

Ambiguity and the historical equity premium¹

Fabrice Collard

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

Sujoy Mukerji²

School of Economics and Finance
Queen Mary University of London, London E1 4NS

Kevin Sheppard

Department of Economics
University of Oxford, Oxford OX1 3UQ

Jean-Marc Tallon

Paris School of Economics and CNRS
48, boulevard Jourdan, 75014 Paris

January 1, 2018

Forthcoming in *Quantitative Economics*

¹We thank three referees and the editor for thoughtful comments and suggestions. We thank R. Bansal, P. Beaudry, H. Bhamra, J. Borovicka, T. Cogley, H. Chen, H. d'Albis, V. Gala, C. Gollier, L. Hansen, P. Klibanoff, H. Liu, T. Ramadorai and R. Uppal for helpful discussions. We also thank seminar and conference participants at Adam Smith Asset Pricing conference, AEA, RUD, Northwestern (MEDS), Warwick, Leicester, Transatlantic Theory Workshop, EUI (Florence), UBC, Workshop on Ambiguity and Robustness in Macroeconomics and Finance (Becker-Friedman Inst.). Tallon thanks support from the Investissement d'Avenir Program (ANR-10-LABX-93).

²Contact author: s.mukerji@qmul.ac.uk

Abstract

This paper assesses the quantitative impact of ambiguity on historically observed financial asset returns and growth rates. The single agent, in a dynamic exchange economy, treats the conditional uncertainty about the consumption and dividends next period as ambiguous. We calibrate the agent's ambiguity aversion to match *only* the first moment of the risk-free rate in data and measure the uncertainty each period conditional on the actual, observed history of (U.S.) macroeconomic growth outcomes. Ambiguity aversion accentuates the conditional uncertainty endogenously in a dynamic way, depending on the history; e.g., it increases during recessions. We show the model implied time series of asset returns substantially match the first and second conditional moments of observed return dynamics. In particular, we find the time-series properties of our model generated equity premium, which may be regarded as an index measure of revealed uncertainty, relates closely to those of the macroeconomic uncertainty indices developed recently in Jurado, Ludvigson, and Ng (2015) and Carriero, Clark, and Marcellino (2017).

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

Keywords: Ambiguity aversion, Asset pricing, Equity premium puzzle, Time-varying uncertainty, Uncertainty shocks.

1 Introduction

This paper seeks to assess the quantitative impact of ambiguity on financial asset returns and prices, in particular their dynamic paths, conditioned on observed historical growth rates. 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 his choices. Our goals are 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, we only consider conditional uncertainty at information sets adapted to the path of observed historical macroeconomic growth rates, as opposed to counterfactual, simulated sample paths. Two, our model of agent’s preferences departs from standard expected utility solely by allowing for sensitivity to ambiguity; take that away, and the agent’s preferences reduce to standard expected utility. These two components, together with the demonstration that they *alone* are sufficient to substantially explain a range of asset return dynamics, distinguish the contribution in this paper.

Ambiguity-averse agents are inclined to choose actions whose consequences are more robust to the perceived ambiguity, e.g., a portfolio position whose (ex-ante) value is relatively less affected by the uncertainty about probability distribution governing the future payoffs.¹ An important reason why ambiguity may be pervasive in economic and financial decision making is model uncertainty. For example, a typical professional investor may have different forecasting models for the same variable or different parameter estimates for the same model, all of which are plausible on the basis of historical data. If the models make distinct (probabilistic) forecasts about key variables of interest, it is natural to seek a portfolio that accounts for differences in the agent’s outcome across the range of forecasts rather than optimizing exclusively to the forecast from a single model as argued, e.g., in Hansen (2007).

This paper considers a standard single agent, Lucas-tree, pure-exchange economy with two less standard assumptions. First, the agent’s belief about the consumption and dividend process is ambiguous, i.e., in each period, he is uncertain about the exact probability distribution governing the realization of consumption and dividends in the following period. Furthermore, this belief is dynamic, evolving as the agent

¹See Dow and Werlang (1992), Epstein and Wang (1994), Mukerji and Tallon (2001), Caballero and Krishnamurthy (2008), Chen, Ju, and Miao (2014), Gollier (2011), Boyle, Garlappi, Uppal, and Wang (2011), Hansen and Sargent (2010), Maccheroni, Marinacci, and Ruffino (2013) and Uhlig (2010), *inter alia*. Surveys of the related literature may be found in Gilboa and Marinacci (2016) (decision theory) and Mukerji and Tallon (2004) (applications).

learns from history. Second, the agent’s preferences are ambiguity-sensitive, modeled using the smooth ambiguity model of Klibanoff, Marinacci, and Mukerji (2005, 2009) (henceforth KMM2005, KMM2009).

The assumed source of the ambiguity in the agent’s beliefs is the occurrence of periodic, temporary changes in the probability distribution governing next period’s growth outcome due to the effect of the business cycle. These transient deviations are assumed to be governed by an auto-regressive ($AR(1)$) latent variable. The agent is, however, unsure about the value of the persistence parameter of the $AR(1)$ process since, even with a large sample of growth rates, it is difficult to distinguish the case where the latent growth state is highly volatile but moderately persistent, from the case where the state is less volatile but highly persistent. Uncertainty about persistence, in turn makes it harder to estimate the evolving location of the latent variable precisely. Furthermore, depending on the observed history, the imprecision of the estimate of the location will vary over time, making the uncertainty about the probability distribution governing next period’s growth vary over time.

The ambiguity-averse agent’s robustness concerns generate, endogenously, *doubt* and *pessimism*, to use the language of Abel (2002). The portfolio choice of the ambiguity-averse agent in the model may be understood as that of an expected utility agent with an “as if” (probabilistic) belief that is more uncertain and pessimistic than the one obtained by objective inference, in the standard fashion, from data. Moreover, the endogenous accentuation of doubt depends on the observed history and the level of ambiguity aversion, making the severity of the effect of uncertainty *endogenously time-varying*. For instance, after a negative shock that follows a series of “normal” ones, the agent behaves as if the uncertainty is more severe and more persistent than what is implied by pure Bayesian inference (and the opposite, if it were a positive shock that broke the normal sequence). The level of ambiguity aversion is calibrated to match the average risk free rate (no other moment is used); all other parameters are either inferred/estimated from the history or fixed at values widely used in the literature.

We present two kinds of results on model implied conditional moments of rates of return and price-dividend ratio: (time-) averages of the moments over the sample period (1978-2011) and time series of the moments over the same sample period, all based on conditional uncertainty at information sets reconcilable with historical growth data. 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.

The model generated conditional equity premium is a measure of conditional macroeconomic uncertainty as revealed by the behavior of the agent in the model. We show its time-series properties match those of the purely statistical index of macroeconomic

uncertainty, recently developed in Jurado, Ludvigson, and Ng (2015) and Carriero, Clark, and Marcellino (2017). Our model gives a theory of why an agent makes decisions following a positive shock that (endogenously) underplays the uncertainty and its persistence, while following a negative shock, behaves *as if* a more severe *and* a more persistent shock were in play, thus explaining a key feature of the index and related findings of the recent literature on *uncertainty shocks*. In particular, the counter-cyclical persistence of equity premium and (revealed) uncertainty speaks directly to the mechanism of ambiguity aversion in our model.

Altogether, our contribution is to demonstrate that model/parameter uncertainty and learning coupled with ambiguity aversion, by themselves, create a quantitatively plausible and intuitively meaningful mechanism for explaining the relationship between macroeconomic uncertainty and the dynamics of equity prices and returns. The time-averaged conditional moments predicted by the model match data moments as well as the best matches in the literature (e.g., in Collin-Dufresne, Johannes, and Lochstoer (2016) and papers cited therein). Our more distinctive results are those on the predicted time series of conditional moments statistics. Two key stylized facts our model matches are the counter-cyclical of conditional equity premium and the pro-cyclical of conditional (excess) return volatility. Models in the literature have found it hard to explain these facts without introducing at least one of the following elements: (a) some *exogenously* time varying uncertainty, such as, time dependent, stochastic volatility; (b) aversion to later resolution of risk via an intertemporal elasticity of substitution (IES) that is significantly greater than unity; (c) habit formation; all elements that are not part of our mechanism.² A reason to be interested in the mechanism posited in the present paper alongside these “best performing” alternatives in the recent literature is that the alternatives rest on assumptions that have been empirically questioned and hence cannot be regarded as the “last word” on the subject. At the same time, the first findings on the estimation and calibrations of ambiguity aversion in the context of asset pricing are promising.

The route of relying on exogenously posited stochastic volatility of aggregate consumption has been questioned because “the evidence for heteroskedasticity in aggregate consumption is fairly weak,”(Campbell (2000)). In a similar vein, Lettau and Ludvigson (2010) and Ludvigson (2012) in their surveys argue that the evidence for stochastic volatility suggests it has neither the inter-temporal shape nor the size required for mod-

²Bansal and Yaron (2004) incorporate (a) and (b); Campbell and Cochrane (1999) have (c); Drechsler (2013) incorporates model uncertainty, learning, ambiguity aversion (a) and (b); Collin-Dufresne, Johannes, and Lochstoer (2016), model uncertainty, learning and (b); Ju and Miao (2012) and Hansen and Sargent (2010) incorporate model uncertainty, learning, ambiguity aversion and (b). We discuss more details of this related literature in Section 5.

els based on stochastic volatility to fit facts about inter-temporal variation in return moments. A more fundamental difference between stochastic volatility based asset pricing models and ours is that in the former there is no explanation, as such, of the variation in volatility: in those models agents are more uncertain when they believe they are in a state where future economic shocks are assumed (exogenously) to be more volatile. In contrast, our model gives a theory *why* an agent makes decisions following a positive shock that underplays the inferred uncertainty and, after a negative shock that follows a series of “normal” ones, behaves as if the uncertainty is more severe and persistent, than pure Bayesian inference would suggest.

It is well documented that the empirical evidence on whether IES is greater than 1 is very mixed (see discussions, e.g., in Beeler and Campbell (2012) and Bansal, Kiku, and Yaron (2012).) Furthermore, recently, Epstein, Farhi, and Strzalecki (2014) argue using a calibration exercise, that the $IES > 1$ values applied in the recent asset pricing literature imply a very implausible premium for early resolution of uncertainty. While we do not know of conclusive direct evidence for or against habit formation, there is some evidence against the key underlying mechanism. Neither in data (nor in the model in the present paper) does lagged consumption growth predict the future price-dividend ratio, while in the habit-formation model it predicts the future price-dividend with an R^2 of over 40%.

A recent study, Gallant, Jahan-Parvar, and Liu (2015), which uses macroeconomic and financial data to estimate the size of ambiguity aversion (as a parameter in a consumption-based asset pricing model based on an elaborated version of the smooth ambiguity model), finds that the estimate “suggests ample scope for ambiguity aversion” to explain asset pricing facts. In the present paper, in Section 4, we conduct a calibration exercise to argue that the size of the ambiguity aversion parameter we apply has very plausible implications for uncertainty premia.

We view the preceding discussion about alternative models and ours as not an argument for considering the approach taken here to be the best, but as showing that it merits careful study and development. 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, assuming a general form of beliefs. In a subsection, we describe and motivate the specific model of ambiguous beliefs we adopt. Section 3 first outlines the numerical solution method we employ, then identifies the key qualitative mechanisms at work in our model and finally presents and explains the quantitative implications of our model for asset prices and returns in the light of the mechanisms identified. In Section 4, using a thought experiment, we show that a decision maker with preferences and beliefs calibrated to match those of our agent’s

will demand a total uncertainty premium (for the Lucas tree) that is well within the bounds of the amounts widely considered as plausible. Section 5 discusses the more closely related literature. A final section concludes. The Appendix gathers several items, including, details of parameter values used in the model, details of the model including the specification of beliefs, how they are updated, and the formulae for rates of return.

2 The Model

2.1 Agent's preferences: recursive smooth ambiguity

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, 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, $s^1 \equiv (s^0, s_1)$ where $s_1 \in \mathcal{S}_1$, the set of possible observations at time $t = 1$; and, so on. The decision maker (DM) chooses between consumption plans f , each of which associates a payoff to a node s^t in the event tree. The DM is uncertain about which stochastic process governs the probabilities on the event tree. The domain of this uncertainty is given by a parameter space $\Theta \ni \theta$, the set of unobservable parameters, over which the DM makes inference at each s^t . We denote by $\pi_\theta(s_{t+1} | s^t)$ the probability under *likelihood* distribution π_θ that the next observation will be s_{t+1} , given that node s^t is reached. The decision maker's prior on Θ is denoted by μ . KMM2009 give assumptions such that *recursive smooth ambiguity preferences* over plans f at a node s^t are updated and represented as:

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

where $V_{s^t}(f)$ is a recursively defined (*direct*) value function, u characterizes attitude to risk, β is a discount factor, ϕ is a function characterizing the decision maker's ambiguity attitude, while $\mu(\cdot | s^t)$ denotes the Bayesian posterior. A concave ϕ characterizes ambiguity aversion, which is defined to be an aversion to mean preserving spreads in the distribution over expected utility values. In general, the model does not impose reduction between the second-order belief μ and the first-order probabilities π_θ 's; reduction only applies when ϕ is affine, representing an ambiguity neutral Bayesian expected utility maximizer.

Ambiguity aversion in this model is equivalent to the DM behaving as more risk averse when choosing between bets on θ than when choosing between objective lotteries.

That is, the DM strictly prefers a lottery which yields a unit payoff with objective probability m (and 0 with probability $1 - m$) to a (same stakes) bet on an event $T \subset \Theta$, where $\mu(T) = m$ and also strictly prefers the complementary lottery to the bet on the complementary event.³ The behavior is exactly analogous to the modal behavior in the Ellsberg two-urn example: preference for betting on a draw from the urn with a known 50:50 mix over betting on a draw from the urn with unknown mix. Hence, the second-order measure μ cannot be calibrated with a lottery; behaviorally, μ is not treated as an objective probability. The standard interpretation is that the DM views his belief about events such as T to be less reliable than an objective probability.

2.2 A Lucas-tree economy and Euler equations with general beliefs

There is an infinitely-lived agent, with recursive smooth ambiguity preferences, consuming a single good. He can trade in a short lived 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 (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 .⁴ Equilibrium will require that at each time $C_t = l_t + D_t$.

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

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

subject to the budget constraint and the law of motion of the two “state” variables (wealth and beliefs), where $J(W_t, \mu_t)$ denotes a recursively defined *indirect* value function (as opposed to the direct value function in eq. (1)). An equilibrium of this economy

³See Section D in the Appendix for details.

⁴In other words, we assume directly a stochastic process for C_t and D_t , leaving labor income l_t implicit.

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 the market clears, i.e., $e_t = 1$, $b_t = 0$, $C_t = D_t + l_t$ at each 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 function ξ_t is a Radon–Nikodym derivative effecting a *node specific* change of measure, or “distortion”, on the posterior μ_t , akin to martingale distortions arising in robust control problems considered by Hansen and Sargent. The distortion is a function of the continuation expected values obtained at successor nodes. In this paper we assume $\phi(x) = -\exp(-\alpha x)/\alpha$, where the parameter α represents ambiguity attitude. This specification simplifies the expressions significantly, since we now have $\Upsilon_t = 1$. It is also assumed that $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$. With these specifications, the Euler equations are as follows:

$$\beta R_t^f E_{\mu_t} [\xi_t(\theta) E_{\pi_\theta} [\exp(-\gamma g_{t+1})]] = 1 \quad (6)$$

$$\beta E_{\mu_t} [\xi_t(\theta) E_{\pi_\theta} [R_{t+1} \exp(-\gamma g_{t+1})]] = 1 \quad (7)$$

$$\Leftrightarrow \beta E_{\mu_t} \left[\xi_t(\theta) E_{\pi_\theta} \left[\left(\frac{\exp(z_{t+1}) + 1}{\exp(z_t)} \right) \exp(d_{t+1} - \gamma g_{t+1}) \right] \right] = 1 \quad (8)$$

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 *Observe, these Euler equations look identical to ones obtained in a standard Bayesian model except for the inclusion of the distortion function, ξ_t . The distortion, in the case of ambiguity aversion, increases the (posterior) weight on likelihoods π_θ with lower expected continuation values, $E_{\pi_\theta}(J(W_{t+1}, \mu_{t+1}))$. One could splice together the one-period ahead predictive distributions, $[\xi_t(\theta) \times \mu_t(\theta)] \otimes \pi_\theta(g_{t+1}, d_{t+1})$, and construct an overall “as if” unconditional probability distribution over the event tree which could be reinterpreted as coming from a Bayesian model. However, seen by itself, the constructed as if distribution cannot be linked to the given set of likelihoods $\{\pi_\theta\}_{\theta \in \Theta}$;*

indeed, typically, it is not possible to obtain the constructed distribution by starting at the initial node with a different prior $\mu'_1 \neq \mu_1$ on Θ with μ'_t , $t > 1$, obtained by updating in the usual way. Hence, an understanding of the role of ambiguity aversion in the modeling exercise is that it provides a link between the subjective as if distribution and a specification of beliefs about possible data generating the processes $(\{\mu_t\}_t, \{\pi_\theta\}_{\theta \in \Theta})$; beliefs which, in principle, can be objectively reconciled with data.

