Fragile beliefs and the price of model uncertainty

Lars Peter Hansen[*]
Thomas J. Sargent[‡]

March 8, 2007

Abstract

An enduring model selection problem in which one of the models has long run risks gives rise to countercyclical risk premia. We use two risk-sensitivity operators to construct the stochastic discount factor for a representative consumer who evaluates consumption streams in light of model selection and parameter estimation problems that over time can aggravate or attenuate long run risks. The arrival of signals induces the consumer to alter his posterior distribution over models and parameters. The consumer copes with doubts about probabilities by slanting them in directions that have pessimistic consequences for value functions. His twisted model probabilities give rise to model uncertainty premia that contribute a time-varying component to what is ordinarily measured as the market price of risk.

KEY WORDS: Learning, Bayes' law, robustness, risk-sensitivity, market price of model uncertainty, time-varying risk premia.

1 Introduction

1.1 Fragility

We make the beliefs of a representative consumer fragile by modifying the usual rational expectations assumption that endows him with a unique probability measure over sequences of payoff relevant outcomes. To represent doubts about his approximating model, we endow the representative consumer with a *set* of probability measures near his approximating model, a set whose members are difficult to distinguish statistically. We assume that he is unwilling or unable to reduce that set to a singleton by putting a prior over it.

Our representative consumer values consumption streams as if he were an expected utility maxi-minimizer á la Gilboa and Schmeidler (1989) and thereby computes valuations that are cautious with respect to the set of measures. Caution inspires him to seek *bounds* on his valuations that are satisfied for every member of a set of distributions. He constructs those bounds by minimizing his value function over the set of probability measures.

At time t, our representative consumer evaluates payoffs according to a particular measure that emerges from a time t minimization problem over a set of continuation measures. *Ex post*, he acts 'as if' he is an ordinary rational expectations investor, except that the probability measure that he uses is twisted in a pessimistic direction relative to what it would be in a comparable rational expectations model. The 'fragility' in our title refers to movements over time in the worst-case continuation measures that underlie his valuations.[1] Relative to the conventional rational expectations case in which the representative consumer has complete confidence in his statistical model, our representative consumer's inability to specify a unique model puts 'model uncertainty premia' into asset prices.

1.2 Fragile expectations as sources of time-varying risk premia

To make these ideas operational, we endow our representative consumer with a hidden Markov model that confronts him with ongoing model selection and parameter estimation problems. Time-varying risk premia emerge as model uncertainty adjustments to the probability distribution over hidden states. Our representative consumer wants to know components of a hidden state vector, some of which stand for unknown parameters within a submodel and others of which index submodels. A distribution over that hidden state becomes part of the state vector in the representative consumer's value function. Bayes' law describes its motion over time. The submodels are difficult to distinguish with limited amounts of historical evidence. To make forecasts of consumption growth that are robust with respect both to his model specification and to his prior over submodels, the representative consumer slants probabilities towards the submodel that has the lowest utility, subject to some constraints on the relative entropy between the approximating model and the resulting slanted model. We show how variations over time in the probabilities attached to the submodels and other state variables put volatility into the model uncertainty premia, and how they are enhanced by the representative consumer's adjustments for model misspecification.

Our application builds on ideas of Tallarini (2000), Bansal and Yaron (2004), and Hansen and Sargent (2006a). We follow Tallarini by adopting an easy to compute logarithmic preference specification and an endowment economy.[2] But we specify a different consumption growth process than Tallarini. Our representative consumer's model of consumption growth consists of a probability weighted mixture of two competing submodels, each of which is a hidden Markov model. And while Tallarini applied one risk sensitivity operator, we apply two. Our second risk-sensitivity operator, taken from Hansen and Sargent (2006a), adjusts for uncertainty about hidden Markov states that include both unknown parameters and indexes of submodels. We interpret both risk-sensitivity operators as capturing the

---

[1]Harrsion and Kreps (1978) and Scheinkman and Xiong (2003) explore a very different setting in which difficult to detect departures from rational expectations lead to interesting asset price dynamics that cannot occur under rational expectations.

[2]Tallarini studies both a production economy and a pure endowment economy.

representative consumer's concerns about robustness instead of the enhanced risk aversion featured by Tallarini.[3]

Along with Bansal and Yaron (2004), we impute to the representative consumer a model for consumption growth that contains a difficult to detect persistent component. We differ from Bansal and Yaron by endowing the representative consumer with multiple doubts about the model specification. Our representative consumer assigns positive probabilities to two submodels that Bansal and Yaron tell us are difficult to distinguish, namely, an i.i.d. consumption growth model and Bansal and Yaron's long-run risk model. But he doubts those model-mixing probabilities as well as the specification of each of those submodels. In contrast, Bansal and Yaron assume that the representative consumer assigns probability one to the long-run risk model even though sample evidence is indecisive in selecting between them.[4] Nevertheless, as we shall see in section 5.3, our framework gives a new possible justification for why a consumer might want to act *as if* he puts probability (close to) one on the long-run risk model even though he knows that it is difficult to discriminate between these models from finite data records.

Our logarithmic utility specification makes the representative consumer's value function linear, conditional on a component of the hidden state that indexes submodels. Value functions conditional on the submodels can be computed via special cases of the linear-quadratic calculations in Hansen and Sargent (2006b).

1.3 Relation to other asset pricing models with Bayesian learning

Bayesian learning about hidden states is part of other asset pricing models. But we exploit substantially more of the hidden state structure than do earlier researchers, for example, Detemple (1986), David (1997), Veronesi (2000), Brennan and Zia (2001), Ai (2006), and Croce et al. (2006).[5] Those papers use learning about a hidden state simply to generate a particular exogenous process for distributions of future signals conditional on past signals, an essential input into any consumption based asset pricing model. Those papers specify the evolution of a primitive stochastic process of technology or endowments in terms of an information structure that conceals hidden Markov states. By applying Bayesian learning, typically as embodied in recursive filtering methods, they construct a less informative state vector that consists of sufficient statistics for the distribution of hidden states conditioned on signal histories, as well as a recursive law of motion for it. The evolution equation for that less informative state vector is used as an input in decision making and asset pricing. Except that they are based on coarser information structures, decision making and asset pricing in these models is standard. Thus, those applications of Bayesian learning to asset pricing reduce the representative consumer's information set without any reference to actions

---

[3]Barillas et al. (2006) reinterpret some of Tallarini's results in terms of concern about model misspecification instead of risk aversion.

[4]Bansal and Yaron (2004) incorporate other features in their specifications of consumption dynamics, including stochastic volatility. They also use a recursive utility specification with an intertemporal elasticity of substitution greater than 1.

[5]The learning problems in those papers share the feature that learning is passive, there being no role for experimentation so that prediction can be separated from control. Cogley et al. (2005) apply the framework of Hansen and Sargent (2006a) in a setting where decisions affect future probabilities of hidden states and experimentation is active.

or utility consequences.[6] Therefore, the asset pricing implications of such learning models depend only on the distributions of future signals conditioned on past signals, and not on the underlying structure with hidden states that the model builder used to deduce that distribution. In such rational expectations models, the only thing that learning contributes is a justification for those conditional distributions: we would get equivalent asset pricing implications if we had started by just *assuming* those distributions.

As we shall see, the way that we activate our second risk-sensitivity operator means that equivalence is not true in our model, because for us, asset prices continue also to depend on the evolution of the hidden states and not just on the distribution of future signals conditioned on signal histories under the consumer's approximating model. This occurs because of how, following Hansen and Sargent (2006a), we make the representative consumer explore potential misspecifications of the distributions of hidden Markov states and of future signals conditioned on those hidden Markov states and on how he therefore refuses to reduce compound lotteries. In a sequel to this paper, Hansen (2007), among other things, (a) expands the model uncertainty faced by the representative consumer by effectively confronting him with a bigger class of models parameterized by introducing multiple parameter configurations into Bansal and Yaron's model, and (b) studies the consequences of applying a risk-sensitivity adjustment to this additional source of model uncertainty.

1.4 Outline

Let $EV(x)$ be a scalar value function, where E is the mathematical expectation over a random vector x. Rather than applying E to $V(x)$, this paper follows Hansen and Sargent (1995), Hansen et al. (1999), Tallarini (2000) and others by applying a *risk-sensitivity* operator to V, namely, $\mathsf{T}[V(x)] = -\theta \log E \exp\left[\frac{-V(x)}{\theta}\right]$, where $\theta > 0$ is a parameter that controls an extra adjustment for risk that is beyond what is captured by the shape of V. When $\theta = +\infty$, $\mathsf{T} = E$.

