \tilde{\Omega})$ . As the numbers (reporting skewness and excess kurtosis) in Table 7 generated using the accurate numerical approximation demonstrate, Normality is a fairly accurate description in the case of the single- $\rho$  model.

Single- $\rho$ Model				
	$x_t$		$g_{c,t}$	
	$E$	$\sigma$	$E$	$\sigma$
Rat. Exp.	-	-	0.018	0.028
Bayesian	-0.002	0.023	0.018	0.032
Twisted	-0.023	0.024	-0.003	0.032
	$sk$	$\kappa$	$sk$	$\kappa$
Rat. Exp.	-	-	0.000	0.000
Bayesian	0.000	-0.000	0.000	-0.000
Twisted	0.000	-0.000	0.000	0.000

Two- $\rho$ Model				
	$x_t$		$g_{c,t}$	
	$E$	$\sigma$	$E$	$\sigma$
Bayesian	-0.001	0.024	0.019	0.034
Twisted	-0.022	0.028	-0.002	0.037
	$sk$	$\kappa$	$sk$	$\kappa$
Bayesian	-0.003	0.013	-0.003	0.017
Twisted	-0.005	-0.053	-0.038	-0.029

Table 7: **Conditional moments of distributions.** In each case,  $\gamma = 2.5$  and  $\alpha$  was set such that the model generates an average risk-free rate of 1.5%.  $C_t$ ,  $\hat{x}_{\ell,t}$ ,  $\hat{x}_{h,t}$  and  $\eta_t$  are set equal to their mean in the data.  $sk$  and  $\kappa$  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 (Approximating assumption 1)**  $\tilde{\mu}_t = N(\tilde{x}_t, \Omega)$ .

This is equivalent to assuming that eq. (22) is exactly a normal density with the same variance as the Bayesian posterior  $\Omega$  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  $k \in \mathbb{R}$ ,

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

Also,  $\widetilde{\text{Var}}_t(x_t) \equiv \text{Var}_{\tilde{\mu}_t}(x_t) = \Omega$  and  $\text{Var}_t(x_t) = \text{Var}_{\mu_t}(x_t) = \Omega$  and all  $\varepsilon$  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:

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

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

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

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 (Approximating assumption 2) :**

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

where  $z_t = \ln\left(\frac{P_t}{D_t}\right)$  and  $\kappa_0$  and  $\kappa_1$  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, which holds well in our data, see Figure 8.

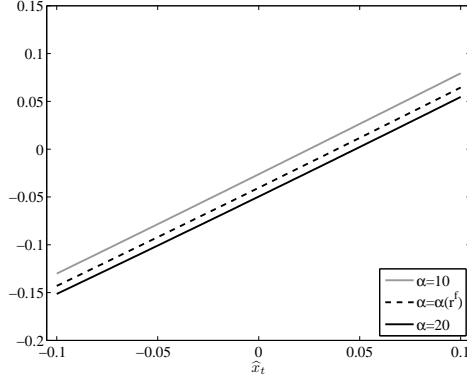


Figure 8:  $\tilde{x}_t = E_{\tilde{\mu}_t}(x_t)$  plotted against  $\hat{x}_t$ . The level of consumption is set to the average value between 1978 and 2011. In each case,  $\gamma = 2.50$ .

**Assumption 3 (Approximating assumption 3)**  $\tilde{x}_t = \delta_0 + \delta_1 \hat{x}_t$  for  $t = 1, 2, \dots$ ,  $\delta_1 \neq 0$ .

Note this assumption implies trivially that  $\hat{x}_t = (\tilde{x}_t - \delta_0)/\delta_1$ . Hence, we obtain a second order approximation of the second Euler equation as follows:

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

Plugging the guess for  $z_t$  and using the processes of growth rates, and using Assumptions 1 and 3, we obtain

$$1 = \tilde{E}_t \exp \left[ \ln(\beta) + \bar{d} - \gamma \bar{g} + \kappa_0 + (\kappa_1 - 1)A_0 + \kappa_1 A_1 (\delta_0 + \delta_1 \hat{x}_{t+1}) - A_1 \tilde{x}_t + (\psi - \gamma)\rho x_t + (\psi - \gamma)\sigma_x \varepsilon_{x,t+1} + \sigma_d \varepsilon_{d,t+1} - \gamma \sigma_g \varepsilon_{g,t+1} \right]. \quad (38)$$

In the expression for  $\hat{x}_{t+1}$  from the Kalman filter, let  $K = [K_g, K_d]$ . Then, we have now an expression for  $\hat{x}_{t+1}$  which is equal to (substituting  $d_{t+1}$  and  $g_{t+1}$  using their dynamics in the model):

$$\hat{x}_{t+1} = \rho \hat{x}_t (1 - K_g - \psi K_d) + (K_g + \psi K_d)\rho x_t + (K_g + \psi K_d)\sigma_x \varepsilon_{x,t+1} + K_g \sigma_g \varepsilon_{g,t+1} + K_d \sigma_d \varepsilon_{d,t+1}$$