Remark 2 If $\phi(\cdot)$ were different from an exponential, e.g., a power function, then $\Upsilon_t \neq 1$, in general, and hence in such a case the difference between these Euler equations and the standard set would not simply be the change of measure term $\xi_t(\theta)$. Thus it is down to our choice of the specification of $\phi(\cdot)$ and of $u(\cdot)$ that we may interpret our Euler equations as arising from an agent we see in standard macro-finance models (with the preference over consumption given by a power function who has non-standard (though Savage-Bayes rational) beliefs which may be justified by appealing to robustness/model uncertainty concerns. This way we can embed our model within that standard literature and, very much in terms of that literature, motivate and explain its point of departure and intuition. Given the specification, the point of departure is just the non-standard beliefs that can be motivated entirely in terms of robustness concerns, arguably very reasonable, even normatively compelling, given the model and parameter uncertainty faced by a typical agent in the real world. Furthermore, the fact that the (non-standard part of) beliefs is entirely shaped by the history dependent $\xi_t(\theta)$ is the key that will allow us to make transparent (as will be seen in Section 3.2) the two key mechanisms driving the results, the higher time averages and the endogenously dynamic fluctuation of returns.

A drawback of this specification is that our value function misses a homogeneity property; note the dependence on (W_t) in (2) and, equivalently, on (C_t) in (10). Thus, the curse of dimensionality makes numerical analysis of the decision maker's dynamic programming problem more complicated. Numerically, we already have a relatively high dimension problem. If we were to use a power function specification for ambiguity preference, wealth in (2) and consumption level in (10) will be factored out. This will not only reduce the dimension by 1, consumption level will drop out of all pricing equations.

Our modeling choice reflects our belief that the two advantages of the adopted specification in providing economic motivation and intuition, outweigh the disadvantage of the ensuing numerical complication. Furthermore, dispensing with homogeneity, a departure from standard practice, required us to be innovative with our numerical method; these innovations might prove useful in future research.

2.3 Beliefs and how they are applied in the evaluation of the Lucas tree

2.3.1 Description

We now describe the specific belief about the Lucas tree economy that we apply in our analysis. It is assumed the agent believes the growth rate of consumption (g_t) and dividends (d_t) are partly driven by a common latent state, x_t , which evolves according to an $AR(1)$ process with persistence ρ . While it is assumed there is a single persistence parameter operating through history, the agent is unsure what it is, believing there are two possible values of the parameter, high (ρ_h) or low (ρ_l). At time t the agent puts probability η_t on persistence being low and $(1 - \eta_t)$ on persistence being high. Each possible process is:⁵

$$\begin{aligned} x_{k,t+1} &= \rho_k x_{k,t} + \sigma_{x_k} \varepsilon_{x_k,t+1} \\ d_{k,t+1} &= \bar{d} + \psi x_{k,t+1} + \sigma_{d_k} \varepsilon_{d_k,t+1} = \bar{d} + \psi (\rho_k x_{k,t} + \sigma_{x_k} \varepsilon_{x_k,t+1}) + \sigma_{d_k} \varepsilon_{d_k,t+1} \\ g_{k,t+1} &= \bar{g} + x_{k,t+1} + \sigma_{g_k} \varepsilon_{g_k,t+1} = \bar{g} + \rho_k x_{k,t} + \sigma_{x_k} \varepsilon_{x_k,t+1} + \sigma_{g_k} \varepsilon_{g_k,t+1} \end{aligned} \quad (9)$$

where $(\varepsilon_{g_k,t+1}, \varepsilon_{d_k,t+1}, \varepsilon_{x_k,t+1})' \sim N(0, I)$, for $k = l, h$. We denote using, \bar{g} , \bar{d} the long-run growth rate of consumption and dividend, respectively. The shock $x_{k,t}$ is the temporary deviation from the trend (identified by the long-run growth rate). The interpretation is that the mean of the distribution on growth is partly fixed by the long-run trend and partly by a temporary shock to productivity due to the business cycle. The business cycle effect on the productivity across the economy is not observed directly. Though an innovation in each period, today's business cycle shock is, naturally, related to previous period's shock, and, so, is modeled by a auto-regressive latent variable. The factor ψ accounts for the empirically observed greater volatility of dividend relative to that of consumption.⁶ Note, there is a *different* tuple of volatility parameters $(\sigma_{g_k}, \sigma_{d_k}, \sigma_{x_k})$ associated with each possible value of persistence, ρ_k .

The agent is assumed to know the values of parameters $(\bar{g}, \bar{d}, \sigma_{g_k}, \sigma_{d_k}, \sigma_{x_k}, \psi)$. The agent observes, contemporaneously, the consumption and dividend growths. Given $x_{k,t}$, ρ_k and the current node $\{(C_\tau, D_\tau)\}_{\tau=0}^t$ the probability distribution over the immediate successor nodes, identified by (g_{t+1}, d_{t+1}) , is the product of two conditionally independent, given $x_{k,t}$ and ρ_k , Normal distributions,

$$g_{k,t+1} \sim N(\bar{g} + \rho_k x_{k,t}, \sigma_{g_k}^2 + \sigma_{x_k}^2) \quad \text{and} \quad d_{k,t+1} \sim N(\bar{d} + \psi \rho_k x_{k,t}, \sigma_{d_k}^2 + \sigma_{x_k}^2).$$

This product distribution is the typical first-order distribution, the object $\pi_\theta(\cdot | s^t)$ in the abstract KMM formulation, with $(\rho_k, x_{k,t})$ playing the role of the unobserved

⁵When $\eta_0 = 0$, the model reduces to the CASE I in Bansal and Yaron (2004).

⁶This modeling device was introduced in Abel (1999) and is followed widely in the finance literature and may be interpreted as the "leverage ratio" on (expected) consumption growth.

parameter “ θ ”. (Note, since the volatilities $\sigma_{g_k}, \sigma_{d_k}, \sigma_{x_k}$ may vary with k , the parameter fixes *both* mean and variance.)

Thus, the domain (i.e., the support) of the second-order uncertainty at time t is an union of two component sets, $\{\rho_l x_{l,t} \mid x_{l,t} \in \mathbb{R}\} \cup \{\rho_h x_{h,t} \mid x_{h,t} \in \mathbb{R}\}$. The agent’s prior belief ascribes a measure to each component set: the measure on the first component is given by $\eta_0 \otimes N(0, \sigma_0^2)$ and that on the second by $(1 - \eta_0) \otimes N(0, \sigma_0^2)$. The agent updates beliefs using Bayes rule, based on the history of growth realizations and the presumption that the economy conforms to one of the two processes described in (9). Let $\hat{x}_{k,t} \equiv E[x_{k,t} \mid g_{k,1}, \dots, g_{k,t}, d_{k,1}, \dots, d_{k,t}]$ denote the expectation of $x_{k,t}$ conditional on the history of growth rates up to t if the beliefs were updated assuming $\rho = \rho_k$ is the data generating process. The filtered latent state corresponding to process k , $\hat{x}_{k,t}$, is obtained by applying the (steady state) Kalman filter that takes the process with $\rho = \rho_k$ as the “true” data generating process. The agent’s posterior belief then ascribes a measure on the first component set given by $\eta_t \otimes N(\hat{x}_{l,t}, \Omega_l)$ and that on the second by $(1 - \eta_t) \otimes N(\hat{x}_{h,t}, \Omega_h)$, where $\Omega_k, k = l, h$, denotes the steady state variance associated with the Kalman filter based on the process with $\rho = \rho_k$ and η_t shows the posterior belief on ρ_l . Hence, the agent’s posterior may be summarized by the tuple, $(\hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t)$.⁷

We now turn to the evaluation of the Lucas tree with the specified beliefs. Denote by $\hat{x}_{k,t+1}^{(i)}$, $i = l, h$, $k = l, h$, the agent’s forecast for the (one period ahead) update using a Kalman filter which assumes the model with $\rho = \rho_k$ as the data generating process, when the data is actually generated by the $\rho = \rho_i$ model. Correspondingly, $\eta_{t+1}^{(l)}$ (respectively $\eta_{t+1}^{(h)}$) is the posterior probability that the low persistence process is the correct model when the low (high) persistence model is the data generating process. The direct continuation value is a function of the current node but does not distinguish between two histories which have the same current consumption and same current belief, summarized by $(\hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t)$. The function is defined by the following recursion:

$$V(C_t; \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) = u(C_t) + \beta \phi^{-1}(\mathcal{V}_{t+1},) \quad (10)$$

where

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

⁷See Section B in the Appendix for further details about the updating.

To see how the KMM representation is being implemented, we note the following. The argument of a $\phi(\cdot)$ is an expectation of the continuation value/utility at successor nodes, where the expectation $E_{x_{k,t}}$ is taken with respect to the typical first-order distribution described earlier, defined by fixing the “parameter pair” $(\rho_k, x_{k,t})$. The measure on $(\rho_k, x_{k,t})$ is given by $\eta_{t,k} \otimes N(\widehat{x}_{k,t}, \Omega_k)$ and we calculate the expectation of the functions $\phi(\cdot)$ by applying this measure, which corresponds to the second-order measure μ_t in the KMM representation.

2.3.2 Motivation for the beliefs model and parameter choice

Hamilton (1989) pioneered the idea of modeling consumption growth as an auto-regressive process, with parametric shifts occurring through Markovian transitions on latent states. That paper also showed that the idea was a particularly good fit for the U.S. growth experience through the improved facility of capturing the effect of business cycles. Hence, the basic functional form of (9) with a given ρ_k , is a plausible starting point for describing the beliefs of an investor for whom the key source of uncertainty is the business cycle.

Adding uncertainty about ρ_k to (9) is empirically justified and improves it as a framework for understanding and quantifying ambiguity about macroeconomic growth; this enables (9) to encapsulate a theory of why it is difficult to precisely estimate the probability distribution of growth, and of why and how that imprecision will vary with history. The key is that the two uncertainties, about persistence ρ_k , and about $x_{k,t}$, which controls the mean of the distribution, go hand in hand: they interact and reinforce each other to make the belief about the “true” growth distribution unreliable and inference about it imprecise. Shephard and Harvey (1990) explains that it is very difficult (in that it would take an inordinately long series of observations) to determine whether the true growth process is a very persistent process where the persistent component has a small volatility or whether it is a moderately persistent process with a persistent component that has a large volatility. Thus, uncertainty about the volatility of the latent variable makes the persistence parameter difficult to estimate. Indeed, even after almost a century of data the learning, far from settling down on one value of ρ_k , produces posteriors η_t that have varied continually between 0.3 and 0.7. In turn, the uncertainty about ρ_k degrades the inference on the evolving latent variable $x_{k,t}$. The expectation of this variable is tracked by the Kalman filter, but the specification of the Kalman filter is determined by the value of the persistence parameter. Since that is not reliably known, the Kalman forecast is imprecise.

This understanding of the uncertainty described by (9) motivates how it is represented in the different parts of the KMM preference functional. Given $(\rho_k, x_{k,t})$, the

uncertainty about the parameters of the distribution on growth is almost objective since the other parameters fixing the distribution may be reliably estimated given this knowledge and the run of data. On the other hand, the uncertainty about $(\rho_k, x_{k,t})$, though probabilistically represented, may be viewed as a deeper uncertainty, far less reliably estimated and more variable. Thus, the former uncertainty appears as a first-order belief in the KMM functional (i.e., “inside” the ϕ) whereas the latter uncertainty is treated as a second-order uncertainty (i.e., “outside” the ϕ).

There are two reasons for choosing a two point support for the uncertainty about persistence. One is computational limitation (with more than two points the number of “state variables” in the dynamic problem that we have to solve goes beyond the state of art capabilities, as we not only need to introduce other persistence parameters, but also to keep track of the updated latent variable in each regime). The second is that a two-point support is an efficient way of capturing Shephard and Harvey (1990)’s key insight that the crucial empirical confound underlying the uncertainty is the confound between a high persistence combined with low volatility parameters on one hand and low persistence combined with larger volatility parameters, on the other.

We were guided in part by findings in the literature, and in part by our own empirical investigations, in choosing the values of ρ_k . One substantial strand of literature (the long run risk literature, pioneered in Bansal and Yaron (2004)) argues there is strong justification, based on asset pricing moments, for assuming a high value of ρ . Another strand points out that pure consumption growth data suggest a more moderate value. It is generally agreed the estimates are quite fragile. Using annual data, we set $\rho_h = 0.85$ as the standard case (and 0.90 for robustness checks), which correspond to the endpoints of the interval of (annualized) values suggested by this literature,⁸ and $\rho_l = 0.30$, motivated by studies in Beeler and Campbell (2012) and Constantinides and Ghosh (2010).⁹ Our own investigations found, setting $\rho_l = 0.3$ and $\rho_h = 0.85$, η_t is approximately 50% in 1977, the beginning of the model evaluation period, and is consistently in the interval $[0.3, 0.7]$ throughout the period 1978-2011, demonstrating how difficult it is to separate the two persistence models on the basis of growth data.¹⁰

⁸For example, Bansal and Yaron (2004) calibrate ρ for a monthly frequency which corresponds to the .85 annualized value. We used annual series rather than monthly/quarterly to enable us to include the Great Depression in the agent’s memory (through the time series estimates of the parameters and evolution of η). Findings of Pohl, Schmedders, and Wilms (2015) show that price-dividend ratio is approximately log-linear for persistence below .95. For higher level of persistence (such as for monthly data), important non-linear effects appear. Thus our results might be missing some of these effects that appear at higher frequency.

⁹Constantinides and Ghosh (2010) provide a GMM estimate (based on the years 1931-2006) of $\rho = 0.32$ (see their Table 4). Though we set $\rho_l = 0.30$, (we found) values between 0.25 and 0.40 have virtually identical posteriors (and implications for rates of returns).

¹⁰Bidder and Dew-Becker (2016) argues a case for embedding the LRR model into an ambiguity

In our model, the domain of ambiguity consists of the x 's and ρ 's. It is evident from the first line of (9) the x 's are Markovian states. So, the agent would never learn the contemporary x and this part of the ambiguity persists even in the steady state. However, the uncertainty about ρ will not last in steady state, the true value will be learnt eventually. In practice, because of the problem pointed out by Shephard and Harvey, learning does not occur fully even with long runs of data (as we see in our sample and even in Hansen and Sargent (2010), which applies a quarterly series from 1947). Having a Markovian process for the persistence would be better to the extent that the entire ambiguity then will persist in the steady state and our analysis would be a full steady state analysis. However, for the moment this proves to be technically intractable since we lose the linear updating formulation the present model allows.

The time-series parameters of the model (except for the persistence parameters ρ_k , and the leverage-ratio parameter ψ) were estimated using maximum likelihood on annual U.S. data from 1930 to 1977 (see Section A in the Appendix for details about the data set and the parameter values.) The remaining years in the data set, 1978-2011, were used in the evaluation of the model. Our aim was to have the longest run of data for the evaluation of the model. Parameter estimates change significantly through the 70s because of the macroeconomic events. By starting the evaluation at 1978, the maintained assumption that the agent behaves as if he knows the parameter values of the model becomes more credible.¹¹ Turning to preference parameters, in all cases the ambiguity aversion parameter α 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. Section 4 discusses whether the calibrated level is plausible for an individual agent. No other moments were used in the choice of α . Choice of the other preference parameters follows the standard practice in the literature.

setting as follows:

“A criticism of the long-run risk model has always been that it depends on a process for consumption growth that is difficult to test for. We turn that idea on its head and argue that it is the difficulty of testing for and rejecting long-run risk that actually makes it a sensible model for investors to focus on.”

¹¹This sample split is the benchmark split and our results everywhere in the paper are based upon this, unless stated otherwise. Table 3 reports on the robustness of our results to alternative learning assumptions that would be implied by splitting the sample differently. Different sample splits imply the learning is different because the parameter estimate the agent learns is potentially different due to being based on a different sample and furthermore, the different estimates imply different filtered values of x and updates of η because the formula of the Kalman filter is a function of these estimates (see equations (21) to (26).)

3 Implications of the model for asset returns and prices

3.1 Solution Method

We pursue the following methodology to numerically solve our model. We solve the model using a projection method, which allows us to express asset prices as a function of the state variables. We then evaluate the state variables at the observed exogenous variables (the history of GDI and dividend growths), which allows us to generate a time-series of predicted asset prices.¹² Using this, we are able to match the equity premium, the risk-free rate and the price-dividend ratio. We now give more detail on how the solution method works.