Tallarini (2000) interprets θ as a risk-aversion parameter and T as an extra adjustment for risk in continuation values. We instead interpret $-\theta \log E \exp\left[\frac{-V(x)}{\theta}\right]$ as the indirect value function for a max-min expected utility decision maker who makes conservative evaluations of probability densities and therefore regard θ as a parameter that measures concern about misspecification of probabilities. We use T as a device from which we can deduce how the market prices a representative consumer's uncertainty about model specification.

In a representative consumer asset pricing model, replacing E with T adds a potentially volatile multiplicative adjustment to the stochastic discount factor. Hansen et al. (1999) and Tallarini (2000) show how this adjustment increases the theoretical value of the market price of risk and thereby helps move the theoretical stochastic discount factor closer to the bounds of Hansen and Jagannathan (1991). In this paper, we consider three models with successively deeper specification doubts. We study how these doubts affect the market price of model uncertainty. We summarize these effects in table 1 and the last column of table 2. Where $A(\iota), C(\iota), D(\iota), G(\iota)$ are matrices in a state space model ι for the consumption growth rate, table 1 displays formulas for objects that compose the market price of model

---

[6]One exception to this statement is our earlier paper Cagetti et al. (2002) that incorporates a robustness correction in a continuous time model.

uncertainty that are reported in table 2. The first row of table 2 shows a time invariant market price of model uncertainty that comes from the model of Tallarini (2000) after we have adapted it by incorporating the long run risk specification of Bansal and Yaron (2004). The second row reports the market price of model uncertainty for a non-stationary version of a model in the spirit of Hansen et al. (2002) that we form by restricting the representative consumer's information set to be the history of consumption growth rates. This model puts a smoothly time-varying component into the market price of model uncertainty. The third row pertains to the model that is the principal focus of this paper. Here the representative consumer faces an enduring and ambiguous model selection problem that contributes an erratic time varying component of the market price of risk that imparts countercyclical risk premia.[7]

1.5   Timing of resolution of uncertainty

We will use two T operators. The first operator T^1 conditions on knowledge of all hidden state variables. Hansen and Sargent (2006a) pointed out that T^1 has an alternative interpretation of relaxing the hypothesis that intertemporal compound lotteries can be reduced. As Kreps and Porteus (1978) and Epstein and Zin (1989) show, the timing of the resolution of uncertainty matters in the representation of preferences. This aspect of preferences is present when we use T^1 to evaluate consumption profiles conditioned on a hidden Markov state.

Our second operator, T^2, has an alternative interpretation as relaxing the assumption that consumption lotteries conditioned on a hidden Markov state can be reduced to lotteries conditioned only on the signal history when depicting preferences. This interpretation follows Segal (1990) and Klibanoff et al. (2003). We use the T^2 operator to adjust for risk about hidden states in a different way than the usual Bayesian model averaging approach does.

1.6   Organization

To make formulas for time-varying uncertainty prices more accessible, we assemble them by constructing a sequence of models that make the representative consumer's inference problems and specification doubts more and more interesting. Section 2 describes how we can blend features of the models of Tallarini (2000) and Bansal and Yaron (2004). Section 3 conceals the state from the representative consumer and, relative to the model of section 2, describes an additional adjustment for model uncertainty. Section 4 adds another type of hidden state that indexes possible submodels. Section 5 contains a description of our application and an account of the sources of time-variation in market prices of model uncertainty. Section 6 contains some concluding remarks.

---

[7]Subsection 5.2 presents table 4 as a convenient way of showing the relationship of our three models to other models that have appeared in the literature.

2 Tallarini meets Bansal and Yaron

This section modifies the model of Tallarini (2000) by replacing his endowment process for log consumption growth with a long run risk specification like those used by Bansal and Yaron (2004). We also reinterpret the risk-sensitivity operator used by Tallarini as the indirect utility function for someone who is concerned about model misspecification, then give a formula for what Hansen et al. (1999) call the market price of model uncertainty (MPU). In this way, we fill in the first row of table 2.

2.1 A warmup model without risk sensitivity

To set the stage, we first consider a model with no risk-sensitivity adjustments. Then to incorporate the representative consumer's fear of model misspecification, sections will add risk-sensitivity adjustments to this basic model.

Let c_t be the log of consumption and define the expected present value of the log of consumption recursively by

$$V_t = (1 - \beta)c_t + E_t \beta V_{t+1} \tag{1}$$

where the growth in log consumption $s_{t+1} = c_{t+1} - c_t$ is governed by the state space system

$$\zeta_{t+1} = A\zeta_t + C\epsilon_{t+1} \tag{2}$$
$$s_{t+1} = D\zeta_t + G\epsilon_{t+1} \tag{3}$$
$$c_{t+1} = c_t + s_{t+1} \tag{4}$$

where $\epsilon_{t+1} \sim \mathcal{N}(0, I)$. Solving recursion (1) shows that V_t is given by the value function

$$V_t = \lambda'\zeta_t + c_t \tag{5}$$

where

$$\lambda' = \beta D(I - \beta A)^{-1}. \tag{6}$$

Proof. Define $v_t = V_t - c_t$ and write (5) as $v_t = E_t(\beta v_{t+1} + s_{t+1})$. This implies $v_t = \beta E_t v_{t+1} + \beta D\zeta_t$. Solving this expectational difference equation forward and using the state space system (2), (3), (4) gives $v_t = \beta D(I - \beta A)^{-1}\zeta_t$. $\square$

Let * denote a next period value and express (1) as the Bellman equation

$$V(\zeta, c) = (1 - \beta)c + E\left[\beta V(\zeta^*, c^*)|\zeta, c\right]. \tag{7}$$

In the next section, we modify the model of Tallarini (2000) by replacing his specification of consumption growth with (2), (3), (4).

2.2 A model with one risk sensitivity operator

We get our version of the model of Tallarini (2000) by replacing (7) with a value function

$$W(\zeta, c) = \lambda'\zeta + \kappa + c \tag{8}$$

that we obtain by solving the recursion

$$W(\zeta, c) = (1 - \beta)c + \mathsf{T}^1 \left[\beta W(\zeta^*, c^*) \right] \tag{9}$$

where T^1 is a *risk-sensitivity* operator defined in the following two mathematically equivalent ways:

$$\mathsf{T}^1[\beta W(\zeta^*, c^*)] = -\theta_1 \log E\left[\exp\left(\frac{-\beta W(\zeta^*, c^*)}{\theta_1} \right) \Big| \zeta, c \right] \tag{10}$$

$$= \min_{m(\epsilon^*) \geq 0, Em(\epsilon^*)=1} E\left(m(\epsilon^*) \left[\beta W(\zeta^*, c^*) + \theta_1 \log m(\epsilon^*) \right] \Big| \zeta, c \right). \tag{11}$$

where

$$\zeta^* = A\zeta + C\epsilon^*$$
$$c^* - c = D\zeta + G\epsilon^*.$$

While Tallarini (2000) emphasizes (10) and interprets T^1 as enhancing risk-aversion, we emphasize (11) and interpret T^1 as a way of implementing a version of the max-min expected utility theory of Gilboa and Schmeidler (1989) and thereby making a cautious evaluation due to the representative consumer's distrust of the model's specification of the distribution of ζ^* conditional on ζ, c. Minimization over the likelihood ratio m expresses that distrust. We allow the function m to depend on ζ and c. The nonnegative function m distorts the density of ϵ^* conditional on ζ, while $E\left[m(\epsilon^*) \log m(\epsilon^*) | \zeta, c \right]$ is the conditional entropy of the distortion. Here θ_1 is a positive risk-sensitivity parameter.

Proposition 2.1. *The fixed point of (9) is attained by a value function (8) with λ satisfying (6) and*

$$\kappa = \frac{-\beta^2}{2(1 - \beta)\theta_1} |\lambda' C + G|^2. \tag{12}$$

The associated worst distribution for ϵ^* is normal with mean

$$w^* = \frac{-\beta}{\theta_1}(C'\lambda + G') \tag{13}$$

and covariance matrix I.