Taking the log of eq. 38 and using  $\hat{x}_t = \frac{\tilde{x}_t - \delta_0}{\delta_1}$ . Hence,

$$\begin{aligned} 0 = & \ln(\beta) + \bar{d} - \gamma \bar{g} + \kappa_0 + (\kappa_1 - 1)A_0 + \kappa_1 A_1 \delta_0 - \delta_0 (\kappa_1 A_1 \rho (1 - K_g - \psi K_d)) \\ & + [\kappa_1 A_1 \rho (1 - K_g - \psi K_d) + \rho \kappa_1 A_1 \delta_1 (K_g + \psi K_d) + (\psi - \gamma)\rho - A_1] \tilde{x}_t \\ & + \rho^2 (\kappa_1 A_1 \delta_1 (K_g + \psi K_d) + \psi - \gamma)^2 \widehat{Var}_t(x_t)/2 \\ & + (\psi - \gamma + \kappa_1 A_1 \delta_1 (K_g + \psi K_d))^2 \sigma_x^2 / 2 \\ & + (\kappa_1 A_1 \delta_1 K_d + 1)^2 \sigma_d^2 / 2 + (\kappa_1 A_1 \delta_1 K_g - \gamma)^2 \sigma_g^2 / 2 \end{aligned}$$

Since this approximation must be valid for any  $\tilde{x}_t$ , we collect the  $\tilde{x}_t$  terms, set the expression equal to zero and we have

$$\kappa_1 A_1 \rho (1 - K_g - \psi K_d) + \rho \kappa_1 A_1 \delta_1 (K_g + \psi K_d) + (\psi - \gamma) \rho - A_1 = 0$$

which must hold for all  $\tilde{x}_t$ . Hence,

$$A_1 = \frac{\rho(\psi - \gamma)}{1 - \rho \kappa_1 (1 - (1 - \delta_1)(K_g + \psi K_d))} \quad (39)$$

Doing the same for the constant terms, we have

$$\begin{aligned} (1 - \kappa_1) A_0 = & \ln(\beta) + \bar{d} - \gamma \bar{g} + \kappa_0 + \kappa_1 A_1 \delta_0 - \delta_0 (\kappa_1 A_1 \rho (1 - K_g - \psi K_d)) \\ & + \rho^2 (\kappa_1 A_1 \delta_1 (K_g + \psi K_d) + \psi - \gamma)^2 \overline{Var}_t(x_t) / 2 \\ & + (\psi - \gamma + \kappa_1 A_1 \delta_1 (K_g + \psi K_d))^2 \sigma_x^2 / 2 \\ & + (\kappa_1 A_1 \delta_1 K_d + 1)^2 \sigma_d^2 / 2 + (\kappa_1 A_1 \delta_1 K_g - \gamma)^2 \sigma_g^2 / 2 \end{aligned} \quad (40)$$

Using eq. 37 and that  $E_t \tilde{x}_{t+1} = \delta_0 + \delta_1 E_t \hat{x}_{t+1}$  where  $E_t \hat{x}_{t+1} = \rho \hat{x}_t (1 - K_g - \psi K_d) + (K_g + \psi K_d) \rho E_t x_t = \rho \hat{x}_t$ , we obtain

$$E_t r_t = \kappa_0 + A_0 (\kappa_1 - 1) + \kappa_1 A_1 \delta_0 (1 - \rho) + \bar{d} + A_1 (\kappa_1 \rho - 1) \tilde{x}_t + \psi \rho \hat{x}_t \quad (41)$$

and so the Equity premium is then

$$\begin{aligned} E_t r_t - r_t^f = & \kappa_0 + A_0 (\kappa_1 - 1) + \kappa_1 A_1 \delta_0 (1 - \rho) + \bar{d} + A_1 (\kappa_1 \rho - 1) \tilde{x}_t + \psi \rho \hat{x}_t \\ & + \ln(\beta) - \gamma \bar{g} - \gamma \rho \tilde{x}_t + \frac{\gamma^2}{2} \left( \sigma_x^2 + \sigma_g^2 + \rho^2 \overline{Var}_t(x_t) \right) \end{aligned} \quad (42)$$

Note that when  $\delta_1 = 1$ , as is true in our data (see Figure 8),  $A_1$  simplifies to  $-\rho(\psi - \gamma) / (\kappa_1 \rho - 1)$ .