An approximate solution to the model is obtained by using the minimum weighted residuals method proposed by Judd (1992). We compute two approximation functions: one for the risky rate, R , and one for the value function, V , which is required to compute the belief distortion. Both 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 $X_t \equiv (C_t, \hat{x}_{h,t}, \hat{x}_{\ell,t}, \eta_t)$ denotes the vector of state variables and $y \in \{V, R\}$. We use a complete basis of orthogonal polynomials, such that the set of indices \mathcal{I} is defined¹³ by

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

Because the model is not homogeneous, we could not deflate the variables, and the support for the approximation is potentially non bounded from above. We therefore rely on Hermite polynomials which are defined over \mathbb{R}_+ . Accordingly, $H_\iota(\cdot)$ denotes a Hermite polynomial of order ι and $\varphi_z(\cdot)$ is a strictly increasing function that maps \mathbb{R} into \mathbb{R} . This function is used to map Hermitian nodes into values for the vector of state variables, $X_t \equiv (C_t, \hat{x}_{h,t}, \hat{x}_{\ell,t}, \eta_t)$.¹⁴ The parameters θ^y , $y \in \{V, R\}$, are then

¹²Hence, we evaluate the model on the actual, observed, history. If we chose to simulate history, we would need to present results based on simulations of two sets of histories, generated by assuming the true persistence to be H and L , respectively.

¹³In our application, we use $(n_c, n_{x_h}, n_{x_\ell}, n_\eta) = (5, 2, 2, 2)$ for the value function and $(n_c, n_{x_h}, n_{x_\ell}, n_\eta) = (3, 3, 3, 3)$ for the interest rate. We use 8 nodes in each dimension (4096 nodes).

¹⁴We 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. The transform functions $\varphi(\cdot)$ are assumed to be linear $\varphi_z(x) = \kappa_z x$ where κ_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 $\kappa_\eta = 1$.

determined by a minimum weighted residuals method. More precisely, θ^y corresponds to the vector of parameters that solve the projection equation of the residuals¹⁵ of, respectively, the Bellman equation, $\mathcal{R}_R(\theta^V; X_t)$, and Euler equation, $\mathcal{R}_R(\theta^R, \theta^V; X_t)$ on Hermite polynomials.¹⁶ More precisely, we solve¹⁷

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

where $\mathcal{H}(X_t) \equiv H_{i_c}(\varphi_h(C_t)) H_{i_h}(\varphi_h(\hat{x}_t^h)) H_{i_j}(\varphi_\ell(\hat{x}_t^\ell)) H_{i_k}(\varphi_\eta(\eta_t))$ 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. This system of equations is solved by means of a Gauss-Newton algorithm.

This problem involves computing various integrals. These 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).¹⁸ 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.

We follow Judd (1992) and assess the accuracy of our approximation by looking at the Euler equation error

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

Since we are mostly interested in the empirical properties of the model, we mainly evaluate the accuracy of the solution for the data. 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

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

¹⁵See online appendix for more details on these residuals.

¹⁶Note that while the Bellman equation only depends on the parameters θ^V , the Euler equation depends both on θ^R and θ^V , through the belief distortion. We therefore first solve the value function approximation problem, and use the result vector of parameters θ^V to solve for the risky rate problem.

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

¹⁸See Judd (1998), chapter 7.

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 invested in the risky asset. These measures are evaluated using the data, and therefore 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. Results for both models are reported in Table 1 and show that the approximation is accurate.

γ	Known persistence				Unknown Persistence			
	α	E_1	E_2	E_∞	α	E_1	E_2	E_∞
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 1: **Accuracy of the Numerical Solution:** This table reports the measure of accuracy for the Euler equation. In each case, α was set such that the model generates a risk-free rate of 1.5%.

Let us first consider a special case of our model where $\eta_t = 0$, “known persistence”, that is the agent acts as if he knew the persistence parameter ρ were equal to .85. In this case, taking $\gamma = 2$ for example, an agent who uses the approximate solution based on consumption claims would make, on average, a 1 dollar mistake for every \$95,500 invested in the assets, while the maximal error would be of the same order. In the general case of the model, with unknown persistence, the performances of the approximation slightly deteriorate. This accuracy loss is essentially due to the structure of the problem. When persistence is known, the model is almost log-linear, and our approximation performs remarkably well. In the full model, 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.

3.2 Understanding the mechanism of ambiguity aversion

A good way to understand the key channels through which ambiguity aversion affects asset returns in our model is by understanding how the distortion function, ξ , shown

in eqn (5) shapes the “as if” belief of the agent, i.e., the (probabilistic) belief which supports the action chosen by the agent in equilibrium. We identify two main mechanisms. The first works through the endogenous pessimism and added doubt that the “as if” belief embodies, at any one point in time, compared to the belief of an agent with rational expectations based on the processes underlying the specified belief model. The second mechanism is an endogenous accentuation of the cyclical variation in uncertainty.

3.2.1 Endogenous pessimism and doubt

The intuition behind the first channel can be more transparently understood in the special case of the model of beliefs where there is no uncertainty about the persistence (e.g., $\eta_0 = 0$). Under this assumption the argument $(x_{l,t}, \eta_t)$ drops out of the value function described in (10), and the distortion is given as (suppressing “ k ” subscripts):¹⁹

$$\xi_t(x_t | C_t, \hat{x}_t; \alpha) \equiv \frac{\exp(-\alpha(E_{x_t}(V(C_{t+1}; \hat{x}_{t+1}))))}{E_{\hat{x}_t}[\exp(-\alpha(E_{x_t}(V(C_{t+1}; \hat{x}_{t+1}))))]}. \quad (11)$$

The effect of ξ_t is to create an “as if” posterior on x_t , i.e., a distorted posterior, $\tilde{\mu}_t \equiv \xi_t(x_t) \otimes N(\hat{x}_t, \Omega)$, where \hat{x}_t is the filtered value at time t . In the case of ambiguity aversion, i.e., $\alpha > 0$, it is evident from eq. (11) that $\tilde{\mu}_t$ puts relatively greater probability mass (compared to μ_t) on x_t ’s that generate probability distributions associated with lower expected continuation values, $E_{x_t}(V(C_{t+1}; \hat{x}_{t+1}))$. The distorted posterior gives rise to an “as if” conditional one-step-ahead distribution on growth which we call the *twisted (predictive) distribution*

$$g_{t+1} \sim \xi_t(x_t) \otimes N(\hat{x}_t, \Omega) \otimes N(\rho x_t + \bar{g}, \sigma_x^2 + \sigma_g^2). \quad (12)$$

When $\xi_t(x_t) = 1$ the formula (12) 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 belief about growth described by a mixture of normals. The twisted distribution, on the other hand, describes the predictive “as if” belief of an ambiguity-sensitive agent.

Another useful benchmark is the predictive belief of an agent with “rational expectations”, narrowly defined. This distribution, $N(\rho \hat{x}_t + \bar{g}, \sigma_x^2 + \sigma_g^2)$, arises from a posterior that is *degenerate* on \hat{x}_t . As Figure 1 shows, compared to the rational expectations distribution, the twisted distribution has a lower mean *and* a larger spread.

¹⁹Henceforth, we shall write ξ_t as a function of direct continuation value $V(\cdot)$ instead of the indirect value, $J(W_{t+1}, \mu_{t+1})$. In a single agent economy consumption is 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.

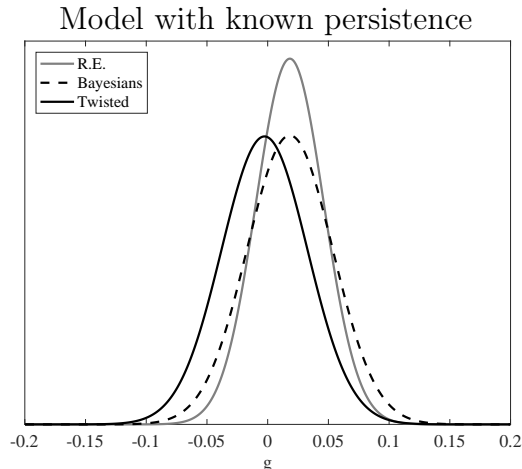


Figure 1: **Beliefs and “as-if” beliefs:** The agent’s “as-if” belief about the conditional distribution of consumption growth with no uncertainty about the latent state (R.E.), with uncertainty about the latent state but without ambiguity aversion (Bayesian) and with ambiguity aversion about the uncertainty of the latent state (Twisted). The distributions were computed using $\rho = 0.85$, and the level of consumption and latent state as the average over 1978–2011.

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 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 while the Bayesian agent only incorporates the doubt. These observations will be the key to understanding our results on time averages of conditional returns moments.

3.2.2 Endogenous accentuation of cyclical variation in uncertainty

To understand the second mechanism we return to the beliefs model without the restriction of $\eta_0 = 0$. Learning about persistence leads to time-varying mixing of the two processes through η_t . This produces a posterior *predictive* belief about consumption growth which is heteroskedastic across time, even though in each process (with a given persistence) the growth distribution is homoskedastic. The mean and variance of the mixture distribution on the latent state are,

$$\hat{x}_t = \eta_t \hat{x}_{l,t} + (1 - \eta_t) \hat{x}_{h,t}, \quad (13)$$

$$\text{Var}_t(x_t) = \eta_t \Omega_l + (1 - \eta_t) \Omega_h + \eta_t (1 - \eta_t) (\hat{x}_{h,t} - \hat{x}_{l,t})^2. \quad (14)$$

It is as if the agent has two forecasting models. When the history is such that both models explain that history just as well (i.e., η_t is close to 0.5) and yet their core forecasts markedly disagree (i.e., $(\hat{x}_{h,t} - \hat{x}_{l,t})^2$ is large), the uncertainty, as shown by the variance, rises. In contrast in the case with $\eta_0 = 0$, what happens over time to the posterior is that its mean \hat{x}_t may change but not its variance, ensuring a homoskedastic predictive distribution.²⁰

The endogenously time varying uncertainty in our model, due to learning about the persistence, creates a potential for *uncertainty shocks*, sudden sharp increases in uncertainty about consumption growth. One way an uncertainty shock can come about is as follows. A sequence of moderately positive growth realizations, being quite consistent with high and low persistent processes, brings η_t close to 1/2. If one or more negative realizations arise after such a sequence, $(\hat{x}_{h,t} - \hat{x}_{l,t})^2$ increases, thus increasing $Var_t(x_t)$. Ambiguity aversion exacerbates the time-variation of the Savage-Bayes uncertainty by endogenously accentuating that uncertainty asymmetrically between positive and negative shocks, creating “as if” uncertainty shocks that are far sharper than what is reflected by the dynamics of $Var_t(x_t)$.

To see how, consider the following. The distorted posterior is a mixture of two component distorted posteriors, $\xi_t^k \otimes \eta_t \otimes N(\hat{x}_{k,t}, \Omega_k)$ for $k = h, l$, where ξ_t^k is as in eq. (28) in Section B.1.2 in the Appendix. Let $\tilde{x}_{k,t}$ denote the mean of a distorted component posterior, $\xi_t^k \otimes N(\hat{x}_{k,t}, \Omega_k)$. Due to the greater persistence, the aggregate uncertainty around $\hat{x}_{h,t}$ – captured by Ω_h – is larger than that around $\hat{x}_{l,t}$. Since the distortion function is proportional to a negative exponential, it has more bite on a distribution which has more probability mass on the left tail by whipping up that mass even more; hence, we have $\hat{x}_{h,t} - \tilde{x}_{h,t} > \hat{x}_{l,t} - \tilde{x}_{l,t}$. Which means that $(\hat{x}_{h,t} - \hat{x}_{l,t})^2 > (\tilde{x}_{h,t} - \tilde{x}_{l,t})^2$ when $\hat{x}_{h,t} > \hat{x}_{l,t}$ (as would be, following a positive shock) and $(\hat{x}_{h,t} - \hat{x}_{l,t})^2 < (\tilde{x}_{h,t} - \tilde{x}_{l,t})^2$ when $\hat{x}_{h,t} < \hat{x}_{l,t}$ (following a negative shock). Hence, when $\hat{x}_{h,t} < \hat{x}_{l,t}$, the components of the mixture yielding the “as if” posterior are further apart compared to the components of the Bayesian posterior (and, conversely, when $\hat{x}_{h,t} > \hat{x}_{l,t}$). This has two implications. One, $\widetilde{Var}_t(x_t)$, the variance of the distorted posterior²¹ understates that of the Bayesian posterior following a positive shock, and exaggerates it following a negative shock, making it more pronouncedly counter-cyclical than $Var_t(x_t)$. Two, the distorted posterior demonstrates a significant negative skewness compared to the Bayesian posterior in recessionary periods, but not in good times.

²⁰The time-varying heteroskedasticity generated endogenously in our model is a *forecast* uncertainty, of beliefs, empirically driven by the history of growth outcomes and consistent with a stationary volatility of consumption shocks.

²¹See Section B.1.1 in the Appendix for an analytical expression.

The left panel in Figure 2 shows how $\hat{x}_{h,t}$ and $\hat{x}_{l,t}$ have moved with the business cycle. The right panel compares the variance of the posterior and the variance of the distorted posterior showing that the latter greatly amplifies movements in the former, especially at downturns. Figure 2 also shows that in 1992 $\hat{x}_{h,t} < \hat{x}_{l,t}$ while in 1999 $\hat{x}_{h,t} > \hat{x}_{l,t}$, though $|\hat{x}_{h,t} - \hat{x}_{l,t}|$ were similar in these two years. Figure 3 demonstrates how much more significant the effect of the distortion was on the posterior in the latter year.

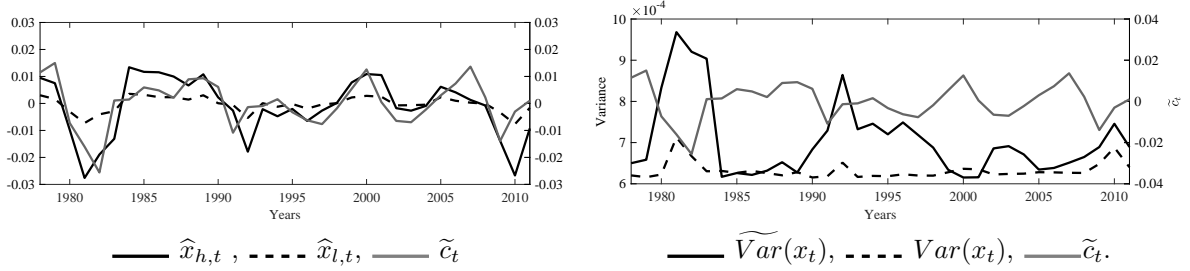


Figure 2: **Explaining time-varying ambiguity:** The left panel shows the filtered latent variables assuming that the high ($\hat{x}_{h,t}$) and low ($\hat{x}_{l,t}$) persistence as the DGP. The right panel graphs the conditional variance of the latent state variable ($Var_t(x_t)$) and the “as if” conditional variance ($\widetilde{Var}_t(x_t)$). In both panels the gray line shows the HP-filtered consumption growth, indicating the business cycle.

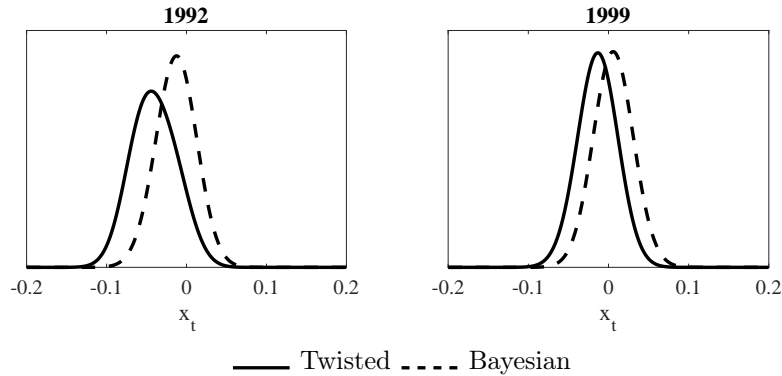


Figure 3: **Time-varying distortion:** The two panels plot beliefs about the latent state without ambiguity aversion (Bayesian) and with ambiguity aversion. The left panel shows a “bad” year where $\hat{x}_{h,t} < \hat{x}_{l,t}$, and the right panel shows a “good” year where $\hat{x}_{h,t} > \hat{x}_{l,t}$.