Proof. The T^1 operator maps a function $\lambda'\zeta + c + \kappa$ into a value function with the same functional form. Since the value function is linear in ζ and c, and the original distribution for ϵ^* is normal, the minimizing value of m in (11) is

$$\hat{m}(\epsilon^*) = \exp\left(\epsilon^* \cdot w^* - \frac{1}{2} w^{*\prime} w^* \right) \tag{14}$$

where w^* is given by (13). This follows because the worst-case density distortion $\hat{m}(\epsilon^*)$ is proportional to

$$\exp\left[-\frac{\beta}{\theta_1}(\lambda'\zeta^* + c^* + \kappa) \right]$$

Object	Formula
$\lambda(\iota)$	$[I - \beta A(\iota)']^{-1}\beta D(\iota)$
$\kappa(\iota)$	$-\frac{\beta^2}{2(1-\beta)\theta_1}\left\|C(\iota)'\lambda(\iota) + G(\iota)\right\|^2$
$\Sigma(\iota)$	$E[\zeta(\iota) - \check{\zeta}(\iota)][\zeta(\iota) - \check{\zeta}(\iota)]'$ from Kalman filter
$w^*(\iota)$	$-\frac{\beta}{\theta_1}\left[C'(\iota)\lambda(\iota) + G(\iota)'\right]$
$u^*(\iota)$	$-\frac{1}{\theta_2}\Sigma(\iota)\lambda(\iota)$

Table 1: Components of value functions and market prices of model uncertainty.

when viewed as a function of ϵ^*. Thus, while $\hat{m}$ changes the distribution of ϵ^*, the distorted distribution is normal with mean w^* instead of zero and the same covariance matrix. The relative entropy for this mean distortion is $\frac{1}{2}w^{*\prime}w^*$. Thus, the problem is solved by: (1) computing the expectation of the discounted value function where ϵ^* has mean w^*; and (2) replacing $E[m\log m|\zeta,c]$ with $\frac{w^{*\prime}w^*}{2}$. These computations imply that

$$\mathsf{T}^1\left[\beta W(\zeta^*,c^*)\right] = \beta(\lambda'A + D)\zeta + \beta\kappa + \beta c - \frac{\beta^2}{2\theta_1}(C'\lambda + G')'(C'\lambda + G').$$

Substitute this into (9) and solve for λ and κ. $\qquad\square$

Remark 2.2. *While the minimization problem allows the distortion to the mean of the shock ϵ^* to be state dependent, the minimizing distortion turns out not to depend on the state because the value function is linear and shocks are homoskedastic. The mean distortion solves:*

$$\min_{w^*}\beta\left[\lambda'(A\zeta + Cw^*) + \kappa + c + D\zeta + Gw^*\right] + \frac{\theta}{2}w^{*\prime}w^*.$$

Remark 2.3. *The state space model (2), (3), (4) implies that a moving average representation for consumption growth is*

$$c_{t+1} - c_t = [D(I - AL)^{-1}CL + G]\epsilon_{t+1} \tag{15}$$

where L is the backward shift operator. Notice that $C'\lambda + G'$ in (13) equals the present value of the moving average coefficients in representation (15).

2.2.1 The market price of model uncertainty

Using (14), the stochastic discount factor for our model is

$$\left(\beta\exp\left(\frac{c}{c^*}\right)\right)\left(\exp(\epsilon^*w^* - .5w^{*\prime}w^*)\right), \tag{16}$$

where the first term is the ordinary stochastic discount factor without concerns about robustness for the CRRA specification with risk-aversion parameter equal to unity (i.e., logarithmic preferences) and the second term is the Radon-Nikodym derivative $\hat{m}$ given by equation (14). Hansen et al. (1999) and Hansen et al. (2002) show that the Radon-Nikodym derivative $\hat{m}$ is

a potentially volatile factor that multiplies the ordinary stochastic discount factor and that it contributes the following amount to what is ordinarily measured as the market price of risk:[8]

$$\text{std}_t(\hat{m}_{t+1}|\zeta_t) = [\exp(w'_{t+1}w_{t+1}) - 1]^{\frac{1}{2}} \approx |w_{t+1}|.$$

This approximation can be formalized by taking a continuous time limit. As explained in Hansen et al. (1999) and Hansen et al. (2002), we prefer to interpret this quantity as the market price of model uncertainty. Thus, for this model, the contribution of a concern about robustness to what is usually interpreted as the market price of risk is approximately

$$|w_{t+1}| = |w^*|,$$

which is time-invariant, as emphasized by Anderson et al. (2003). This formula explains the entry for the market price of model uncertainty MPU_t for model 1 in table 2.

2.2.2 The market prices of uncertainty for three models

In section 3, we explain the entry in table 2 for MPU_t from a model 2 that, by withholding knowledge of ζ, puts the representative consumer in a situation that requires him to filter. In model 2, the decision maker applies another risk sensitivity operator to account for his uncertainty about the distribution $\zeta \sim \mathcal{N}(\hat{\zeta}, \Sigma)$ of the hidden state that emerges from the filtering problem. The model 2 MPU_t has a time-varying component due to the presence of Σ_t as a determinant of u_t^*, but it is a smooth and deterministic function of time. The third row of table 2 summarizes the outcome of a model 3 that we form in section 4 by endowing the decision maker with two submodels for the consumption growth process and a prior probability distribution $\check{p}_t$ for averaging them. That the representative consumer doubts the specification of his prior distribution contributes adjustments that make the market price of uncertainty vary erratically over time as the arrival of new information unsettles his worst-case valuations and his worst case probabilities $\tilde{p}_t$ over submodels.[9]

Model 1 contains as many shocks as appear in ϵ_{t+1}, two in our application, while models 2 and 3 contain as many shocks as occur in the innovations representation for s_{t+1}, one in our application.

3 A model with a hidden state and a second risk-sensitivity adjustment

The model in section 2.2 assumes that the representative consumer observes (ζ_t, c_t) at time t. We now consider a model in which ζ_t is a hidden state at time t. The representative consumer observes c_t and the history $s_t, s_{t-1}, \ldots, s_0$ at time t, an information structure that

---

[8]The date $t + 1$ is used on w_{t+1} because it is the distortion of the mean of the shock ϵ_{t+1} conditional on date t information.

[9]Models 2 and 3 both implement versions of the robustness recursions that Hansen and Sargent (2006b) refer to as Game I. That paper also discusses alternative recursions in a Game II that apply $\mathsf{T}^2 \circ \mathsf{T}^1$ at each iteration and so work with value functions that depend on observed states and sufficient statistics for estimated states only, not the hidden states themselves. Hansen and Sargent (2006b) discuss how Game II focuses the decision maker's concerns about robustness in different directions than Game I.

Model	hidden states	T	# shocks	MPU_t
1	none	T^1	2	$\|w^*\|$
2	ζ	$\mathsf{T}^2, \mathsf{T}^1$	1	$\frac{1}{\sqrt{GG'}}\|Gw^* + Du_t^*\|$
3	$\iota, \zeta(\iota)$	$\mathsf{T}^2, \mathsf{T}^2, \mathsf{T}^1$	1	$\frac{1}{\sqrt{GG'}}\big\|(\check{p}_t - \tilde{p}_t)[D(1)\check{\zeta}_t(1) - D(0)\check{\zeta}_t(0)]$
				$-\tilde{p}_t[G(1)w^*(1) + D(1)u_t^*(1)]$
				$-(1 - \tilde{p}_t)[G(0)w^*(0) + D(0)u_t^*(0)]\big\|$

Table 2: Market prices of uncertainty in models with risk sensitivity. Model 1 is our robustness version of Tallarini (2000). Model 2 is a modification of a model of Hansen et al. (2002) to allow for different θ's in distinct risk-sensitivity corrections. Model 3 activates model uncertainty about the prior distribution over submodels. MPU_t is the conditional standard deviation of a multiplicative adjustment to the stochastic discount factor that is induced by concern about model misspecification. $\iota = 1$ is Bansal and Yaron's long run risk model; $\iota = 0$ is an i.i.d. consumption growth model.

confronts him with a filtering problem. We proceed to construct a model with learning that differs from those in the literature described in subsection 1.3 because we add two corrections for the representative consumer's concerns about model misspecification.[10]

3.1 Filtering implies an innovations representation

Application of the Kalman filter yields the following *innovations representation* for system (2), (3):

$$
\begin{aligned}
\check{\zeta}_{t+1} &= A\check{\zeta}_t + K(\Sigma_t)a_{t+1} \\
\Sigma_{t+1} &= A\Sigma_t A' + CC' - K(\Sigma_t)(A\Sigma_t A' + CG')' \\
s_{t+1} &= D\check{\zeta}_t + a_{t+1}
\end{aligned}
$$

where $\check{\zeta}_t = E[\zeta_t | s_t, \ldots, s_1]$, $a_{t+1} = s_{t+1} - E[s_{t+1} | s_t, \ldots, s_1]$, $\Sigma_t = E(\zeta_t - \check{\zeta}_t)(\zeta_t - \check{\zeta}_t)'$, $\Delta_t \equiv E a_{t+1}a_{t+1}' = D\Sigma_t D' + GG'$, and

$$
K(\Sigma) = (A\Sigma D' + CG')(D\Sigma D' + GG')^{-1}.
$$

At time t, $(\check{\zeta}_t, \Sigma_t)$ are sufficient statistics for the history of signals $s_t, \ldots, s_0$.