We need values of the approximating constants,  $\kappa_0$  and  $\kappa_1$ , to compute the log price-dividend ratio. Beeler and Campbell (2009) obtain the constants as follows

$$\begin{aligned} \bar{z} &= \frac{\sum z_t}{N} \\ \kappa_1 &= \frac{\exp \bar{z}}{1 + \exp \bar{z}} \\ \kappa_0 &= \ln(1 + \exp \bar{z}) - \kappa_1 \bar{z}. \end{aligned}$$

## B Details of the model where agents beliefs are given by the two- $\rho$ 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 BY type specification with either a low persistence ( $\rho_l$ ) or a high persistence ( $\rho_h$ ), but does not know for sure which. That is, she believes either of the models described in equation (28) 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:<sup>21</sup>

$$\begin{aligned}
 V(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) = & (1 - \beta) \frac{C_t^{1-\gamma}}{1-\gamma} \\
 & - \frac{\beta}{\alpha} \ln \left[ \eta_t \left\{ \int_{-\infty}^{\infty} \exp \left( -\alpha \iiint_{-\infty}^{\infty} V \left( C_t \exp(g_{l,t+1}), \hat{x}_{l,t+1}^{(l)}(\vec{\epsilon}_{l,t+1}), \right. \right. \right. \right. \\
 & \left. \left. \left. \hat{x}_{h,t+1}^{(l)}(\vec{\epsilon}_{l,t+1}), \eta_{t+1}^{(l)}(\vec{\epsilon}_{l,t+1}) \right) dF(\vec{\epsilon}_{l,t+1}) \right) dF(x_{l,t}) \right\} \\
 & + (1 - \eta_t) \left\{ \int_{-\infty}^{\infty} \exp \left( -\alpha \iiint_{-\infty}^{\infty} V \left( C_t \exp(g_{h,t+1}), \hat{x}_{l,t+1}^{(h)}(\vec{\epsilon}_{h,t+1}), \right. \right. \right. \\
 & \left. \left. \left. \hat{x}_{h,t+1}^{(h)}(\vec{\epsilon}_{h,t+1}), \eta_{t+1}^{(h)}(\vec{\epsilon}_{h,t+1}) \right) dF(\vec{\epsilon}_{h,t+1}) \right) dF(x_{h,t}) \right\} \Big]
 \end{aligned} \tag{43}$$

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  $\eta_t$  is the posterior probability at time  $t$  that the model with  $\rho_l$  is the data generating process.  $F(\vec{\epsilon}_{l,t+1})$  and  $F(\vec{\epsilon}_{h,t+1})$  are both trivariate independent standard normal

<sup>21</sup>Note 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.



distributions. The updates for  $\widehat{x}_{j,t+1}^{(i)}$  are obtained as follows:

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

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  $\rho_j$ ,  $j = h$ , the surprise is defined

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

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

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

Finally,  $\Omega_i$ ,  $i = l, h$ , is defined as the solution to

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

The Bayes update of  $\eta_t$  is obtained as follows :

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

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- $\rho$  model the risky rate of return is a function of four state variables,  $C_t, \widehat{x}_{l,t}, \widehat{x}_{h,t}, \eta_t$ , just like  $V$  and  $\xi_t$ . 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

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

and the risk-free rate is

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

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

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

## 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 8. All parameters, except  $\rho$  and  $\psi$  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,  $\psi$ , was set to either 3 or 2.5, which is slightly lower than values which maximize the likelihood.

**Parameter Estimates**

Parameter	$\psi = 3$				$\psi = 2.5$	
	$\rho = .25$	$\rho = .3$	$\rho = .85$	$\rho = .9$	$\rho = .3$	$\rho = .85$
$\bar{g}$	1.92 (0.302)	1.92 (0.302)	1.92 (0.302)	1.92 (0.302)	1.92 (0.302)	1.92 (0.302)
$\bar{d}$	2.31 (2.21)	2.31 (2.21)	2.31 (2.21)	2.31 (2.21)	2.02 (2.21)	2.02 (2.21)
$\sigma_g$	0.048 (0.016)	0.046 (0.016)	0.25 (0.010)	0.20 (0.007)	0.47 (0.017)	0.20 (0.008)
$\sigma_d$	4.49 (0.893)	4.51 (0.892)	4.75 (0.909)	4.73 (0.902)	4.64 (0.914)	4.77 (0.918)
$\sigma_x$	0.054 (0.013)	0.54 (0.013)	0.051 (0.019)	0.059 (0.021)	0.054 (0.013)	0.059 (0.021)

Table 8: **Parameter estimates** (standard errors below in parentheses) using annual data and the long-run risk model, shown above, using data from 1930 until 1977. All standard deviation estimates and their standard errors have been multiplied by 100.