The following argument focused on the uncertainty about ρ_k offers another, and perhaps pithier, intuition. The ambiguity-averse agent behaves as if he forecasts consumption growth putting more weight (compared to the Bayesian posterior) on the “worst case” persistence, i.e., the ρ_k that minimizes the expected continuation utility. When consumption growth is below the mean, the worst case persistence parameter

is ρ_h , suggesting that we will remain below the mean for a long time. In contrast, when consumption growth is above the mean, the worse case is that the persistence is ρ_l , so we revert quickly to the mean. Thus, the ambiguity-averse agent, endogenously behaves as if the uncertainty is more persistent and severe following negative shocks than in normal times (*even though* $\eta_t \simeq 1/2$). These insights about the asymmetric reaction to good and bad news will be key to understanding how ambiguity aversion affects conditional returns and their variation over time, in particular, over the business cycle.

3.3 Comparing model implications with data

We use annual data on real per-capita consumption C_t and estimates of $\widehat{x}_{k,t}$ corresponding to the filtration imposed by the observed history of growth of real consumption and of real dividends to obtain a time series of model implied *conditional moments* of the annual rates of return using our numerical solution technique (see Section 1 in the Online Appendix.) We compute the model implied price-dividend ratio applying the relationship

$$R_{t+1} = \frac{\exp(p_{t+1} - d_{t+1}) + 1}{\exp(p_t - d_t)} \exp(d_{t+1} - d_t) \quad (15)$$

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 ($t = 0$). Throughout the exercise, the level of ambiguity aversion was calibrated so that the average risk-free rate was 1.5%.

We present and discuss two kinds of results on model implied conditional moments of rates of return and price-dividend ratio: averages of the moments over the sample period, 1978-2011 in Section 3.3.1, and time series (and time series properties) of the moments over the same sample period in Section 3.3.2. In Section 3.3.3 we compare the time-series of our model implied equity premium with the leading macroeconomic uncertainty index in the literature.

3.3.1 Time averages of moments

Table 2 reports the model implied conditional moments of returns and price-dividend ratio, time averaged over the sample period.²² The panels in Figure 4 show the comparative statics of ambiguity aversion and risk aversion on the conditional rates of return.

²²When trying to infer how ambiguity aversion affects returns/prices from the entries in Table 2, bear in mind α is calibrated so that r_f is equal to 1.5%. So, when γ is changed, α does too to ensure calibration. From Table 2, if one wants to infer anything about a change in ambiguity aversion alone, then one can compare what happens in the Bayesian case ($\gamma = 2.5$ and $\alpha = 0$) with our benchmark case ($\gamma = 2.5$, $\alpha = 11.3$).

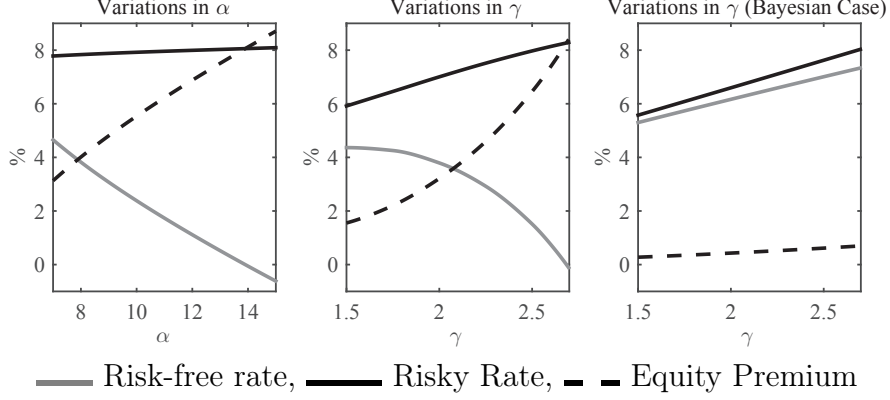


Figure 4: **Comparative statics** : In the left panel, α varies with γ fixed at 2.5. In the middle panel, α was fixed at 11.3 and γ varies. The average comparative statics are constructed by first computing the comparative statics for each year using the filtered values \hat{x}_t and then averaging across $t = 1978, \dots, 2011$. The right panel depicts the Bayesian case, i.e., with $\alpha \approx 0$. (The graphs correspond to our model with unknown persistence.)

The model’s match of the first moments of returns is quite perfect and second moments are predicted to a large extent.

Table 3 reports on some further robustness checks. In particular, it checks for the consequences from alternative learning assumptions that would be implied by splitting the sample differently. As was explained, our benchmark calculations are obtained by splitting the 1930-2011 sample between a “learning period”, 1930-1977, where the time-series parameters were estimated, and the remaining period where the model was evaluated, i.e., the benchmark split is (1930-1977;1978-2011). Table 3 considers three alternative splits: (1930-1968; 1969-2011), (1930-1959; 1960-2011), (1930-1950; 1951-2011). In addition, the Table 3 allows for alternative pairs of values of high and low persistence parameter. As we can see, findings on average of conditional moments of rates of return are remarkably robust to alternative learning assumptions implied by the different sample splits.

To help us understand these results (which were obtained numerically) we consider analytical approximations²³ for the rates of return for the case where persistence is known, e.g., with $\eta_t = 0$. The risk-free rate is approximated as:

$$r_t^f = -\ln \beta + \gamma \bar{g} + \gamma \rho \tilde{x}_t - \frac{\gamma^2}{2} \left(\sigma_x^2 + \sigma_g^2 + \rho^2 \widetilde{Var}_t(x_t) \right) \quad (16)$$

where \tilde{x}_t is the mean of the distorted posterior at time t .

An increase in ambiguity aversion, α , decreases \tilde{x}_t making the agent behave as if he were expecting a lower endowment income in future states. Implying, a rise

²³See Appendix C for details of the derivation.

Returns and Volatility								
	γ	α	$E(r)$	$E(r - r^f)$	$\sigma(r^f)$	$\sigma(r)$	$\sigma(r - r^f)$	
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	≈ 0	7.62	0.62	1.70	23.1	23.2	
Price-Dividend Ratio								
	γ	α	$\mathbb{E}(P/D)$	$\sigma(P/D)$	$\mathbb{E}(p - d)$	$\sigma(p - d)$	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	≈ 0	40.0	11.5	3.65	0.30	0.82	0.75
Bayesian, $\beta = .97$	2.5	≈ 0	46.3	16.5	3.77	0.37	0.86	0.79

Table 2: The top panel contains the average of the predicted *conditional* moments of rates of return (on dividend claim) for different values of γ and calibrated α . Immediately below is a series of robustness checks where the parameter in the left-most column was changed from the basic specification ($\rho_h = 0.85$, $\rho_l = 0.3$, $\psi = 3$, $\beta = 0.975$), taking $\gamma = 2.5$ as part of the baseline specification. The bottom panel contains the time-averaged model implied price/dividend ratio statistics over the period 1978–2011. AC1 and AC2 denote the first and second order autocorrelation of $p - d$.

Returns and Volatility							
(ρ_l, ρ_h)	γ	α	$E(r)$	$E(r - r^f)$	$\sigma(r^f)$	$\sigma(r)$	$\sigma(r - r^f)$
Data			8.08	6.68	2.20	16.5	16.1
<i>1969-2011 evaluation period</i>							
(0.30,0.85)	2.00	17.100	7.475	5.981	2.398	23.056	23.056
(0.30,0.85)	2.50	10.900	8.084	6.575	3.061	23.661	23.653
(0.30,0.85)	3.00	6.400	8.775	7.277	3.699	24.352	24.329
(0.25,0.85)	2.50	11.050	8.096	6.602	2.937	23.766	23.730
(0.30,0.90)	2.50	7.390	8.036	6.534	3.420	23.881	23.842
<i>1960-2011 evaluation period</i>							
(0.30,0.85)	2.00	15.050	7.064	5.555	2.276	24.412	24.409
(0.30,0.85)	2.50	9.500	7.592	6.089	2.863	25.088	25.074
(0.30,0.85)	3.00	5.535	8.196	6.690	3.430	25.845	25.815
(0.25,0.85)	2.50	9.150	7.634	6.128	2.671	25.244	25.204
(0.30,0.90)	2.50	6.030	7.630	6.121	3.105	25.401	25.354
<i>1951-2011 evaluation period</i>							
(0.30,0.85)	2.00	24.270	6.270	4.763	4.612	27.183	27.464
(0.30,0.85)	2.50	15.100	6.888	5.389	5.594	28.150	28.490
(0.30,0.85)	3.00	8.695	7.587	6.085	6.473	29.254	29.634
(0.25,0.85)	2.50	15.100	6.957	5.456	4.738	28.327	28.538
(0.30,0.90)	2.50	7.840	7.075	5.574	3.859	28.711	28.760

Table 3: Robustness of returns moments to the sample split

in demand for the risk-free asset (a “flight to quality”, as termed by Caballero and Krishnamurthy (2008)) driving up its equilibrium price and lowering the risk-free rate. The accentuation of doubt, working through $\widetilde{Var}_t(x_t)$ reinforces the effect. This is a key effect of ambiguity aversion. Note, when $\alpha > 0$ the term $\gamma\rho\tilde{x}_t$ acts to dampen the effect of $\gamma\bar{g}$, making the comparative static of γ on the risk free rate very different, qualitatively and quantitatively, depending on whether $\alpha > 0$ or $\alpha = 0$, as a comparison of the middle and right panels of Figure 4 shows. Hence, it is not possible to replicate the effect of ambiguity aversion by turning it off and simply varying γ .

The first moment of the risky rate is approximated as

$$E_t r_t = \text{Const}_1 + \rho(\gamma - \psi)\tilde{x}_t + \psi\rho\hat{x}_t - \frac{\rho^2}{2} [(\gamma - \psi)^2 \text{Const}_2] \widetilde{Var}_t(x_t) \quad (17)$$

where $E_t \equiv E_{\hat{x}_t} E_{x_t}$ 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 $t+1$. Const_1 and Const_2 collect terms which are constant across time and not affected by ambiguity aversion. An increase in α has two countervailing effects. The first effect, given by $\rho\gamma\tilde{x}_t$, was also present in the expression for the risk-free rate; the intuition here is analogous. The second effect is in the term $-\rho\psi\tilde{x}_t$. As α increases \tilde{x}_t decreases, hence decreasing the (“as if”) expected future dividend payoff from the asset causing the agent to want to pay less for the asset. With $\gamma \leq 3$ and $\psi = 3$, as we have here, the second effect dominates (very slightly) and equilibrium risky rate varies positively (but quite minimally) with ambiguity aversion.

The approximation for the equity premium may be written as

$$E_t r_t - r_t^f = \text{Const}_3 + \psi\rho(\hat{x}_t - \tilde{x}_t) + \frac{\rho^2}{2} [\gamma^2 - (\psi - \gamma)^2 \text{Const}_2] \widetilde{Var}_t(x_t). \quad (18)$$

where we have explicitly left the two terms which are affected by ambiguity aversion, $(\hat{x}_t - \tilde{x}_t)$ and $\widetilde{Var}_t(x_t)$. The first term shows that the premium increases with ambiguity aversion (the difference $(\hat{x}_t - \tilde{x}_t)$ increases when α is increased) and the magnitude of this effect is accentuated by persistence and leverage. A doubt factor also comes into play (principally) through its effect on the risk free rate, discussed earlier. Since, the risk free rate is conditionally non-stochastic, the conditional volatility of equity premium coincides with that of the risky rate. The overwhelming factor fixing the (average) conditional volatility of risky return is the volatility of the dividend claim, in turn determined by the volatility of the latent state multiplied by ψ and ρ .

To summarize, ambiguity aversion gets the first moment of equity premium right by holding the risk free rate down while affecting the risky rate only very marginally. The volatility comes from two sources, the uncertainty about the latent state accentuated by the uncertainty about the persistence and the leverage factor.

3.3.2 Time series profiles of conditional rates of return and price-dividend ratio

Perhaps the more distinctive results of the analysis in this paper concerns the time series of conditional moments. These are largely driven by dynamics of the “as if” belief explained in Section 3.2.2. Figure 5 demonstrates this quite vividly in the case of the equity premium, especially the crucial role of the uncertainty about persistence. Studies have estimated conditional moments of equity premium on historical data, notably Whitelaw (1994) and Lettau and Ludvigson (2010). The former summarizes a key finding as follows (pp 526; in the quote “expected return” is the conditional first moment of equity returns in excess of risk free rate) :

The expected return seems to reach a maximum at the trough of the business cycle and reach a minimum before, or at, the peak of the business cycle. Expected returns appear to decrease during economic expansions and increase during economic contractions. In contrast, the conditional volatility appears to reach a maximum earlier in the business cycle, at or slightly after the peak in the cycle, and to reach a minimum just after the business cycle trough.

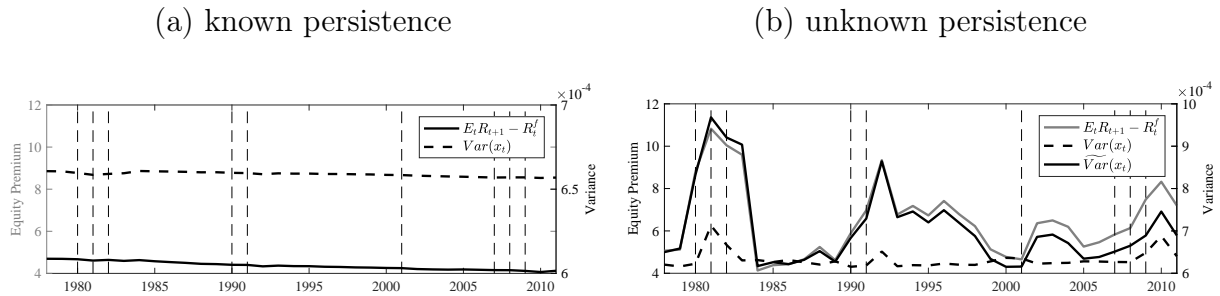


Figure 5: **Movements in variance and model implied equity premium:** Panel (a) shows the conditional equity premium and the conditional variance of the “as-if” posterior from the model with known persistence, $\rho = 0.85$. Panel (b) shows the same as well as the variance of the undistorted posterior for the model with unknown persistence. Vertical dashed lines indicate years featuring a recession.

Figures 5(b) and 6(a) show how well the series predicted by our model match the above quote. Equity premium, as predicted by the model, is counter-cyclical; its correlation with H-P filtered consumption growth is -0.59. Whitelaw (1994) estimates the contemporaneous correlation between the first and second (conditional) moments (of equity premium) to be -0.34; based on the data considered for this paper, which pertains to a different time period and frequency, the correlation of the same two statistics in our model is -0.86.

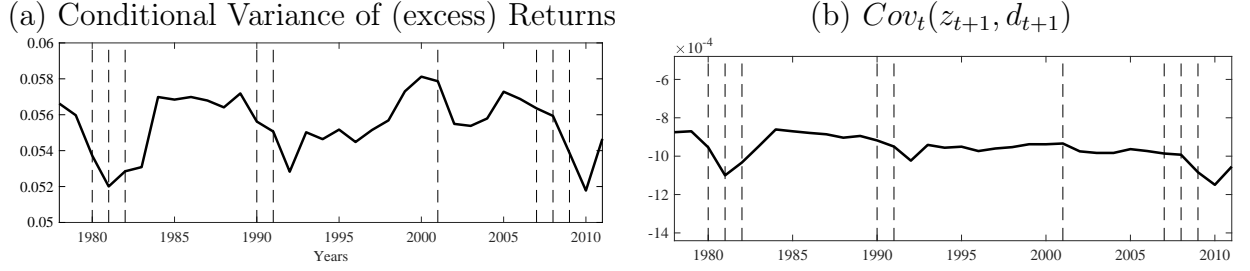


Figure 6: The two panels depict the conditional variance of (excess) returns and $Cov_t(z_{t+1}, d_{t+1})$, implied by the model, demonstrating the close link between the dynamics of the two. Vertical dashed lines indicate years featuring a recession.

What accounts for the pro-cyclical volatility of returns in our model? Starting from the standard approximation for the risky rate (eqn. 33 in Section C in the Appendix), its variance may be seen to be composed as:

$$Var_t(r_t) \simeq \kappa_1^2 Var_t(z_{t+1}) + Var_t(d_{t+1}) + 2\kappa_1 Cov_t(z_{t+1}, d_{t+1}).$$

(In our data, $\kappa_1 = 0.98$.) It turns out the time averaged variance is completely swamped by the term $Var_t(d_{t+1})$ ($\overline{Var}_t(r_t) = 0.0555, \overline{Var}_t(d_{t+1}) = 0.0541$ and $\overline{Var}_t(z_{t+1}) = 1.17e - 4$). However, as seen from Figure 6, the dynamics of $Var_t(r_t)$ are very largely determined by $Cov_t(z_{t+1}, d_{t+1})$. To see an intuition why this covariance is negative and even more so in recessionary times, note, belief about d_{t+1} is determined by the Bayesian posterior, with mean \hat{x}_t , while z_{t+1} is guided by \tilde{x}_t , the mean of the distorted posterior. As explained in Section 3.2.2, \tilde{x}_t is below \hat{x}_t , and even more so and less mean reverting (i.e., more persistent) than \hat{x}_t in recessions. Hence, in recessions there is a bigger measure of events where d_{t+1} realizes above its mean while z_{t+1} stays below its mean. The price-dividend ratio is function of the agent's view of the longer term prospects while the dividend is just the outcome in the next period; the former may remain relatively downbeat and sluggish, especially in recessionary times, despite a positive outcome of the latter.