3.2 Distrust of the distribution over ζ_t

We first imagine a representative consumer who values consumption streams by using the following two step procedure:

1. Temporarily pretend that ζ_t is observed and proceed as in section 2.2 to obtain the value function

$$
W(\zeta, c) = \lambda'\zeta + \kappa + c. \tag{17}
$$

---

[10]However, rows 1a and 1b of table 4 display a specification without T^2 that does share the features emphasized in subsection 1.3.

2. Now imagine that the representative consumer acknowledges that ζ_t is not observed by taking the conditional expectation of (17) with respect to the conditional distribution of ζ pinned down by $(\check{\zeta}, \Sigma)$ to get

$$W(\check{\zeta}, c) = \lambda'\check{\zeta} + \kappa + c.$$

To introduce our second risk-sensitivity operator, we amend this two-step procedure by replacing the conditional expectation in the second step with a risk-sensitivity operator that is designed to make an evaluation that is cautious with respect to the distribution of ζ conditioned on $(\check{\zeta}, \Sigma)$. This risk-sensitivity adjustment is

$$\mathsf{T}^2(W(\zeta, c)|\check{\zeta}, c, \Sigma)|\theta_2 = -\theta_2 \log \int \exp\left(\frac{-W(\zeta, c)}{\theta_2}\right)\phi(\zeta|\check{\zeta}, \Sigma)d\zeta$$

$$= \min_{h(\zeta)\geq 0, \int h(\zeta)\phi(\zeta|\check{\zeta},\Sigma)=1} \int \left[W(\zeta, c) + \theta_2 \log h(\zeta)\right]h(\zeta)\phi(\zeta|\check{\zeta}, \Sigma)d\zeta,$$

where θ_2 is another positive risk-sensitivity parameter and $\phi(\zeta|\check{\zeta}, \Sigma)$ is a Gaussian density with mean $\check{\zeta}$ and covariance matrix Σ. The minimizing h is

$$h(\zeta) \propto \exp\left(\frac{-(\lambda'\zeta + \kappa + c)}{\theta_2}\right) \tag{18}$$

where the factor of proportionality is chosen to make h have expectation one when integrated against ϕ. The worst case density for ζ is $h(\zeta)\phi(\zeta|\check{\zeta}, \Sigma)$. This density is normal with mean $u^* + \check{\zeta}$ and covariance matrix Σ.

Proposition 3.1. *The adjusted value function is:*

$$\mathsf{T}^2(W(\zeta, c)|\check{\zeta}, c, \Sigma|\theta_2) = \lambda'\check{\zeta} + \kappa - \frac{1}{2\theta_2}\lambda'\Sigma\lambda + c \tag{19}$$

and the mean of the worst-case normal distribution is:

$$u^* = -\frac{1}{\theta_2}\Sigma\lambda. \tag{20}$$

Proof. From (18), the worst-case density for ζ is proportional to,

$$h(\zeta)\exp\left[-\frac{1}{2}\left(\zeta - \check{\zeta}\right)'\Sigma^{-1}\left(\zeta - \check{\zeta}\right)\right] = \exp\left[\frac{-(\lambda'\zeta + \kappa + c)}{\theta_2} - \frac{1}{2}\left(\zeta - \check{\zeta}\right)'\Sigma^{-1}\left(\zeta - \check{\zeta}\right)\right].$$

It follows from a standard complete the square argument that this density is normal with mean $u^* + \check{\zeta}$ given by (20) and covariance matrix Σ. The entropy of the minimizing distortion is:

$$Eh \log h = \frac{1}{2}u^{*'}\Sigma^{-1}u^* = \frac{1}{2\theta_2^2}\lambda'\Sigma\lambda.$$

After substituting for the distorted mean and this measure of entropy, we obtain the value function (19). □

Remark 3.2. *The worst case distribution for ζ is normal with a distortion to the mean only. The distorted mean and resulting value function can be computed by solving:*

$$\mathsf{T}^2(W(\zeta, c)|\check{\zeta}, c, \Sigma)|\theta_2 = \min_u\left(\lambda'\check{\zeta} + \lambda'u + \kappa + c + \frac{\theta_2}{2}u'\Sigma^{-1}u\right).$$

11

3.3 Worst case density and the market price of model uncertainty

The worst case density for next period's consumption growth rate signal s_{t+1} that is induced by applying T^2 to the fixed point of recursions on (9) using T^1 is:

$$
\begin{aligned}
&\mathcal{N}(D\check{\zeta} + Gw^* + Du_t^*, GG' + D\Sigma_t D') \\
&= \mathcal{N}\left(D\check{\zeta} - \frac{\beta}{\theta_1}G(C'\lambda + G') - \frac{1}{\theta_2}D\Sigma_t\lambda, GG' + D\Sigma_t D'\right)
\end{aligned}
$$

Thus, $Gw^* + Du_t^*$ is the distortion to the mean of the scalar random variable s_{t+1} that has variance $GG' + D\Sigma_t D' = \Delta_t$. It follows the market price of uncertainty is approximately

$$
\text{std}_t[g(s_{t+1}|s^t)] \approx \frac{|Gw^* + Du_t^*|}{\sqrt{\Delta_t}} \tag{21}
$$

where g is the ratio of the distorted conditional density for s_{t+1} to the original conditional density. This approximation can be justified by taking a continuous time approximation, in which case $\Delta_t = GG'$. T^1 contributes w^* and T^2 contributes u_t^*. Formula (21) is the entry in the second row of table 2. The component u_t^* reflects the representative consumer's caution with respect to the distribution of the hidden state and contributes a time-varying component to the price of risk. That this component is time-varying is attributable to our having assumed that the representative consumer observes a finite history of consumption growth signals. Hansen et al. (2002) considered a model in which the length of the history of signals is infinite, allowing u_t^* to converge to a constant in their model.[11]

4 A model with two submodels and yet another risk-sensitivity operator

4.1 Two submodels without another risk-sensitivity operator

We now consider two submodels for consumption growth s_{t+1}, each of which has the form (2), (3), (4). We denote the submodels $\iota \in \{0, 1\}$ and suppose that the representative consumer attaches probability $\check{p}_t = E(\iota)|\mathcal{S}_t$ to model 1 at time t, where $\mathcal{S}_t$ denotes the sigma algebra generated by the signal history. As we are about to describe, these probabilities can be computed by using Bayes' rule and data $s^t = [s_t, s_{t-1}, \ldots, s_1]$.

4.2 Submodel ι and its innovations representation

Submodel ι has a state space representation

$$
\begin{aligned}
\zeta_{t+1}(\iota) &= A(\iota)\zeta_t(\iota) + C(\iota)\epsilon_{t+1} \\
s_{t+1} &= D(\iota)\zeta_t(\iota) + G(\iota)\epsilon_{t+1}.
\end{aligned}
$$

---

[11]When $\theta_1 = \theta_2$, the mean distortion for the signal depends directly on the covariance of the signal with the hidden state, $GC' + G\Sigma_t A'$, and on the covariance matrix of the signal, $GG' + D\Sigma_t D'$, both conditioned on the signal history. When θ_1 and θ_2 are distinct, the decomposition of these objects into the contributions from the underlying shock vector w_{t+1} and $\zeta_t(\iota) - \check{\zeta}_t$ matters.