## D Details of the numerical solution procedure

### D.1 Solution Method: two- $\rho$ 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

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

where  $X_t \equiv (C_t, \widehat{x}_{h,t}, \widehat{x}_{\ell,t}, \eta_t)$  denotes the vector of state variables of our two- $\rho$  case<sup>22</sup> 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.<sup>23</sup>  $H_\ell(\cdot)$  is a Hermite polynomial of order  $\ell$  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,  $X_t \equiv (C_t, \widehat{x}_{h,t}, \widehat{x}_{\ell,t}, \eta_t)$ ,<sup>24</sup> The parameters  $\theta^y$ ,  $y \in \{V, R\}$ , 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

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

where

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

and

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

where

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

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

<sup>22</sup>In the single  $\rho$  case, the vector of state variables reduces to  $X_t = (C_t, x_t)$  and the approximant takes the simpler form  $\Phi_y(X_t) = \exp\left(\sum_{i_c, i_x \in \mathcal{I}} \theta_{i_c, i_x}^y H_{i_c}(\varphi_c(C_t)) H_{i_x}(\varphi_x(\widehat{x}_t))\right)$ .

<sup>23</sup>See Judd (1998), Chapter 7.

<sup>24</sup>We 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.

also define

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

with

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

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 expressions are simplified in the single- $\rho$  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  $\theta^V$  and  $\theta^R$  are then determined by projecting the residuals on Hermite polynomials. This then defines a system of orthogonality conditions which is solved for  $\theta^V$  and  $\theta^R$ . More precisely, we solve<sup>25</sup>

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

where

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

and

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

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  $\theta^V$  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

<sup>25</sup>It should be clear to the reader that the integral refers to a multidimensional integration problem, as we integrate over  $C$ ,  $x^h$ ,  $x^\ell$  and  $\eta$ .

and projections) and a Gauss Hermitian quadrature approach when dealing with uni-dimensional integrals (outer integrals in the computation of expectations).<sup>26</sup>

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- $\rho$  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 uni-dimensional 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.<sup>27</sup> 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  $\theta^V$  and  $\theta^R$
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

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

<sup>26</sup>See Judd (1998), chapter 7.

<sup>27</sup>More 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.

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

$$\langle \mathcal{R}_V(\theta^V; X_t) | \mathcal{H}(X_t) \rangle \text{ and } \langle \mathcal{R}_R(\theta^R, \theta^V; X_t) | \mathcal{H}(X_t) \rangle .$$

4. If the orthogonality conditions are satisfied, in the sense the residuals are lower than the stopping criterion  $\varepsilon$ , then the vector of parameters are given by  $\theta^V$  and  $\theta^R$ . Else update  $\theta^V$  and  $\theta^R$  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  $\{X_t\}_{t=t_1}^{t=t_2} = \{C_t, \hat{x}_{h,t}, \hat{x}_{l,t}, \eta_t\}_{t=t_1}^{t=t_2}$  of annual observations of aggregate per-capita consumption, beliefs and prior probabilities in the time periods  $t = t_1$  through  $t = t_2$  we compute the conditional  $n^{\text{th}}$  order moment of the risky rate in period  $t$  as

$$E_t^n R_{t+1} = \iiint_{-\infty}^{\infty} \Phi(X_{t+1})^n dF(\vec{\varepsilon}_{t+1}) dF(x_t) \quad (44)$$

The model average  $n$ -th order moment is then computed as

$$ER^n = \frac{1}{t_2 - t_1} \left[ \sum_{t=t_1}^{t=t_2} E_t^n R_{t+1} - (E_t^1 R_{t+1})^n \right] \quad (45)$$

Similarly, given a sequence  $\{C_t, \hat{x}_{h,t}, \hat{x}_{l,t}, \eta_t\}_{t=t_1}^{t=t_2}$ , 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 uni-dimensional 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. (45).



### 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 9 and show that the approximation is accurate.

$\gamma$	Single- $\rho$ Model				Two- $\rho$ Model			
	$\alpha$	$E_1$	$E_2$	$E_\infty$	$\alpha$	$E_1$	$E_2$	$E_\infty$
2.0	11.51	-4.98	-8.18	-4.52	17.75	-3.63	-5.63	-3.34
2.5	7.24	-5.54	-9.29	-5.09	11.35	-4.07	-6.50	-3.77
3.0	4.21	-8.66	-15.59	-8.05	6.65	-5.78	-9.93	-5.48

Table 9: **Accuracy of the Numerical Solution:** This table reports the measure of accuracy for the Euler equation. In each case,  $\alpha$  was set such that the model generates a risk-free rate of 1.5%.

For example, let us consider the single  $\rho$  case with  $\gamma = 2$ , an agent who uses the approximate solution based on consumption claims would make, on average, a 1 dollar mis-

take for every 95,500 dollars invested in the assets, while the maximal error would be of the same order. Good performances are valid for the two values of persistence ( $\rho$ ) 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  $\rho$  case, the model is almost log-linear, such that our approximation performs remarkably well. In the two  $\rho$  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.<sup>28</sup>

---

<sup>28</sup>Accuracy is actually improved by increasing persistence, lowering the leverage and the discount factor.