Together, the countercyclical variation of the mean and the increase in volatility during recessions leads to countercyclical variation of the conditional Sharpe ratio, $E_t(r - r^f)/\sigma_t(r - r^f)$. The Sharpe ratio rises from the peak to the trough of every completed business cycle in the data and in our model implied series. Lettau and Ludvigson (2010) investigate how leading, established asset pricing models explain this time-series behavior of the conditional Sharpe ratio. They find that neither the Bansal and Yaron model nor the standard model with constant relative risk aversion and time-varying consumption volatility matches the dynamic behavior of the empirical Sharpe ratio: the models predict a conditional Sharpe ratio that is negatively correlated with

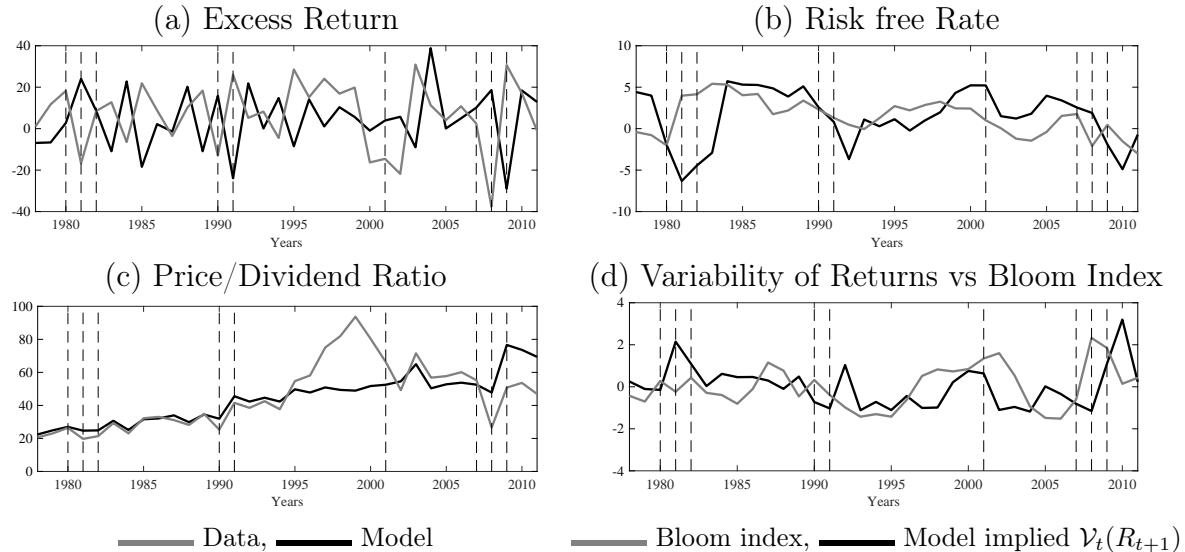


Figure 7: **Returns and Price–Dividend Ratio**: Panel (a) contains a plot of the model-implied excess return along with the actual excess return. Panel (b) shows the model-implied risk-free rate along with the actual real risk-free rate. Panel (c) contains the actual and model implied price-dividend ratios. Panel (d) shows the time-series of $\mathcal{V}_t(R_{t+1}) \equiv \sqrt{E_t(R_{t+1} - ER_{t+1})^2}$ and stock market volatility index constructed in Bloom (2009). For comparison purposes, both the Bloom index and $\mathcal{V}_t(R_{t+1})$ are normalized by their respective mean levels. So, on the vertical axis, we measure the (signed) percentage deviations from the respective means.

the empirical Sharpe ratio, “because both models are linear functions of the consumption volatility, which itself is negatively correlated with the Sharpe ratio for the U.S. stock market”. The prediction of our model is very similar to Campbell and Cochrane’s Habits model, it has the right shape over time and in relation to business cycles but amplitudes are less pronounced than in data. However, different from Campbell and Cochrane, ours has a lower and more realistic autocorrelation.

While equity premium is not directly observed, 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).²⁴ This sets out a stark, stiff test for the model. The predictions are evidently good, especially for returns but reasonably good too for the price-dividend ratio. The correlation of the realized risky rate and excess return with \hat{x}_t is -0.08 and -0.1 in data compared to -0.07 and -0.21, respectively, in the model

²⁴The 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.

prediction. The instantaneous correlation between R and $(p - d)$ is positive in the data (0.54) and in the model 0.66. The correlation of the linearly detrended (in logs), HP-filtered (in logs) and unfiltered predicted price-dividend ratio and the correspondingly treated price-dividend observed in data are 0.67, 0.77 and 0.83, respectively. However, the prediction does not match the data in 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 in our model, where prices are determined in general equilibrium entirely based on the stochastic evolution of *real* output. In this respect, it is significant that the predicted price-dividend returns to the actual path following the collapse of the bubble.

Panel (d) 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 unconditional²⁵ mean. One may consider this a measure of variability of risky returns and as shown, it is a good match with Bloom (2009)'s measure of stock market volatility (the correlation between the two is 0.38).

Excess returns 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 4. 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. Table 5 allows a comparison, between the data and the model implications, of coefficients from predictive regressions of annual returns on lagged price-dividend ratio. The estimated coefficients match sign and while the model implied coefficients are smaller they are within the 95% confidence interval of the corresponding estimates in the data.²⁶

Thus there is suggestive but not strong evidence of stock return predictability by $p - d$ ratio in our model. However, it is worth noting that stock return predictability is not always visible in subsamples (see Goyal and Welch (2008)). As Koijen and Van Nieuwerburgh (2011), p.8, remarks, “significant instability over time (...) in other words, for thirty year sample ending in between 1965-1995, there was evidence

²⁵More 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 .

²⁶The estimates of coefficients from model implied values are fragile since the nature of the exercise limits us to historical sample points and hence very few observations. In the literature, predictability regressions are typically run on data obtained from model simulations; Beeler and Campbell (2012), e.g., use a million such data points.

Cumulative Autocorrelation					
Lag, in years	1	2	3	5	7
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 4: **Mean reversion of returns:** Autocorrelation structure of excess returns in the data and as implied by the model (baseline specification). The cumulative autocorrelation is defined as $\left(\sum_{i=1}^j \text{Corr}l((R_t - R_t^f), (R_{t-i} - R_{t-i}^f))\right)$.

for stock return predictability but this evidence disappeared after 1995. It was absent for pre-war period as well.” Nevertheless, a significant part of the literature gives return predictability a status of stylized fact (see Cochrane (1999) and Cochrane (2008) who argue it’s hard to make sense of the time series properties otherwise.) We take the view, given the short period of model evaluation, it is difficult for the model to really address stock return predictability.

We turn now to some other indicators that shed light on the question whether the model implies the right variation in expected stock returns and expected dividend growth rates. Too little persistence in the $p - d$ ratio is usually taken as a sign of too little variation in expected stock returns. However, as shown in Table 6 the model implied $p - d$ has a high persistence that matches the data very well. Does the model generate too much predictability in dividend growth rates by the $p - d$ ratio? Table 7 reports results of running a regression of dividend growth on the lagged $p - d$ ratio at various horizons and compares the outcomes to the data, and demonstrates, if anything, the model implications have slightly less predictability than in the data. Consistent with this lack of dividend growth predictability is the evidence from a Campbell-Shiller variance decomposition that the estimate of proportion of variation in model implied $p - d$ explained by variation in dividend is about as much as it is in data (for an 8-year horizon, along our sampled history, in the model it is 21% and in the data 29%, approximately). However, the evidence is not conclusive because the standard errors of the estimates are quite high. Relatedly, as shown in Table 7, consumption growth too is unpredictable in data and in our model, unlike in the Bansal and Yaron (2004) model, for example, which implies significant predictability of consumption growth by price-dividend ratios. This excess predictability, which has been seen as a weakness of long-run risk models (see Beeler and Campbell (2012)), is not present when there is uncertainty about the persistence parameter and learning. Finally, as Table 8 shows, price-dividend ratio is not predicted by consumption growth, neither in data nor in our model, drawing a sharp distinction with the implication of habit formation models

(e.g., Campbell and Cochrane (1999)) where consumption growth strongly predicts price-dividend ratio.

$$\sum_{n=1}^N r_{t+n} = \theta_0 + \theta_p(p - d)_t + \varepsilon_{t+n}$$

		Data	Model
N	θ_p	95% C.I.	θ_p
3	-0.56	[-1.30;0.18]	-0.07
5	-1.03	[-2.03;-0.02]	-0.14

Note: standard errors are robust to heteroskedasticity and autocorrelation.

Table 5: **Predictability Regression Coefficients (1978–2011):** The table reports coefficients from predictive regressions of annual returns on lagged price-dividend ratios over the sample period, 1978-2011, in the data and in the time-series implied by the model. The third column shows the 95% confidence interval on the estimated regression coefficient.

k	P/D in Level			P/D in Logs.		
	1	2	4	1	2	4
Data	0.84	0.76	0.58	0.80	0.76	0.64
Model	0.82	0.72	0.59	0.85	0.78	0.66

Table 6: **Price/Dividend Ratio, autocorrelation**

3.3.3 Equity premium and macro-uncertainty measures

In this section we show one more way to assess the model’s performance in matching historical data. The model implied equity premium is the conditional expectation of the model implied return of a share of the equity in excess of the (model implied) risk free return. The risk free return may be understood as the return under the assumption the asset delivers the conditionally expected (or, the forecasted) payoff *for sure*. The premium is the compensation for the uncertainty that the equity may deliver a payoff different from what is forecasted, hence a compensation for possible forecast error. Since the taste parameters (e.g., attitudes toward time and uncertainty) have been held fixed across time in the model, we may interpret the movements in equity premium to be driven by coincident movements in the perceived macroeconomic uncertainty. Thus, the model generated conditional equity premium is an index measure of the conditional macroeconomic uncertainty revealed by equilibrium behavior, the

Dividend Growth					Consumption Growth			
Data			Model		Data		Model	
k	R^2	p-value	R^2	p-value	R^2	p-value	R^2	p-value
Price/Dividend in Logs.								
1	0.19	0.0096	0.09	0.0911	0.06	0.1558	0.00	0.8484
2	0.55	0.0000	0.29	0.0050	0.11	0.1509	0.00	0.9608
4	0.57	0.0000	0.35	0.0108	0.28	0.0404	0.14	0.3212
8	0.69	0.0001	0.52	0.0092	0.54	0.0059	0.30	0.2675
Price/Dividend in Levels								
1	0.15	0.0215	0.09	0.0875	0.07	0.1302	0.00	0.9598
2	0.46	0.0001	0.21	0.0258	0.13	0.1111	0.01	0.9156
4	0.48	0.0006	0.26	0.0602	0.22	0.1152	0.13	0.3828
8	0.61	0.0011	0.54	0.0062	0.53	0.0074	0.36	0.1343

Table 7: **Predictability Regressions:** This table reports the R^2 and the p-value of the global significance test of the regression $y_t = \alpha_0 + \sum_{i=1}^k \alpha_i PD_{t-i}$, $y = d, g$, where $H_0 : \alpha_i = 0 \forall i = 1 \dots k$. PD_{t-i} is the i -th lag of the price dividend ratio.

Data			Model	
L	p-val.	R^2	p-val.	R^2
1	0.5369	0.01	0.3319	0.03
2	0.8138	0.01	0.5821	0.03
4	0.9462	0.02	0.7677	0.06
8	0.9963	0.04	0.9637	0.08

Table 8: **Price-Dividend Ratio and Backward Consumption Growth:** This table reports results for the regression $(p - d)_{t+1} = \alpha_0 + \sum_{j=1}^L \alpha_j g_{t+1-j} + u_{t+1}$. p-val denotes the p-value associated to the joint significance test of $H_0 : \alpha_j = 0$ for $j = 1 \dots L$. Predictability is rejected at any lag.

perceived uncertainty. Thus, how well the model predicted historical *perceived uncertainty* matches actual indices of such uncertainty available in the literature presents a test of the model.

Jurado, Ludvigson, and Ng (2015) (henceforth JLN) construct an index of macroeconomic uncertainty by averaging the (conditional) uncertainty of the forecast errors of 132 variables selected to represent broad categories of macroeconomic time series: ranging from real output, employment, real retail, labor compensation, price indexes to financial market indexes. The conditional uncertainty in each variable is a moment measure: the conditional volatility of the *unforecastable* component of the future value of the series, with the property that if the conditional expectation of the squared error in forecasting the future value rises, uncertainty in the variable increases. The average of these uncertainties captures the common variation in uncertainty across the many series, and hence the macro-uncertainty. Note, since the measure accounts for the endogenous economic response to uncertainty (of endogenous price and quantity variables) it is a measure of perceived uncertainty. Hence, it is comparable to our computed equity premium, which is also a measure of endogenous perceived uncertainty revealed through equilibrium asset prices. Actually, in footnote 2, JLN speculate that their measure could be a result of Knightian uncertainty, “in which agents are uncertain about the probability distribution itself”.

As panels (a) and (b) of Figure 8 show, JLN’s conjecture is largely vindicated since the JLN index and our model implied conditional equity premium are closely related: the correlation is 0.58 for both levels and differences.²⁷ Contrastingly, the conditional equity premium implied by the Bayesian case (i.e., by setting $\alpha \simeq 0$) yields correlations of -0.02 and 0.11 , for levels and differences, respectively. We have already noted the pronounced counter-cyclical of the model implied conditional equity premium (a correlation of -0.61 with the Kalman filtered latent variable). The JLN index is similarly counter-cyclical, with a correlation of -0.60 with the filtered value of our latent variable. Another salient feature is persistence: both series are persistent but the *persistence is significantly greater in years with recessionary episodes*, as the numbers reported in the first two rows of Table 9 show. The final row of the table shows, in contrast, that the model generated conditional equity premium in the Bayesian case demonstrates no significant difference in persistence across the business cycle.

Figure 9 demonstrates the close dynamic relationship between price dividend ratio

²⁷The JLN uncertainty measure is available monthly, whereas our conditional equity premium is an annual measure. We compare the equity premium to the trailing 12 month average of the monthly JLN measure. This is done to facilitate a more realistic alignment of when data is made available to market participants. However, the adjustment is far from perfect. Our equity premium variable, by construction, is based on the annual GDP growth report, and hence effectively shows the uncertainty lagged by about a year. This is worth bearing in mind when looking at the graphs.

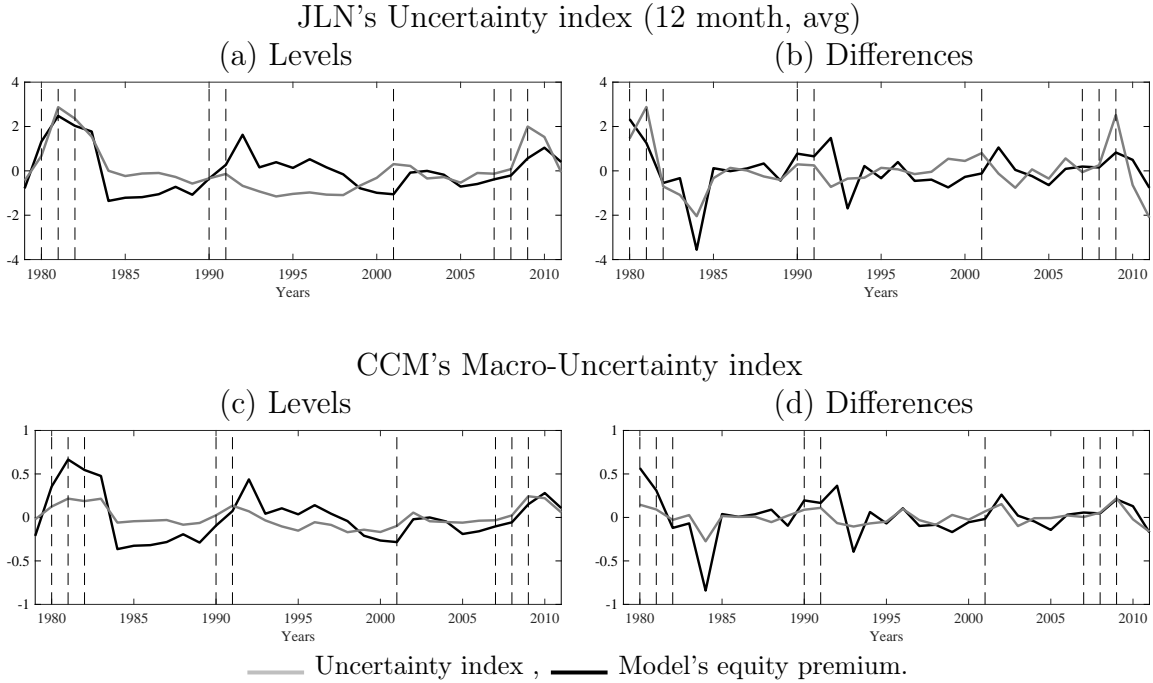
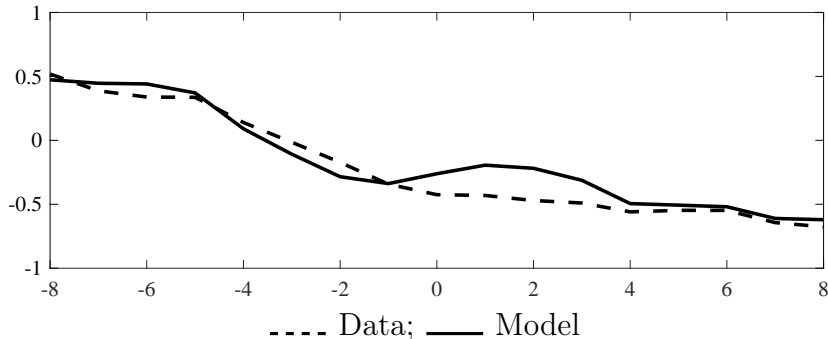


Figure 8: Comparing time-series (levels in panels (a) and (c) and differences in panel (b) and (d)) of Model implied Equity Premium and Uncertainty Indices. For comparison purposes, both the uncertainty index and the equity premium are normalized by their respective mean levels. So, on the vertical axes, we measure the (signed) percentage deviations from the respective means. The (dashed) vertical bars indicate years with at least one NBER declared recession episode.