As does the single ι model of section 3, submodel ι has an innovations representation:

$$\begin{aligned}
\check{\zeta}_{t+1}(\iota) &= A(\iota)\check{\zeta}_t(\iota) + K[\Sigma_t(\iota), \iota]a_{t+1}(\iota) \\
s_{t+1} &= D(\iota)\check{\zeta}_t + a_{t+1}(\iota)
\end{aligned} \tag{22}$$

where $\check{\zeta}_{t+1}(\iota) = E[\zeta_{t+1}|s_{t+1}, \ldots, s_1, s_0, \iota]$,

$$K[\Sigma_t(\iota), \iota] \doteq [A(\iota)\Sigma_t(\iota)D(\iota)' + C(\iota)G(\iota)'][D(\iota)\Sigma_t(\iota)D(\iota)' + G(\iota)G(\iota)']^{-1},$$

$$\Sigma_{t+1}(\iota) = A(\iota)\Sigma_t(\iota)A(\iota)' + C(\iota)C(\iota)' - K(\Sigma_t(\iota))(A(\iota)\Sigma_t(\iota)A(\iota)' + C(\iota)G(\iota)')',$$

$a_{t+1}(\iota)$ is the forecast error for the signal (i.e., the 'innovation'), and $\Sigma_t(\iota)$ is the covariance matrix for $\zeta_t(\iota) - \check{\zeta}_t(\iota)$ conditioned on ι and the signal history through date t.

4.3 The likelihood of s^t

To construct a likelihood function for a history $s^t = s_t, s_{t-1}, \ldots, s_1$ conditioned on model ι, we use (22) to compute the date $t + 1$ signal innovation $a_{t+1}(\iota)$ for model ι recursively from the innovations representation and form the conditional variance for the signal based on date t information, namely, $\Delta_t(\iota) = G(\iota)G(\iota)' + D(\iota)\Sigma_t(\iota)D(\iota)'$. Then the log-likelihood of model ι can be represented recursively as

$$\mathfrak{L}_T(\iota) = \sum_{t=1}^{T} \ell_t(\iota)$$

where

$$\ell_t(\iota) = -\frac{1}{2}\left[\log(2\pi) + \log\det[\Delta_{t-1}(\iota)] + a_t(\iota)'[\Delta_{t-1}(\iota)]^{-1}a_t(\iota)\right].$$

We compute $\check{p}_t$ by

$$\check{p}_t = \frac{\check{p}_0 \exp[\mathfrak{L}_t(1) - \mathfrak{L}_t(0)]}{(1 - \check{p}_0) + \check{p}_0 \exp[\mathfrak{L}_t(1) - \mathfrak{L}_t(0)]}.$$

Alternatively, we can update $\check{p}_t$ recursively using:

$$\check{p}_{t+1} = \frac{\check{p}_t \exp[\ell_{t+1}(1) - \ell_{t+1}(0)]}{(1 - \check{p}_t) + \check{p}_t \exp[\ell_{t+1}(1) - \ell_{t+1}(0)]}.$$

In the appendix, we deduce the continuous time limit of this updating equation:

$$d\check{p}_t = \check{p}_t(1 - \check{p}_t)\left[\check{\mu}_t(1) - \check{\mu}_t(0)\right]\Delta^{-1}(ds_t - \check{\mu}_t dt)$$

where $\check{\mu}_t(\iota) = \check{\zeta}_t(\iota)'D(\iota)'$, $\check{\mu}_t = \check{p}_t\check{\mu}_t(1) + (1 - \check{p}_t)\check{\mu}_t(0)$, $\Delta = GG'$, and G is independent of ι.[12] Here $\check{\mu}_t(\iota)$ is the estimated conditional mean of the signal conditioned on model ι. The probability assigned to the model with the larger conditional mean increases when the signal (i.e., the consumption growth rate) exceeds its mean across models.

---

[12]To justify the continuous time limit, we need to assume that $G(0) = G(1)$ so that the volatility in the signal equation is the same for each model. Our estimates suggest that this is a good approximation for our two models. Without this independence, the probabilities would be degenerate. A difference across models in local variances would be detected by the high frequency data and that would reveal the statistically preferred model.

4.3.1 Two submodels with another risk-sensitivity operator

A hidden state vector $z = [\zeta(0), \zeta(1), \iota]$. In the previous section, we prepared all of the objects that we need to form the joint density of z conditioned on the history of the signal. Conditioned on submodel ι, the density of $\zeta(\iota)$ is normal with mean $\check{\zeta}(\iota)$ and covariance matrix $\Sigma(\iota)$. We also showed how to compute the marginal probabilities $1 - \check{p}$ and $\check{p}$ of the two submodels, conditional on the signal history. Multiply these densities to obtain the joint density for $z = [\zeta(0), \zeta(1), \iota]$ conditioned on the signal history.

The representative consumer evaluates discounted streams of c_t by using a value function formed in the following three step process:

1. To account for specification doubts about the distribution of s_{t+1} conditional on both the state $\zeta(\iota)$ and model ι, apply T^1 to each submodel ι. Thus, conditional on $z = [\zeta(0), \zeta(1), \iota]$, we form two value functions of the form that we presented in section 2.2:

$$W(\zeta(\iota), c, \iota) = \lambda(\iota) \cdot \zeta(\iota) + \kappa(\iota) + c$$

where

$$\lambda(\iota) = [I - \beta A(\iota)']^{-1} \beta D'(\iota) \tag{23}$$

$$\kappa(\iota) = -\frac{\beta^2}{2(1-\beta)\theta_1} |\lambda(\iota)'C(\iota) + G(\iota)|^2. \tag{24}$$

The mean of the associated worst case shock vector for model ι is

$$w^*(\iota) = -\frac{\beta}{\theta_1} \Big[C(\iota)'\lambda(\iota) + G'(\iota) \Big], \tag{25}$$

a distortion that is familiar from section 2.2, Hansen et al. (1999), and Anderson et al. (2003), except now it depends on the model ι. The distorted distribution for the signal s_{t+1} is normal with mean:

$$D(\iota)\zeta_t(\iota) - \frac{\beta}{\theta_1} [C(\iota)'\lambda(\iota) + G(\iota)'],$$

and covariance matrix $G(\iota)G(\iota)'$ conditioned on ι and $\zeta_t(\iota)$.

2. To account for uncertainty about the distribution of the state $\zeta(\iota)$ for model ι, apply T^2 conditional on ι. We proceed as in section 3:

$$\mathsf{T}^2[\lambda(\iota) \cdot \zeta + \kappa(\iota) + c | \iota, \theta_2] \left(\check{\zeta}(\iota), \Sigma(\iota) \right) = U\left(\iota, \check{\zeta}(\iota), \Sigma(\iota) \right) + c$$

where

$$U[\iota, \check{\zeta}(\iota), \Sigma(\iota)] \doteq \lambda(\iota) \cdot \check{\zeta} + \kappa(\iota) - \frac{1}{2\theta_2} \lambda(\iota)'\Sigma(\iota)\lambda(\iota).$$

The mean of the distorted distribution for ζ is $\check{\zeta} + u^*$, where

$$u^*(\iota) = -\frac{1}{\theta_2} \Sigma(\iota)\lambda(\iota). \tag{26}$$

14

The implied worst case distribution of $\zeta_t(\iota) - \check{\zeta}_t(\iota)$ is altered from normal mean zero with covariance $\Sigma_t(\iota)$ to normal with mean $-\frac{1}{\theta_2}\Sigma_t(\iota)\lambda(\iota)$ and covariance matrix $\Sigma_t(\iota)$. As a consequence, s_{t+1} is normal with mean:

$$D(\iota)\check{\zeta}_t(\iota) - \frac{\beta}{\theta_1}G(\iota)\left[C(\iota)'\lambda(\iota) + G(\iota)'\right] - \frac{1}{\theta_2}D(\iota)\Sigma_t\lambda(\iota).$$

The conditional covariance matrix of s_{t+1} is $G(\iota)G(\iota)' + D(\iota)\Sigma_t(\iota)D(\iota)'$ conditioned on the model ι and the signal history.