	Persistence in years		p-value	Level of
	w/ recession	w/o recession	Test	Significance
	(1)	(2)	(1)-(2)	
Model Cond. Eqty. Prm.	0.76	0.55	0.004	100%
	(0.12;6.19)	(0.12;4.75)		
JLN Uncertainty index	0.89	0.48	0.272	73%
	(0.22;4.06)	(0.26;1.84)		
Bayes case Eq Prm ($\alpha \simeq 0$)	0.68	0.69	0.436	— %
	(0.13;5.18)	(0.13;5.16)		

Table 9: **Counter-cyclical Persistence:** Columns 2 and 3 show estimates, corresponding to the time-series indicated in column 1, of the AR(1) parameter and between parenthesis its standard deviation and the associated Student-t statistic in years with and without recessionary episodes, respectively; the final columns show the p-value of the test for statistical significance of the difference in estimates in columns 2 and 3 and the associated level of significance. The final row of the table shows these numbers for the series obtained from the model with ambiguity neutrality (i.e., $\alpha \simeq 0$).

Figure 9: Dynamic Correlations of (log) Price/Dividend ratio with JLN Uncertainty Index



Note: The graphs report the correlations $\text{corr}(JLN_t, (p-d)_{t+k})$, for $k = -8, \dots, 8$, where JLN_t is the JLN uncertainty index and $(p-d)_{t+k}$ is the log of the price/dividend ratio (in data and model implied) evaluated at various leads and lags.

and the JLN uncertainty index *both* in the data and in the model implied series. The interpretation of the graphs is simple. For example, a high uncertainty today (i.e., high JLN_t , $t = 0$) is foreshadowed in a lower price dividend ratio with a lead of up to three periods ($(p-d)_{t+k}$, $k = 0, -1, -2, -3$); and, it depresses prices with a lag of up to 6 periods ($k = 1, \dots, 6$). However, prices are not adversely affected by anticipation of uncertainty at horizons of four and more years, both in our model and in the data.

JLN emphasize in their concluding remarks that the key features of macroeconomic uncertainty are its counter-cyclical and its persistence during recessions. These two features speak directly to the mechanism at work in our model. As we showed, the Bayesian uncertainty does increase, if minimally, following a shock; but *that increase is symmetric* with respect to the sign of shock. It is ambiguity aversion that is responsible for the *asymmetric* behavioral response to good and bad news and for increasing the (“as if”) belief on high persistence in recessionary periods, the key mechanism in our model. Could these features obtain in a model with stochastic volatility but no ambiguity aversion? As discussed earlier, investigations have shown that the evident consumption volatility in data has neither the right variation over time nor the size needed to explain the observed time variation in equity premium and the Sharpe ratio.

Carriero, Clark, and Marcellino (2017) (CCM henceforth) construct a measure of macro-uncertainty based on a large vector autoregression with stochastic volatility driven by common factors representing macroeconomic uncertainty. The index reflects changes in both the conditional mean and volatility of the variables. An advantage of this approach over JLN is that, the authors argue, it reduces the risk of biases

and endogeneity problems stemming from measurement errors and omitted variables. As panels (c) and (d) of Figure 8 show, compared to the JLN measure this index and our model implied conditional equity premium are even more closely related: the correlation is 0.73. both in levels and in differences.

Recently, Orlik and Veldkamp (2014) have constructed a measure of macroeconomic uncertainty which also comes with a theory why such uncertainty is more countercyclical than stochastic volatility alone. In their model the agent does not know the true distribution of macroeconomic outcomes, but estimates its parameters in the way of a Bayesian econometrician using real time (GDP) data. They measure uncertainty as the conditional standard deviation of GDP growth, which captures uncertainty about the distributions' estimated parameters. When the forecasting model admits only normally-distributed outcomes, they find small, acyclical changes in uncertainty. But when the forecasting model is enlarged in a specific way, so that agents also estimate parameters that regulate skewness, uncertainty fluctuations become more pronouncedly counter-cyclical. However, they find the uncertainty diminishes secularly and significantly due to the learning of the parameters. To rectify this they add an exogenously specified stochastic volatility component which, like in Bansal and Yaron (2004), has a persistence that is *independent* of the business cycle. They report that their measure has a correlation of 0.31 with the JLN uncertainty index (recall, for our model this correlation is 0.58).

4 Assessing the calibrated value of ambiguity aversion

Here we discuss a way of assessing the plausibility of the calibrated levels of ambiguity aversion in terms of implied individual (as opposed to market) behavior.

In standard analysis of the equity premium question, the value of (relative) risk aversion parameter is motivated by using a thought experiment; the typical question being how much an agent would pay to avoid a given risk. Arguably, neither the question nor the intuitive answer refers to the expected utility model, or any formal model of decision making for that matter. We now consider as a thought experiment the implied uncertainty premium of an individual investor with preferences and dynamic belief, precisely like the agent in our model, evaluating a Lucas tree prospect. We find the investor is willing to pay an overall uncertainty premium (a sum of the risk premium and the ambiguity premium) that is well within the bounds of what is regarded as intuitively plausible per the *standard* intuition and analysis.

Our thought experiment consists of an offer at time t to our Lucas economy agent,

with preference parameters (γ, α, β) , to replace the uncertain consumption prospect he faces with a fixed consumption in each period, now and for ever. Define the *consumption certainty equivalent*, $c^*(\gamma, \alpha, \beta; c_t)$, to be the c^* that makes the agent indifferent, given information at t , between the plan (c^*, c^*, c^*, \dots) and his endowed stochastic consumption plan (c_t, c_{t+1}, \dots) . Hence, $c^*(\gamma, \alpha = 0, \beta; c_t)$ is the certainty equivalent for the Bayesian agent and $c^*(\gamma = 0, \alpha = 0, \beta; c_t)$ is the certainty equivalent for a risk neutral agent [hence, sum of discounted conditional expectations.] The *risk premium* is $\mathcal{R}(\gamma, 0, \beta; c_t) \equiv c^*(0, 0, \beta; c_t) - c^*(\gamma, 0, \beta; c_t)$, and the *ambiguity premium* is $\mathcal{A}(\gamma, \alpha, \beta; c_t) \equiv c^*(\gamma, 0, \beta; c_t) - c^*(\gamma, \alpha, \beta; c_t)$. Then, the *total uncertainty premium* paid by our agent with preference parameters (γ, α, β) is given by

$$\mathcal{U}(\gamma, \alpha, \beta; c_t) \equiv \mathcal{R}(\gamma, \alpha, \beta; c_t) + \mathcal{A}(\gamma, \alpha, \beta; c_t).$$

Hence, when $\alpha = 0$, $\mathcal{U}(\gamma, 0, \beta; c_t) = \mathcal{R}(\gamma, 0, \beta; c_t)$.

	$\beta = 0.975$			$\beta = 0.965$
γ	2.0	2.5	3.0	2.5
α	17.75	11.35	6.65	13.00
$\gamma^*(\gamma, \alpha, \beta)$	3.48793	3.51117	3.52169	3.76459

Table 10: **Uncertainty premia in the thought experiment:** We report the time-average of $\gamma^*(\gamma, \alpha, \beta; c_t)$ computed at each t on the sample path.

Finally, define $\gamma^*(\gamma, \alpha, \beta; c_t)$, to be the value of the relative risk aversion parameter which solves the following equation:

$$\mathcal{R}(\gamma^*, \alpha = 0, \beta; c_t) = \mathcal{U}(\gamma, \alpha, \beta; c_t) \Leftrightarrow c^*(\gamma^*, 0, \beta; c_t) = c^*(\gamma, \alpha, \beta; c_t). \quad (19)$$

On the left and right of the first equality in (19) we have, respectively, the uncertainty premium of an ambiguity neutral and the uncertainty premium of an ambiguity averse agent, both facing the same uncertain prospect as the agent in our model. Table 10 reports calculations with $\gamma = 2, 2.5, 3$ and α set to the corresponding calibrated values used in our model. Hence, our agent is calibrated to pay as much uncertainty premium (in *total*) as a standard expected utility agent with relative risk aversion around 3.5. Almost every equity premium study in the literature considers this amount of uncertainty premium very much within the range of plausibility in the context of a financial economy (Mehra and Prescott (1985), e.g., had argued on this basis that $\gamma \leq 10$ was plausible). In this sense, the calibrated uncertainty attitude parameters, *taken together*, make a plausible preference configuration for an individual DM in a financial economy.

The fact that our ambiguity averse agent is paying the same overall uncertainty premium as an expected utility maximizer with risk aversion 3.5 may seem odd given how large the average model implied equity premium we find. Note however, that the (conditional) equity premium is calculated, as is standard, by taking the expectation of risky rate (implied by the model) with respect to the posterior predictive distribution, as a Savage-Bayes ambiguity neutral outside observer would evaluate it. If we were to compute $Er - r_f$ as our ambiguity averse agent would, it would be much smaller given the pessimism of his evaluation functional. Thus the equity premium the agent perceives he is paying is consistent with the uncertainty premium values implied by the values of γ^* in the Table 10.

5 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 representative agent frameworks.

Bansal and Yaron (2004) 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 IES from risk aversion) could explain aspects of the observed equity premium. 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 be uncertain about the value of the persistence parameter; (3) letting the agent preferences treat (1) and (2) as ambiguity without separation of IES from risk aversion. We show these changes are sufficient to yield a model of beliefs where the (endogenously accentuated) uncertainty varies enough over time, without resorting to an exogenously specified stochastic volatility. Bansal and Yaron’s Case II model assumes an exogenous stochastic volatility. In our model notice the volatilities, $\sigma_{x_k}, \sigma_{d_k}, \sigma_{g_k}$, are different conditional on the value of persistence, ρ_k , but since there is “one true model” for all time, the volatility is non-stochastic, per se. Our agent is agnostic about the value of persistence and never puts probability one on either value, $k = h, l$. However, as has been explained in Section 3.2.2, depending on history, and because of ambiguity aversion, the agent amplifies the posterior probability mass on one or the other value, therefore creating an endogenously accentuated stochastic volatility, that is, the uncertainty about the value of σ ’s are accentuated endogenously.

In Hansen and Sargent (2010), countercyclical risk prices are driven by a representative investor’s robust model averaging *and* a preference for early resolution of

uncertainty. The investor carries along two difficult-to-distinguish models of consumption growth, one asserting i.i.d. log consumption growth, the other asserting that the growth in log consumption is a process with a slowly moving conditional mean. The investor uses observations on consumption growth to update a Bayesian prior over these two models, starting from an initial prior probability of .5. Each period, the agent expresses his specification distrust by pessimistically exponentially twisting a posterior over the two baseline models. That leads the investor to interpret good news as temporary and bad news as persistent, causing him to put countercyclical uncertainty components into the equilibrium price of risk.

Our framework is inspired by Hansen and Sargent (2010). Where we depart is the role of ambiguity in the driving mechanism and in the quantitative match obtained. Their agent believes the economy evolves according to a model like we have here but processes belief differently, by applying two “risk-sensitivity operators”. The first operator, which may be interpreted as 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 ρ . The other operator may be interpreted as a KMM2005 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 ρ . Hence, while uncertainty about the two values of ρ is treated as ambiguity, the uncertainty about the latent state, given ρ , is *not* processed as ambiguity, unlike in our model. Thus the results they obtain have their origin both in ambiguity aversion and an $\text{IES} > 1$.²⁸

Ju and Miao (2012) use a modified smooth ambiguity framework to assess the effect of ambiguity on dynamics of asset prices. In the model of beliefs there the latent state variable driving the (mean) growth rate in the economy may take only two possible values. The preference model also incorporates an IES effect, in addition to ambiguity aversion, with the IES parameter set at 1.5. They produce statistics on *unconditional* moments of returns and prices, by averaging across simulated, counter-factual paths, which match data 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 period the (second-order) belief has been almost completely stuck (virtually Dirac) on the same latent (high-growth) state. Hence, the results we obtain about predicted time series of moments of conditional returns (even counter-cyclical equity premium) could

²⁸We implemented, on our data set, an amended version their preference model with simply the second (KMM style) operator on the two point belief but excluding the other, Krep-Porteus style operator. We find the predicted time-averaged equity premium (conditional on actual history) is about 0.6% and that the conditional equity premium has a negative correlation with the JLN index.

not be obtained in their model if actual history were applied.²⁹

The part of Collin-Dufresne, Johannes, and Lochstoer (2016) most closely related to ours applies model/parameter uncertainty and Bayesian learning in a framework where the beliefs about the growth process is anchored to an uncertainty about whether the true process is LRR or i.i.d. They show that even a small probability of the LRR model being the true model leads to significant increase in the risk premium compared to the case in which consumption growth is known to be i.i.d. They also show that this uncertainty creates counter-cyclical fluctuations in the equity premium. However, as we underlined in the introduction, the driving force in the agent's preferences is an $IES > 1$ (they consider values 1.5 and 2, together with a relative risk aversion of 10). The mechanism at work is thus different from ours, as ambiguity aversion plays no role in their model. Drechsler (2013) introduces ambiguity aversion alongside model uncertainty and an $IES > 1$. He obtains good matches of time average returns moments. He uses a maxmin approach in which the set of priors, that represents uncertainty, varies over time in an *exogenous* way calibrated to an uncertainty index. Bidder and Dew-Becker (2016) is similar, in that they model ambiguity aversion using a worst case scenario, à la Gilboa and Schmeidler's maxmin expected utility model. The worst case model is literally the homoskedastic version of Bansal and Yaron's (2004) long-run risk model. They also apply an Epstein and Zin style IES effect, in addition to ambiguity aversion.

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. David and Veronesi (2013) studies time varying uncertainty but not the equity premium per se. In their model, agents must learn which regime the economy is in through signals about growth and inflation. The learning mechanism relies on (possibly small) money illusion. 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), Gjordani and Soderlind (2006), Jouini and Napp (2006), show that exogenously introducing

²⁹Recently, Strzalecki (2013) has shown 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 those include separate components explicitly adding preference for early resolution above and beyond what may be already mixed in with ambiguity aversion.

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 counter-cyclical of the equity premia, though, as we have observed, in these models consumption growth predicts the price-dividend ratio, unlike in the data and in our model. Habits effectively allow the risk aversion to vary endogenously over the business cycle. The crucial difference to our paper is that we have constant aversion (to ambiguity and risk) but our agent faces time-varying uncertainty and it is variation in that uncertainty, rather than variation in the aversion to it, which causes the returns and premia to vary.

6 Concluding remarks

Our model applied uncertainty and learning about persistent hidden states describing the cyclical component, and about the level of persistence; treating both these uncertainties as ambiguous and incorporating a level of ambiguity aversion calibrated to match the average risk-free rate. The uncertainty and learning compatible with a Bayesian agent (but not with rational expectations), 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 dynamics of all the variables, especially the first and second moments (conditional) equity premium through the channel of an endogenously accentuated “as if” uncertainty.