3. To capture the representative consumer's distrust of his prior over submodels, we apply a second T^2 operator that makes a robust adjustment of the model probabilities $\check{p}$ and $1 - \check{p}$. T^2 applied to $U[\iota, \check{\zeta}(\iota), \Sigma(\iota)]$ is:[13]

$$
\begin{aligned}
& \mathsf{T}^2 U[\iota, \check{\zeta}(\iota), \Sigma(\iota)]) \hspace{4cm} (27)\\
= {} & -\theta_2 \log\left[(1-\check{p})\exp\left(-\frac{U[0,\check{\zeta}(0),\Sigma(0)]}{\theta_2}\right) + \check{p}\exp\left(-\frac{U[1,\check{\zeta}(1),\Sigma(1)]}{\theta_2}\right)\right]\\
= {} & \min_{0\le\tilde{p}\le 1}\left\{\left[U[0,\check{\zeta}(0),\Sigma(0)] + \theta_2[\log(1-\tilde{p}) - \log(1-\check{p})]\right](1-\tilde{p}) + \right.\\
& \left. \left[U[1,\check{\zeta}(1),\Sigma(1)] + \theta_2[\log(\tilde{p}) - \log(\check{p})]\right]\tilde{p}\right\}.
\end{aligned}
$$

The twisted submodel probabilities are:

$$(1-\tilde{p}_t) \;\propto\; (1-\check{p}_t)\exp\left(-\frac{U[0,\check{\zeta}(0),\Sigma(0)]}{\theta_2}\right) \hspace{2cm} (28)$$

$$\tilde{p}_t \;\propto\; \check{p}_t\exp\left(-\frac{U[1,\check{\zeta}(1),\Sigma(1)]}{\theta_2}\right). \hspace{2cm} (29)$$

Please note how equations (28), (29) shift probability toward the submodel ι that has the worse value at each time t. This aspect of $\tilde{p}$ plays an important role in shaping the evolution of the market price of model uncertainty.

An outcome of this three step process is an implied worst case density for s_{t+1} conditioned on current and past signals that is a mixture of normals with $(1 - \tilde{p}_t, \tilde{p}_t)$ as probability weights. To compute the market price of model uncertainty, we want to deduce the ratio of this density to that implied by the approximating model.

4.4 Multiplicative adjustment to stochastic discount factor

Denote the likelihood of the approximating model as

$$\check{f}(s_{t+1}|s^t) = (1-\check{p}_t)\check{f}(s_{t+1}|s^t,0) + \check{p}_t\check{f}(s_{t+1}|s^t,1) \hspace{2cm} (30)$$

---

[13]We ignore the consumption contribution to value functions conditioned on ι because it is common to both value function and hence serves as a common proportionality factor in the probability distortions.

and let the distorted likelihood be

$$\tilde{f}(s_{t+1}|s^t) = (1 - \tilde{p}_t)\tilde{f}(s_{t+1}|s^t, 0) + \tilde{p}_t\tilde{f}(s_{t+1}|s^t, 1). \tag{31}$$

We have computed all of the objects appearing in (30) and (31). In particular, for $, \iota = 0, 1$, $\tilde{f}(s_{t+1}|s^t, \iota)$ is normal with means and variances that we computed in section 3. The two model-averaged signal densities are mixtures of normals with weights $\check{p}_t$ and $\tilde{p}_t$, respectively.

We now want to study the statistical time behavior of time series for the following Radon-Nikodym derivative:

$$g(s_{t+1}|s^t) = \frac{\tilde{f}(s_{t+1}|s^t)}{\check{f}(s_{t+1}|s^t)}. \tag{32}$$

Under the approximating model, $g(s_{t+1}|s^t)$ serves as a multiplicative adjustment to the stochastic discount factor that accounts for the representative consumer's fear of model misspecification. By construction, the random variable $g(s_{t+1}|s^t)$ has mean one under the approximating model.

In the next section, we describe a Gaussian approximation to the worst case distribution and a formula for the likelihood ratio $g(s_{t+1}|s^t)$ that we shall motivate by a continuous time approximation.

4.5 Continuous time approximation

The mean consumption growth rate under the approximating model is

$$\check{\mu}_s = \check{p}D(1)\check{\zeta}(1) + (1 - \check{p})D(0)\check{\zeta}(0).$$

Appendix B uses a continuous time model and the assumption that $G(1) = G(0)$ to motivate the following approximation to the mean of the growth rate under the worst case model:

$$\tilde{\mu}_s = \tilde{p}[D(1)\check{\zeta}(1) + G(1)w^*(1) + D(1)u^*(1)] + (1 - \tilde{p})[D(0)\check{\zeta}(0) + G(0)w^*(0) + D(0)u^*(0)].$$

Subtracting and rearranging gives the following distortion to the mean of consumption growth:

$$\begin{aligned}
\check{\mu}_s - \tilde{\mu}_s &= (\check{p} - \tilde{p})\Big[D(1)\check{\zeta}(1) - D(0)\check{\zeta}(0)\Big] - \tilde{p}\Big[G(1)w^*(1) + D(1)u^*(1)\Big] \\
&\quad - (1 - \tilde{p})\Big[G(0)w^*(0) + D(0)u^*(0)\Big].
\end{aligned}$$

Appendix B also shows that under both the approximating model and the worst case model, s_{t+1} has conditional covariance GG', so that the market price of model uncertainty in this model is

$$\text{std}_t[g(s_{t+1}|s^t)] \approx \left|\frac{\check{\mu}_{s,t} - \tilde{\mu}_{s,t}}{\sqrt{GG'}}\right|, \tag{33}$$

where g is the distorted conditional density for s_{t+1} relative to the original conditional density. Formula (33) is the entry for the MPU_t for model 3 in table 2. Formulas (33), (33) isolate variations in $\check{\zeta}_t(0), \check{\zeta}_t(1), \check{p}_t, \tilde{p}_t$ as sources of *erratic* time variations in the market price of model uncertainty, while $u_t^*(0), u_t^*(1)$ contribute smooth time variation, and $w^*(0), w^*(1)$ contribute constant components.

ι	$\zeta(\iota)$	meaning
0	$\zeta_0(0)$	$E[(c_{t+1} - c_t)|\iota = 0]$
1	$\zeta_{1,t}(1)$	persistent component
1	$\zeta_{2,0}(1)$	$E[(c_{t+1} - c_t)|\iota = 1]$

Table 3: Interpretation of state variables in two submodels, where $\iota = 1$ is the long-run risk submodel and $\iota = 0$ is the i.i.d. consumption growth submodel.

5 Doubts about long run risk

In this section, we reinterpret some of the observations about long run risk made by Bansal and Yaron (2004) to justify the hypothesis that a representative consumer carries along two submodels for consumption growth, each taking the form of (2), (3), (4). Bansal and Yaron's argument that these two models are difficult to distinguish even with moderately long time series, prompts us to take a different approach than they do and to posit that the representative consumer doubts the stochastic specification of each submodel as well as his posterior distribution over the submodels at each date. We calibrate the two submodels and use them to study how the various components of (33) contribute to the market price of model uncertainty.

5.1 Two submodels for consumption growth

The representative consumer indexes two submodels for the signal $s_{t+1} = c_{t+1} - c_t$ by the time-invariant hidden state $z_{3,t} = z_{3,0} = \iota \in \{0,1\}$, where $\iota = 0$ denotes "model 0" and $\iota = 1$ denotes "model 1". The state vector z_t used in our general formulation consists of $\zeta_t(1), \zeta_t(0), \iota$.[14]

We take $\iota = 1$ to be the long run risk model and $\iota = 0$ to be the iid consumption growth model. In both models, the consumption growth rate is the signal. For easy reference, table 3 describes the state variables in the two submodels. In each submodel, the unconditional mean of consumption growth is an unknown parameter that becomes a constant component of the hidden state.

---

[14]We have defined the random variable $\zeta_t(1)$ only for $\iota = 1$, but we could extend it by letting it be identically zero when $\iota = 0$ and similarly for $\zeta_t(0)$. Such extensions are inconsequential because the consumer cares about the distribution of $\zeta_t(1)$ only when $\iota = 1$ and about the distribution of $\zeta_t(0)$ only when $\iota = 0$.