Our results show that observed levels and movements of moments of asset returns can be explained on the basis of aggregate macroeconomic risk, conditional on the *actual history* of aggregate output growth reports. That both first and second moments of conditional excess returns have the cyclical properties that match the data is a significant finding. As was the finding that the model implied conditional equity premium matches the time series properties of the JLN macroeconomic uncertainty index, thereby giving a theory of uncertainty shocks and the counter cyclical nature of their severity and persistence. Thus, consistent with JLN’s conjecture, we do find that Knightian uncertainty can provide a good explanation of dynamics of macroeco-

conomic uncertainty. Finally, it is worth appreciating the minimality of the departure from expected utility that was sufficient to capture so many aspects of returns data. These observations are very suggestive of the potential for this approach in domains of macro-finance research where effects of endogenously time-varying uncertainty are of interest.

In terms of future work, an interesting next step would be to replace the exchange economy with a production economy. Such a model would allow us to explore at least two important issues. First, it would allow us to understand the effect of ambiguity on output decisions (rather than just asset prices). Secondly, in turn, this would shed light on how uncertainty shocks, when they are endogenously accentuated by ambiguity aversion, contribute to the business cycle properties of the economy (see Backus, Ferriere, and Zin (2015) for a first discussion of these issues.)

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.
- ABEL, A. B. (1990): “Asset Prices under Habit Formation and Catching up with the Joneses,” *The American Economic Review*, 80(2), pp. 38–42.
- ALBERT, J., AND S. CHIB (1993): “Bayesian Analysis via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts,” *Journal of Business and Economic Statistics*, 11(1), 1–15.
- ANDERSEN, S., G. W. HARRISON, M. I. LAU, AND E. RUTSTROM (2008): “Eliciting Risk and Time Preferences,” *Econometrica*, 76(3), 583–619.
- AVRAMOV, D., L. BARRASZ, AND R. KOSOWSKI (2010): “Hedge Fund Predictability Under the Magnifying Glass: The Economic Value of Forecasting Individual Fund Returns,” Discussion paper, Imperial College Business School.
- BACKUS, D., A. FERRIERE, AND S. ZIN (2015): “Risk and ambiguity in models of business cycles,” *Journal Monetary Economics*, 69, 42–63.
- 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., D. KIKU, AND A. YARON (2012): “An Empirical Evaluation of the Long-Run Risks Model for Asset Prices,” *Critical Finance Review*, 1, 183–221.
- 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.
- (2009): “Rare disasters, asset prices, and welfare costs,” *The American Economic Review*, 99(1), 243–264.
- BEELER, J., AND J. CAMPBELL (2012): “The long-run risks model and aggregate asset prices: an empirical assessment,” *Critical Financial Review*, 1(1), 141–182.
- BIDDER, R., AND I. DEW-BECKER (2016): “Long-run risk is the Worst-Case Scenario,” *American Economic Review*, 106(9), 2494–2527.
- BLOOM, N. (2009): “The impact of uncertainty shocks,” *Econometrica*, 77(3), 623–685.

- BOYLE, P., L. GARLAPPI, R. UPPAL, AND T. WANG (2011): “Keynes Meets Markowitz: The Trade-off Between Familiarity and Diversification,” *Management Science*, 58(2), 253–272.
- BURNSIDE, C. (1998): “Solving Asset Pricing Model with Gaussian Shocks,” *Journal of Economic Dynamics and Control*, 22(3), 329–340.
- 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. (1999): “Ambiguity-Aversion and Non-Additive Probability: Experimental Evidence, Models and Applications,” in *Uncertain Decisions: Bridging Theory and Experiments*, ed. by L. Luini, pp. 53–80. Kluwer Academic Publishers.
- 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.
- CAMPBELL, J. Y. (2000): “Asset pricing at the millennium,” *The Journal of Finance*, 55(4), 1515–1567.
- CARRIERO, A., T. CLARK, AND M. MARCELLINO (2017): “Measuring uncertainty and its impact on the economy,” *Review of Economics and Statistics*, Forthcoming.
- 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 (2014): “Dynamic asset allocation with ambiguous return predictability,” *Review of Economic Dynamics*, 17(4), 799–823.
- CHETTY, R. (2006): “A New Method of Estimating Risk Aversion,” *American Economic Review*, 96(5), 1821–1834.
- CICCHETTI, C. J., AND J. DUBIN (1994): “A Microeconomic Analysis of Risk Aversion and the Decision to Self-Insure,” *J. of Political Economy*, 102(1), 169–186.
- COCHRANE, J. (1999): “New facts in finance,” *Economic Perspectives, Federal Reserve Bank of Chicago*, 23(3).

- (2008): “The dog that did not bark: A defense of return predictability,” *Review of Financial Studies*, 21(4), 1533–1575.
- COLLIN-DUFRESNE, P., M. JOHANNES, AND L. LOCHSTOER (2016): “Parameter learning in general equilibrium: Asset pricing implications,” *American Economic Review*, 106(3), 664–698.
- 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.
- DAVID, A., AND P. VERONESI (2013): “What ties return volatilities to price valuations and fundamentals?,” *Journal of Political Economy*, 121(4), 682–746.
- 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.
- DRECHSLER, I. (2013): “Uncertainty, Time-Varying Fear, and Asset Prices,” *The Journal of Finance*, 68(5), 1843–1889.
- 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., E. FARHI, AND T. STRZALECKI (2014): “How Much Would You Pay to Resolve Long-Run Risk?,” *American Economic Review*, 104(2), 2680–2697.
- 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.
- (1991): “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis,” *Journal of Political Economy*, 99(2), 263–86.

- FERNANDEZ-VILLAVERDE, J., AND J. RUBIO-RAMÍREZ (2010): “Macroeconomics and Volatility: Data, Models, and Estimation,” Working Paper 16618, University of Pennsylvania and NBER.
- GALLANT, A. R. (1997): *An Introduction to Econometric Theory: Measure-Theoretic Probability and Statistics with Applications to Economics*. Princeton University Press.
- GALLANT, A. R., M. JAHAN-PARVAR, AND H. LIU (2015): “Measuring Ambiguity Aversion,” Discussion paper, Department of Economics, Penn State University.
- GEWEKE, J. (2001): “A note on some limitations of CRRA utility,” *Economics Letters*, 71(3), 341–345.
- GILBOA, I., AND M. MARINACCI (2016): “Ambiguity and the Bayesian paradigm,” in *Readings in Formal Epistemology*, pp. 385–439. Springer.
- GILBOA, I., AND D. SCHMEIDLER (1989): “Maxmin expected utility with non-unique prior,” *Journal of Mathematical Economics*, 18(2), 141–153.
- GIORDANI, P., AND P. SODERLIND (2006): “Is there evidence of pessimism and doubt in subjective distributions? Implications for the equity premium puzzle,” *Journal of Economic Dynamics and Control*, 30(6), 1027–1043.
- GOLLIER, C. (2009): “Portfolio Choices and Asset Prices: The Comparative Statics of Ambiguity Aversion,” IDEI Working Papers 357, Institut d’Économie Industrielle (IDEI), Toulouse.
- GOLLIER, C. (2011): “Portfolio choices and asset prices: The comparative statics of ambiguity aversion,” *The Review of Economic Studies*, 78(4), 1329–1344.
- GOYAL, A., AND I. WELCH (2008): “A comprehensive look at the empirical performance of equity premium prediction,” *Review of Financial Studies*, 21(4), 1455–1508.
- GRIMMETT, G., AND D. STIRZAKER (2001): *Probability and Random Processes*. Oxford University Press.
- GUVENEN, F. (2009): “A Parsimonious Macroeconomic Model for Asset Pricing,” *Econometrica*, 77(6), 1711–1750.
- HALL, R. (1988): “Intertemporal substitution in consumption,” *The Journal of Political Economy*, 96(2), 339–357.
- HAMILTON, J. (1989): “A New Approach to Economic Analysis of Nonstationary Time Series,” *Econometrica*, 57(2), 357–384.
- 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.
- HARRISON, G., J. LIST, AND C. TOWE (2007): “Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion,” *Econometrica*, 75(2), 433–458.
- JOBERT, A., A. PLATANIA, AND L. ROGERS (2006): “A Bayesian solution to the equity premium puzzle,” Discussion paper, Statistical Laboratory, University of Cambridge.
- 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.
- JURADO, K., S. C. LUDVIGSON, AND S. NG (2015): “Measuring uncertainty,” *American Economic Review*, 105(3), 1177–1216.
- KANDEL, S., AND R. F. STAMBAUGH (1990): “Expectations and volatility of consumption and asset returns,” *Review of Financial Studies*, 3(2), 207–232.
- 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.
- KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (2011): “On the Smooth Ambiguity Model: A Reply,” Discussion paper, Department of Economics, University of Oxford, <http://www.economics.ox.ac.uk/research/WP/pdf/paper524.pdf>.
- KLIBANOFF, P., M. MARINACCI, AND S. MUKERJI (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.
- KOCHERLAKOTA, N. (1996): “The equity premium: It’s still a puzzle,” *Journal of Economic Literature*, 34(1), 42–71.
- KOIJEN, R., AND S. VAN NIEUWERBURGH (2011): “Predictability of returns and cash flows,” *Annual Review of Financial Economics*, 3, 467–491.

- 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.
- LETTAU, M., AND S. C. LUDVIGSON (2010): *Handbook of Financial Econometrics* chap. Measuring and Model Variation in the Risk-Return Trade-off, pp. 617–690. Elsevier. B. V.
- LJUNGQVIST, L., AND T. SARGENT (2004): *Recursive macroeconomic theory*. The MIT Press.
- LUCAS, R. J. (1978): “Asset Prices in an Exchange Economy,” *Econometrica*, 46(6), 1429–1445.
- LUDVIGSON, S. C. (2012): *Handbook of the Economics of Finance* chap. Advances in consumption-based asset pricing: Empirical tests, pp. 799–906. Elsevier. B. V.
- MACCHERONI, F., M. MARINACCI, AND D. RUFFINO (2013): “Alpha as Ambiguity: Robust Mean-Variance Portfolio Analysis,” *Econometrica*, 81, 1075–1113.
- MACCHERONI, F., M. MARINACCI, AND A. RUSTICHINI (2006): “Ambiguity Aversion, Robustness, and the Variational Representation of Preferences,” *Econometrica*, 74(6), 1447–1498.
- MEHRA, R., AND E. PRESCOTT (1985): “The Equity Premium: A Puzzle,” *Journal of Monetary Economics*, 15(2), 145–161.
- MUKERJI, S., AND J. M. TALLON (2001): “Ambiguity aversion and incompleteness of financial markets,” *Review of Economic Studies*, 68(4), 883–904.
- MUKERJI, S., AND J.-M. TALLON (2004): “An overview of economic applications of David Schmeidler’s models of decision making under uncertainty,” in *Uncertainty in economic theory: Essays in honor of David Schmeidler’s 65th birthday*, ed. by I. Gilboa. Routledge.
- ORLIK, A., AND L. VELDKAMP (2014): “Understanding uncertainty shocks and the Role of the Black Swan,” Working paper 20445, NBER working paper, <http://www.nber.org/papers/w20445>.
- POHL, W., K. SCHMEDDERS, AND O. WILMS (2015): “Higher-Order Effects in Asset-Pricing Models with Long-Run Risks,” Discussion Paper 14-68, Swiss Finance Institute Research Paper.
- 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.

- ROUWENHORST, K. (1995): “Asset pricing implications of equilibrium business cycle models,” in *Frontiers of Business Cycle Research*, ed. by Cooley, and Prescott, pp. 294–330. Princeton University Press.
- 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. (2013): “Temporal resolution of uncertainty and recursive models of ambiguity aversion,” *Econometrica*, 81(3), 1039–1074.
- 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.
- (1990): “Nonexpected Utility in Macroeconomics,” *The Quarterly Journal of Economics*, 105(1), 29–42.
- WEITZMAN, M. (2007): “Subjective expectations and asset-return puzzles,” *The American Economic Review*, 97(4), 1102–1130.
- WHITELAW, R. F. (1994): “Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns,” *Journal of Finance*, 49(2), 515–541.
- WHITELAW, R. F. (2000): “Stock market risk and return: An equilibrium approach,” *Review of Financial Studies*, 13(3), 521–547.

A Data and estimation of parameters of the stochastic models

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 α 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 α . The relative risk aversion parameter γ was allowed to range between 1 (log utility) and 3, regarded as plausible in macroeconomic models (Ljungqvist and Sargent, 2004, pg. 426); the “baseline” calibration set $\gamma = 2.5$.³⁰ The discount factor β 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 $\psi = 2.5$.

The long-run risk model was fit to annual data using maximum likelihood. Parameter estimates are shown in Table 11. All parameters, except ρ and ψ were estimated using data 1930–1977. The mean of consumption and dividends, \bar{g} and \bar{d} , respectively were set to their values in the period 1930 – 1977. The variances of the latent state process, consumption growth and dividend growth were estimated using the Kalman Filter. The dividend leverage parameter, ψ , 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 ($\psi = 2.5$).

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)
σ_g^2	0.048 (0.016)	0.046 (0.016)	0.025 (0.010)	0.020 (0.007)	0.047 (0.017)	0.026 (0.008)
σ_d^2	4.49 (0.893)	4.51 (0.892)	4.75 (0.909)	4.73 (0.902)	4.64 (0.914)	4.81 (0.918)
σ_x^2	0.054 (0.013)	0.054 (0.013)	0.051 (0.019)	0.059 (0.021)	0.054 (0.013)	0.050 (0.021)

Table 11: **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 variance estimates and their standard errors have been multiplied by 100.

B Details of the model

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 (ρ_l) or a high persistence (ρ_h), but does not know for sure which. That is, he believes either of the models described in equation (9) represent the true data generating process.

³⁰If the two smooth ambiguity preferences do not share the same risk attitude it is not necessarily true that a more concave ϕ means more ambiguity aversion. Hence α is meaningfully calibrated *given* a value of γ ; not independent of γ .

Define $\hat{x}_{k,t} \equiv E[x_{k,t} | g_{k,1}, \dots, g_{k,t}, d_{k,1}, \dots, d_{k,t}]$, $k = l, h$, 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 $\rho = \rho_k$ 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 $\hat{x}_{k,t+1}^{(i)}$, $i = l, h$, $k = l, h$, the agent's forecast for the (one period ahead) update to his 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 $\rho = \rho_k$ as the data generating process, when the data is actually generated by the i persistence model. The direct value function obtains as follows:

$$\begin{aligned}
V(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) &= (1 - \beta) \frac{C_t^{1-\gamma}}{1 - \gamma} & (20) \\
&- \frac{\beta}{\alpha} \ln \left[\eta_t \left\{ \int_{-\infty}^{\infty} \exp \left(-\alpha \iiint_{-\infty}^{\infty} V \left(C_t \exp(g_{l,t+1}), \hat{x}_{l,t+1}^{(l)}(\vec{\varepsilon}_{l,t+1}), \right. \right. \right. \right. \\
&\quad \left. \left. \left. \hat{x}_{h,t+1}^{(l)}(\vec{\varepsilon}_{l,t+1}), \eta_{t+1}^{(l)}(\vec{\varepsilon}_{l,t+1}) \right) dF(\vec{\varepsilon}_{l,t+1}) dF(x_{l,t}) \right\} \right. \\
&+ (1 - \eta_t) \left\{ \int_{-\infty}^{\infty} \exp \left(-\alpha \iiint_{-\infty}^{\infty} V \left(C_t \exp(g_{h,t+1}), \hat{x}_{l,t+1}^{(h)}(\vec{\varepsilon}_{h,t+1}), \right. \right. \right. \\
&\quad \left. \left. \left. \hat{x}_{h,t+1}^{(h)}(\vec{\varepsilon}_{h,t+1}), \eta_{t+1}^{(h)}(\vec{\varepsilon}_{h,t+1}) \right) dF(\vec{\varepsilon}_{h,t+1}) dF(x_{h,t}) \right\} \right]
\end{aligned}$$

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

$$\hat{x}_{l,t+1}^{(l)}(\vec{\varepsilon}_{l,t+1}) = \rho_l \hat{x}_{l,t} + K_l \nu_{l,t+1}^{(l)} \quad (21)$$

$$\hat{x}_{h,t+1}^{(l)}(\vec{\varepsilon}_{l,t+1}) = \rho_h \hat{x}_{h,t} + K_h \nu_{h,t+1}^{(l)} \quad (22)$$

$$\hat{x}_{l,t+1}^{(h)}(\vec{\varepsilon}_{h,t+1}) = \rho_l \hat{x}_{l,t} + K_l \nu_{l,t+1}^{(h)} \quad (23)$$

$$\hat{x}_{h,t+1}^{(h)}(\vec{\varepsilon}_{h,t+1}) = \rho_h \hat{x}_{h,t} + K_h \nu_{h,t+1}^{(h)} \quad (24)$$

where $\nu_{k,t+1}^{(i)}$, $(i) = (l)$ or $(i) = (h)$ and $k = l, h$, denote the ‘‘surprises’’. For example, when the DGP is $(i) = (l)$ and the filter uses ρ_k , $k = h$, the surprise is defined

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

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

$$K_k = \rho_k \Omega_k [1 \ \psi] \hat{F}_k^{-1}, \quad \text{where} \quad \hat{F}_k = \begin{bmatrix} \Omega_k + \sigma_{g_k}^2 & \psi \Omega_k \\ \psi \Omega_k & \psi \Omega_k + \sigma_{d_k}^2 \end{bmatrix}$$

Finally, Ω_k , $k = l, h$, is defined as the solution to

$$\Omega_k = \rho_k^2 \Omega_k - \rho_k^2 \Omega_k^2 [1 \ \psi] \hat{F}_k^{-1} [1 \ \psi]' + \sigma_{x_k}^2$$

The Bayes update of η_t is obtained as follows :

$$\eta_{t+1}^{(l)}(\vec{\varepsilon}_{l,t+1}) = \frac{\eta_t L(\nu_{l,t+1}^{(l)}, \hat{F}_l)}{\eta_t L(\nu_{l,t+1}^{(l)}, \hat{F}_l) + (1 - \eta_t) L(\nu_{h,t+1}^{(l)}, \hat{F}_h)} \quad (25)$$

$$\eta_{t+1}^{(h)}(\vec{\varepsilon}_{h,t+1}) = \frac{\eta_t L(\nu_{l,t+1}^{(h)}, \hat{F}_l)}{\eta_t L(\nu_{l,t+1}^{(h)}, \hat{F}_l) + (1 - \eta_t) L(\nu_{h,t+1}^{(h)}, \hat{F}_h)} \quad (26)$$

where the likelihood is

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

B.1.1 Mean and variance of the distorted posterior

The mean of the distorted (or, ‘‘as if’’) posterior is given by:

$$\tilde{x}_t = \eta_t \int_{-\infty}^{\infty} (x_{l,t}) \xi_t^{(l)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) dF(x_{l,t}) + (1 - \eta_t) \int_{-\infty}^{\infty} (x_{h,t}) \xi_t^{(h)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) dF(x_{h,t}) \quad (27)$$

and the variance, by:

$$\begin{aligned} \widetilde{Var}_t(x_t) \equiv & \eta_t \int_{-\infty}^{\infty} (x_{l,t}^2) \xi_t^{(l)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) dF(x_{l,t}) \\ & + (1 - \eta_t) \int_{-\infty}^{\infty} (x_{h,t}^2) \xi_t^{(h)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) dF(x_{h,t}) - \tilde{x}_t^2 \end{aligned}$$

B.1.2 The rates of return

The risky rate of return is a function of four state variables, $C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t$, just like V and ξ_t . In the sequel, it should be clear that variables in $t + 1$ are evaluated using the relevant stochastic components. Let $C_{k,t+1} = C_t \exp(g_{k,t+1})$, $k = l, h$. The risk rate,

R_t , will satisfy:

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

where,

$$\xi_t^{(l)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) = \frac{\phi' \left(\iiint_{-\infty}^{\infty} V(C_{l,t+1}, \hat{x}_{l,t+1}^{(l)}, \hat{x}_{h,t+1}^{(l)}, \eta_{t+1}^{(l)}) dF(\vec{\varepsilon}_{l,t+1}) \right)}{\Psi} \quad (28)$$

and

$$\xi_t^{(h)}(C_t, \hat{x}_{l,t}, \hat{x}_{h,t}, \eta_t) = \frac{\phi' \left(\iiint_{-\infty}^{\infty} V(C_{h,t+1}, \hat{x}_{l,t+1}^{(h)}, \hat{x}_{h,t+1}^{(h)}, \eta_{t+1}^{(h)}) dF(\vec{\varepsilon}_{h,t+1}) \right)}{\Psi} \quad (29)$$

with

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

Then, we have

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

and the risk-free rate is

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

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

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

C An analytical approximation for rates of return in the case of known persistence model

This section develops an analytical approximation to the equilibrium rates of return in the model with known persistence. The crucial assumption on which the following second order approximation analysis depends is that $E_{\tilde{\rho}_t}$ operates with respect to some normal distribution $N(\tilde{x}_t, \tilde{\Omega})$. As the numbers (reporting skewness and excess kurtosis) in Table 12 generated using the accurate numerical approximation demonstrate, Normality is a fairly accurate description.

Model with known persistence ($\rho = .85$)				
	x_t		$g_{c,t}$	
	E	σ	E	σ
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	κ	sk	κ
Rat. Exp.	–	–	0.000	0.000
Bayesian	0.000	-0.000	0.000	-0.000
Twisted	0.000	-0.000	0.000	0.000

Model with unknown persistence ($\rho = .85$)				
	x_t		$g_{c,t}$	
	E	σ	E	σ
Bayesian	-0.001	0.024	0.019	0.034
Twisted	-0.022	0.028	-0.002	0.037
	sk	κ	sk	κ
Bayesian	-0.003	0.013	-0.003	0.017
Twisted	-0.005	-0.053	-0.038	-0.029

Table 12: **Conditional moments of distributions.** In each case, $\gamma = 2.5$ and α was set such that the model generates an average risk-free rate of 1.5%. C_t , $\hat{x}_{l,t}$, $\hat{x}_{h,t}$ and η_t are set equal to their mean in the data. sk and κ denote skewness and excess kurtosis (relative to a Gaussian distribution), respectively. The latent state variable is known to a rational expectations agent and so the conditional distribution is degenerate.

1 (Approximating assumption 1) $\tilde{\mu}_t = N(\tilde{x}_t, \Omega)$.

Recall that $\tilde{\mu}_t \equiv \xi_t(x_t) \otimes N(\hat{x}_t, \Omega)$ and thus has density given by

$$\tilde{f}(x_t) = \xi_t(x_t | C_t, \hat{x}_t; \alpha) \frac{1}{\sqrt{2\pi\Omega}} \exp\left(-\frac{(x_t - \hat{x}_t)^2}{2\Omega}\right). \quad (30)$$

This assumption is thus equivalent to assuming that eq. (30) is exactly a normal density with the same variance as the Bayesian posterior Ω but with a different mean (\tilde{x}_t instead of \hat{x}_t). Let $E_t \equiv E_{\hat{x}_t} E_{x_t}$; $\tilde{E}_t \equiv E_{\tilde{\mu}_t} E_{x_t} \equiv E_{\tilde{x}_t} E_{x_t}$. It is useful to recall, if x_t is normally distributed, then for any $k \in \mathbb{R}$,

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

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 ε terms have expectation zero under both \tilde{E}_t and E_t since the terms have expectation zero conditional on x_t .

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

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

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

$$r_t^f = -\ln\beta + \gamma\bar{g} + \gamma\rho\tilde{x}_t - \frac{\gamma^2}{2}(\sigma_x^2 + \sigma_g^2 + \rho^2\widetilde{\text{Var}}_t(x_t)). \quad (31)$$

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

$$\tilde{E}_t \exp\left[\ln\beta + \ln\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) - \gamma \ln\left(\frac{C_{t+1}}{C_t}\right)\right] = 1 \quad (32)$$

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

2 (Approximating assumption 2) :

$$r_t \equiv \ln\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) \simeq \kappa_0 + \kappa_1 z_{t+1} - z_t + d_{t+1} \quad (33)$$

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

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

$$z_t = A_0 + A_1 \tilde{x}_t. \quad (34)$$

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 10.

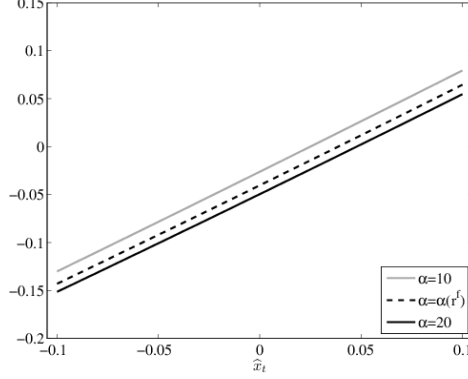


Figure 10: $\tilde{x}_t = E_{\tilde{\mu}_t}(x_t)$ plotted against \hat{x}_t . The level of consumption is set to the average value between 1978 and 2011. In each case, $\gamma = 2.50$.

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

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

$$1 = \tilde{E}_t \exp \left[\ln(\beta) + \bar{d} - \gamma \bar{g} + \kappa_0 + (\kappa_1 - 1)A_0 + \kappa_1 A_1 (\delta_0 + \delta_1 \hat{x}_{t+1}) - A_1 \tilde{x}_t + (\psi - \gamma)\rho x_t + (\psi - \gamma)\sigma_x \varepsilon_{x,t+1} + \sigma_d \varepsilon_{d,t+1} - \gamma \sigma_g \varepsilon_{g,t+1} \right]. \quad (35)$$

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. (35) and using $\hat{x}_t = \frac{\tilde{x}_t - \delta_0}{\delta_1}$. Hence,

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

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,

$$A_1 = \frac{\rho(\psi - \gamma)}{1 - \rho \kappa_1 (1 - (1 - \delta_1)(K_g + \psi K_d))} \quad (36)$$

Doing the same for the constant terms, we have

$$\begin{aligned} (1 - \kappa_1) A_0 = & \ln(\beta) + \bar{d} - \gamma \bar{g} + \kappa_0 + \kappa_1 A_1 \delta_0 - \delta_0 (\kappa_1 A_1 \rho (1 - K_g - \psi K_d)) \\ & + \rho^2 (\kappa_1 A_1 \delta_1 (K_g + \psi K_d) + \psi - \gamma)^2 \widetilde{Var}_t(x_t) / 2 \\ & + (\psi - \gamma + \kappa_1 A_1 \delta_1 (K_g + \psi K_d))^2 \sigma_x^2 / 2 \\ & + (\kappa_1 A_1 \delta_1 K_d + 1)^2 \sigma_d^2 / 2 + (\kappa_1 A_1 \delta_1 K_g - \gamma)^2 \sigma_g^2 / 2 \end{aligned} \quad (37)$$

Using eq. (34) 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

$$E_t r_t = \kappa_0 + A_0 (\kappa_1 - 1) + \kappa_1 A_1 \delta_0 (1 - \rho) + \bar{d} + A_1 (\kappa_1 \rho - 1) \tilde{x}_t + \psi \rho \hat{x}_t \quad (38)$$

and so the Equity premium is then

$$\begin{aligned} E_t r_t - r_t^f = & \kappa_0 + A_0 (\kappa_1 - 1) + \kappa_1 A_1 \delta_0 (1 - \rho) + \bar{d} + A_1 (\kappa_1 \rho - 1) \tilde{x}_t + \psi \rho \hat{x}_t \\ & + \ln(\beta) - \gamma \bar{g} - \gamma \rho \tilde{x}_t + \frac{\gamma^2}{2} (\sigma_x^2 + \sigma_g^2 + \rho^2 \widetilde{Var}_t(x_t)) \end{aligned} \quad (39)$$

Note that when $\delta_1 = 1$, as is true in our data (see Figure 10), A_1 simplifies to $-\rho(\psi - \gamma) / (\kappa_1 \rho - 1)$.

We need values of the approximating constants, κ_0 and κ_1 , to compute the log price-dividend ratio. Beeler and Campbell (2012) obtain the constants as follows

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

D Ambiguity of second-order beliefs

Let T be a second-order event, i.e., $T \subset \Theta$, with $\mu(T) = m$. Consider two prospects. One, a bet on this event, which pays x on the event and y off it, with $x > y$. Two, a lottery, ℓ_m which pays x with probability m and y with probability $1 - m$. Notice, when ϕ is concave, by Jensen's inequality,

$$m\phi(u(x)) + (1 - m)\phi(u(y)) < \phi(m(u(x)) + (1 - m)(u(y))) \quad (40)$$

The LHS of (40) is the evaluation of the bet on T while the RHS is the evaluation of the lottery, per the smooth ambiguity model. Similarly, the bet on the complementary event T^c is dispreferred to ℓ_{1-m} given a concave ϕ . Indeed, ambiguity aversion implies we cannot find a calibrated lottery event such that betting on that lottery event is same as betting on T ; there is no lottery probability that is same as μ . Hence, when ϕ is concave, the second-order measure μ cannot be calibrated with a lottery; behaviorally, μ is not treated as an objective probability.

As shown formally in Section 2.4 in Klibanoff, Marinacci, and Mukerji (2012), this is the heart of the argument that establishes that ambiguity of a first-order event E implies that non-null and non-universal second-order events concerning the probability of E are treated as ambiguous. Hence, the smooth ambiguity model property of expected utility evaluation of second-order acts (e.g., bets on events in Θ) does not mean that the DM treats these acts as based on unambiguous events.

Ambiguity and the historical equity premium

Online Appendix

Collard

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

Sujoy Mukerji¹

School of Economics and Finance
Queen Mary University of London, London E1 4NS

Kevin Sheppard

Department of Economics
University of Oxford, Oxford OX1 3UQ

Jean-Marc Tallon

Paris School of Economics and CNRS
48, boulevard Jourdan, 75014 Paris

January 1, 2018

¹Contact author: s.mukerji@qmul.ac.uk

1 Details of the numerical solution procedure

1.1 Solution Method

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

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

where $X_t \equiv (C_t, \widehat{x}_{h,t}, \widehat{x}_{\ell,t}, \eta_t)$ denotes the vector of state variables¹ 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.² $H_\iota(\cdot)$ is a Hermite polynomial of order ι 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, $X_t \equiv (C_t, \widehat{x}_{h,t}, \widehat{x}_{\ell,t}, \eta_t)$.³ The parameters θ^y , $y \in \{V, R\}$, are then determined by a minimum weighted residuals method. More precisely, we define the residuals associated to both the direct Value function equation, $\mathcal{R}_V(\theta^V; X_t)$, and the Euler equations for risky assets (consumption claims and dividend claims), $\mathcal{R}_R(\theta^V; X_t)$, as

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

where

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

and

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

¹When persistence is known, the vector of state variables reduces to $X_t = (C_t, x_t)$ and the approximant takes the simpler form $\Phi_y(X_t) = \exp \left(\sum_{i_c, i_x \in \mathcal{I}} \theta_{i_c, i_x}^y H_{i_c}(\varphi_c(C_t)) H_{i_x}(\varphi_x(\widehat{x}_t)) \right)$.

²See ?, Chapter 7.

³We use this function in order to be able to narrow down the range of values taken by the state variables, such that the approximation performs better when evaluated on the data.

where

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

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

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, $C_{t+1}^{(\nu)}$, $\widehat{x}_{h,t+1}^{(\nu)}$, $\widehat{x}_{\ell,t+1}^{(\nu)}$, $\eta_{t+1}^{(\nu)}$, $\nu \in \{h, \ell\}$, are obtained using the dynamic equations described in Subsection ???. These expressions are simplified when the agent is certain about the persistence. This case amounts to setting $\eta_t = 0$ for all t in the preceding expressions and consider only one process for \widehat{x}_t .

The vector of parameters θ^V and θ^R are then determined by projecting the residuals on Hermite polynomials. This then defines a system of orthogonality conditions which is solved for θ^V and θ^R . More precisely, we solve⁴

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

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

$$\Omega(X_t) \equiv \omega(\varphi_h(C_t)) \omega(\varphi_h(\widehat{x}_t^h)) \omega(\varphi_\ell(\widehat{x}_t^\ell)) \omega(\varphi_\eta(\eta_t))$$

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

⁴It should be clear to the reader that the integral refers to a multidimensional integration problem, as we integrate over C , x^h , x^ℓ and η .

direct value function in equilibrium, the system can be solved recursively. We therefore first solve the value function approximation problem, and use the result vector of parameters θ^V to solve for the risky rate problem.

Integrals are approximated using a monomial approach whenever we face a multidimensional integration problem (inner integrals in the computation of expectations and projections) and a Gauss Hermitian quadrature approach when dealing with unidimensional integrals (outer integrals in the computation of expectations).⁵

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 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 $\rho_h = 0.85$,
- $(n_c, n_{x_h}, n_{x_\ell}, n_\eta) = (4, 2, 2, 2)$ 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 κ_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.⁶ Finally, the stopping criterion is set to $1e-6$.

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

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

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

⁵See ?, chapter 7.

⁶More precisely, we approximate

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

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 ? for greater details.

- 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

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

- If the orthogonality conditions are satisfied, in the sense the residuals are lower than the stopping criterion ε , then the vector of parameters are given by θ^V and θ^R . Else update θ^V and θ^R using a Gauss Newton algorithm and go back to step 1.

1.2 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^{th} order moment of the risky rate in period t as

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

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

$$ER^n = \frac{1}{t_2 - t_1} \left[\sum_{t=t_1}^{t=t_2} E_t^n R_{t+1} - (E_{t_1}^1 R_{t_1+1})^n \right] \quad (2)$$

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 moments of the equity premium can be computed as in eq. (??).