5.1.1 Long run risk model

The $\iota = 1$ model is:[15]

$$\begin{bmatrix} \zeta_{1,t+1}(1) \\ \zeta_{2,t+1}(1) \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \zeta_{1,t}(1) \\ \zeta_{2,t}(1) \end{bmatrix} + \begin{bmatrix} \sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{bmatrix}$$

$$s_{t+1} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \zeta_{1,t}(1) \\ \zeta_{2,t}(1) \end{bmatrix} + \begin{bmatrix} 0 & \sigma_2(1) \end{bmatrix} \begin{bmatrix} \epsilon_{1,t+1} \\ \epsilon_{2,t+1} \end{bmatrix} \tag{34}$$

where $\zeta_{2,0}(1) = \mu(1)$. We set the initial state covariance matrix to be:

$$\Sigma_0(1) = \begin{bmatrix} \sigma_1^2/(1-\rho^2) & 0 \\ 0 & v_2(1) \end{bmatrix}$$

where $v_2(1) = 3 * .0008^2$, which is three times its counterpart from the maximum likelihood estimates. The initial mean for the state is set at $\begin{bmatrix} \check{\zeta}_{1,0}(1) \\ \check{\zeta}_{2,0}(1) \end{bmatrix} = \begin{bmatrix} 0 \\ .0054 \end{bmatrix}$.

5.1.2 iid growth model

The $\iota = 0$ model has a single state variable $\zeta_t(0)$ that is time invariant with initial condition $\mu(0)$. Thus, we write the state evolution as $\zeta_{t+1}(0) = \zeta_t(0)$ and the signal equation as:[16]

$$s_{t+1} = \zeta_t(0) + \begin{bmatrix} 0 & \sigma_2(0) \end{bmatrix} \epsilon_{t+1}. \tag{35}$$

We set the scalar $\Sigma_0(0) = 3 * .0008^2$, which is three times its maximum likelihood estimate, and we set the initial mean $\check{\zeta}(0)$ for the scalar state to be .0056.

5.2 Comparisons with related models

Table 4 summarizes some assumptions about information sets, robustness parameters θ_1, θ_2, and the initial probability $\check{p}_0$ attached to the long-run risk submodel $\iota = 1$ in the three models in table 2 as well as in models that appear in other papers that can be viewed as special cases of the model of this paper. What we call model $1a$ gives the representative consumer an information set that consists of an infinite record of past consumption growth

---

[15]To capture the 'long run risk' specification of Bansal and Yaron, we imposed $\rho = .98, \sigma_1 = .00025$ and estimated $\sigma_2(1) = .0053(.0003)$ and $\zeta_{2,0}(1) = .0054(.0008)$ by maximum likelihood where standard errors are in parentheses. The estimates are based on 228 quarterly observations for consumption growth per capita of nondurables plus services from 1947.2–2003.3. When we also estimated ρ, σ_1 by maximum likelihood, we obtained a point estimate of ρ that was too low to capture Bansal and Yaron's idea of long run risk, but ρ is estimated very imprecisely. Therefore, we simply imposed the values just mentioned. These values do capture Bansal and Yaron's idea: figure 3 confirms that the resulting long run risk model is difficult to distinguish from the i.i.d. consumption growth model. Hansen (2007) studies a setting where there are multiple parameter configurations of ρ and σ_1 with similar log-likelihoods. Also, Hansen confirms that the likelihood conveys little information about the values of these parameters. Hansen (2007) constructs a risk-sensitivity (exponential tilting) adjustment that expresses the representative consumer's uncertainty about these parameters and studies how it affects the market price of uncertainty.

[16]To obtain parameter values for the i.i.d. consumption submodel $\iota = 0$, we estimated $\zeta_{2,0}(0) = .0056(.0004)$ and $\sigma_2(0) = .0054(.0003)$.

Model	$\check{p}_0$	Info	θs	papers
1	1	$\zeta_t(1), c_t$	θ_1	Tallarini (2000)
				Bansal and Yaron (2004)
1a	1	$\zeta_{2,0}(1), c_t, s^t_{-\infty}$	$\theta_1 = \theta_2$	Anderson et al. (2003)
				Croce et al. (2006)
1b	1	$\zeta_{2,0}(1), c_t, s^t_0$	$\theta_1 = \theta_2$	Hansen et al. (2002)
2	1	c_t, s^t_0	θ_1, θ_2	
3	$\in (0,1)$	c_t, s^t_0	θ_1, θ_2	current paper

Table 4: Comparisons of various models in the literature.

rates, as well as knowledge of the unconditional mean consumption growth rate $\zeta_{2,0}$. To capture this specification, this model initializes $\Sigma(1)$ at the steady state of iterations on the Riccati equation for $\Sigma(1)_{t+1}$). This specification was used by Anderson et al. (2003), Hansen et al. (2002), and Croce et al. (2006).[17] To capture the reduced information specification in model 1b, we can set $\Sigma_0(1)$ at an arbitrary matrix exceeding the conditional covariance matrix used in model 1a. For example, for those components of $\zeta(1)$ that are asymptotically stationary, we can imagine setting the corresponding components of $\Sigma_0(1)$ to those values. Model 2 generalizes model 1b by allowing the representative consumer to distinguish between his doubts about the distribution of s_{t+1} conditional on $\zeta_t(1)$, on the one hand, and his doubts about the distribution of $\zeta_t(1)$ itself, on the other.

5.3 A reinterpretation of the consumer's attachment to the long-run risk model

As in Hansen et al. (1999) or Tallarini (2000), the representative consumer's specification doubts contribute a multiplicative adjustment to the stochastic discount factor that equals a Radon-Nikodym derivative of the representative consumer's worst case distribution of consumption growth relative to the distribution given by his approximating model. The representative consumer assigns worst case probabilities to the iid and long-run risk consumption growth models, respectively, in order to form that Radon-Nikodym derivative. For our calibration, we show how the representative consumer slants his worst case probabilities toward the long run risk model $\iota = 1$. For an interesting setting of (θ_1, θ_2), we show that his worst case model puts almost all of the probability on the long-run risk model. In this case, our representative consumer becomes almost like Bansal and Yaron's by fully embracing the long run risk specification. But that is not the end of the story in our setting. We show that even in this case, the fact that the iid model remains on the table under the approximating model contributes an interesting component of risk-premia that varies over time as the Bayesian probability attached to the long run risk model fluctuates under the consumer's approximating model. Thus, doubts about the model specification continue to affect market prices of risk even when our representative consumer acts 'as if' (under his worst case model) he attaches probability close to one to the long-run risk model.

---

[17]Appendix A describes an alternative recursion on $\mathsf{T}^2 \circ \mathsf{T}^1$ that with $\theta_1 = \theta_2$ leads to our version of the model of Croce et al. (2006), and also explains why in the present context that recursion gives outcomes that are the same as those considered in this paper.
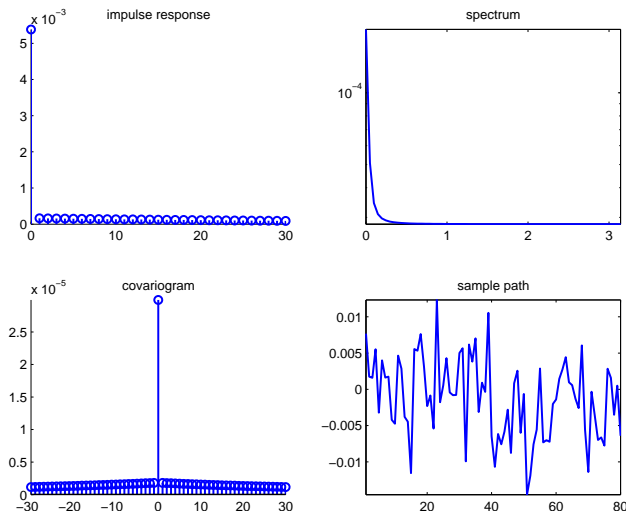
Figure 1: Long run risk consumption growth process

5.4   Features of the long-run risk model

Figure 1 reports the impulse response, spectrum, covariogram, and sample path for the consumption growth process given by the long-run risk model (34). The spectrum is flat at all except very low frequencies. The covariogram and impulse response function closely resemble ones associated with an i.i.d. process. On the basis of outcomes like those summarized in figure 1, Bansal and Yaron argued that consumption growth data alone make it difficult to distinguish a long-run risk model like (34) that makes log consumption growth equal to an i.i.d. process plus a persistent process with low variance from the alternative specification that log consumption growth is an i.i.d. process. They use that observation to justify imputing a model like (34) to a representative consumer.

We start from Bansal and Yaron's observation that the i.i.d. consumption growth model (model $\iota = 0$) and the long run risk model (34) (model $\iota = 1$) are difficult to distinguish except in long data sets. (The probabilities $\check{p}_t = E(\iota|s^t)$ reported in figure 3 below confirm that perception.) As emphasized above, we depart from Bansal and Yaron's approach by assuming that our representative consumer expresses his belief that these two submodels are difficult to distinguish by having him attach positive probability to both. We allow him to learn which model the data favor as time passes. We also depart from Bansal and Yaron by assuming that the representative consumer observes neither the state variable $\zeta_{1,t}(1)$ nor the unconditional mean growth rates $\zeta_{2,t}(\iota)$ for $\iota = 0, 1$. Therefore, the consumer estimates them.

5.5   Sub model probabilities

The top panel of figure 2 plots consumption growth against the mean of consumption growth conditioned on the signal history for the long-run risk model $\iota = 1$, while for comparison the bottom panel shows the conditional mean from the i.i.d. consumption growth model $\iota = 0$. The conditional mean from model $\iota = 0$ is less volatile than the conditional mean from model
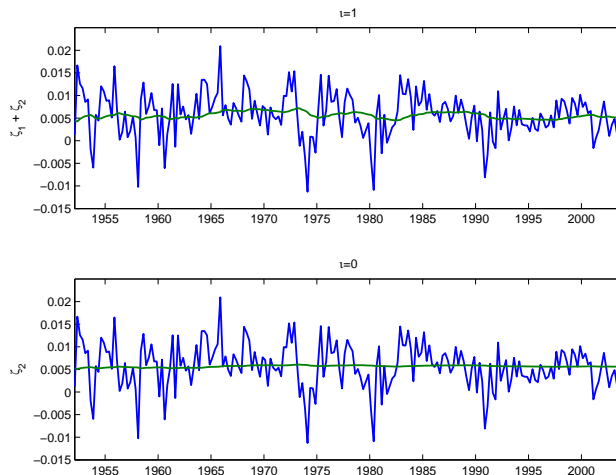
Figure 2: Top panel: Consumption growth and $\check{\zeta}_{1,t}(1) + \check{\zeta}_{2t}(1)$; bottom panel: consumption growth and $\check{\zeta}_t(0)$.

$\iota = 1$.

Figure 3 plots $\check{p}_t$ for our data on the quarterly growth rate of per capita U.S. consumption over the period 1947.2-2003.3 starting from an initial condition $\check{p}_0 = .5$ that puts equal weight on the two models described above. The fact that $\check{p}_t$ wanders in the range $[0, 1]$ confirms Bansal and Yaron's claim that a long-run risk model is difficult to distinguish from an i.i.d. model for consumption growth. In figure 4, we shall plot $\check{p}_t$ along side a twisted version of it that is induced by application of the T^2 operator.

5.6   Robust model averaging

For the calibrated models $\iota = 0$ and $\iota = 1$ in section 5.4 and $\beta = .998$, the dashed (red) line in figure 3 reports the worst case probabilities $\tilde{p}_t$ given by (28) for our calibrated submodels when $\check{p}_{1,0} = .5$. We emphasize concerns about robustness in state estimation and model choice rather than misspecified dynamics for each submodel by setting $\theta_2 = .2$ and $\theta_1 = 20$. While Bansal and Yaron (2004) argue that long-run risk model is hard to distinguish from other time series models of consumption, their representative consumer nevertheless commits to a known long-run risk model. In contrast, our representative consumer leaves both models on the table. The worst-case probabilities $\tilde{p}_t$ that are depicted in figure 3 indicate that the representative consumer's concern for robustness makes him slant model selection probabilities towards the long run risk model because, as we shall document below, relative to the i.i.d. consumption growth model, the long-run risk model has adverse consequences for discounted utility. A cautious investor mixes submodels by slanting probabilities towards the model with the lower discounted expected utility.

Formula (28) tells how the slanting of the model mixing probabilities $\check{p}$ depends on the magnitudes of the robust value functions conditioned on the model ι:

$$U[\iota, \check{\zeta}(\iota), \Sigma(\iota)] + c = \left[\lambda(\iota) \cdot \check{\zeta}_t(\iota) + \kappa(\iota) \right] + \left[-\frac{1}{2\theta_2} \lambda(\iota)' \Sigma_t(\iota) \lambda(\iota) \right] + c. \tag{36}$$
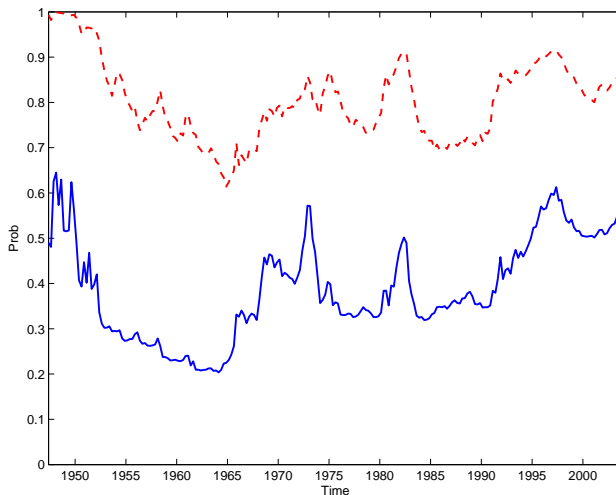
21

Figure 3: Bayesian probability $\check{p}_t$ attached to model 1 for U.S. quarterly consumption (non-durables plus services) per capita for $\check{p}_0 = .5$ (solid blue line) and worst case probability $\tilde{p}_t$ associated with $\theta_1 = 20, \theta_2 = .2$ (dashed red line).

The T^2 operator slants probabilities toward the model for which (36) is smaller, in an amount determined by the parameter θ_2. The worst case probabilities in figure 3 are slanted more when there are large differences in the value functions $U[\iota, \check{\zeta}_t(\iota), \Sigma(\iota)]$ for $\iota = 0, 1$. These differences in turn are driven largely by movements in estimates of the long run risk component. When the estimate of the long run risk component $\zeta_1(1)$ is high, the discounted expected utility is also typically expected to be high because the high growth rate in consumption persists.

The top left panel of figure 4 displays $\tilde{p}_t - \check{p}_t$, while the remaining panels shows objects that contribute to the fluctuations in $\tilde{p}_t - \check{p}_t$. The top right panel displays the difference across models in the key object in formula (28), $U[0, \check{\zeta}_t(0), \Sigma(0] - U[\iota, \check{\zeta}_t(1), \Sigma(1)]$, while the bottom two panels decompose this difference into two parts. The lower left panel shows $\lambda(0) \cdot \check{\zeta}_t(0) - \lambda(1) \cdot \check{\zeta}_t(0)$ while the lower right shows $-\frac{1}{2\theta_2}\lambda(0)'\Sigma_t(0)\lambda(0) + \frac{1}{2\theta_2}\lambda(1)'\Sigma_t(1)\lambda(1)$. These two sum to the top right panel. The difference of $\lambda(\iota) \cdot \check{\zeta}_t(\iota)$ between the submodels plotted in the lower left panel is evidently the important source of the time series variation in the difference between $U[\iota, \check{\zeta}(\iota), \Sigma(\iota)]$ for $\iota = 0, 1$. Recall that $\lambda(\iota) \cdot \zeta_t(\iota) + c$ is the expected discounted utility conditioned on the hidden state. From the Law of Iterated Expectations, $\lambda(\iota) \cdot \check{\zeta}_t(\iota) + c$ is the expected discounted utility conditioned on the signal history. This source of fluctuations would be present even when $\theta_1 = +\infty$. It varies more when $\iota = 1$ than when $\iota = 0$. The difference across submodels in $-\frac{1}{2\theta_2}\lambda(\iota)'\Sigma_t(\iota)\lambda(\iota)$ is the contribution in (36) that comes from the first application of the T^2. This term reflects a smooth secular movement that is induced by learning about the unconditional means of consumption growth in the two submodels. The T^1 operator contributes a distinct constant $\kappa(\iota)$ to the discounted expected utility that depends on θ_1.[18]

---

[18]Given the large value of θ_1, the κ's are small: $\kappa(0) = -.0004$ and $\kappa(1) = -.0020$.
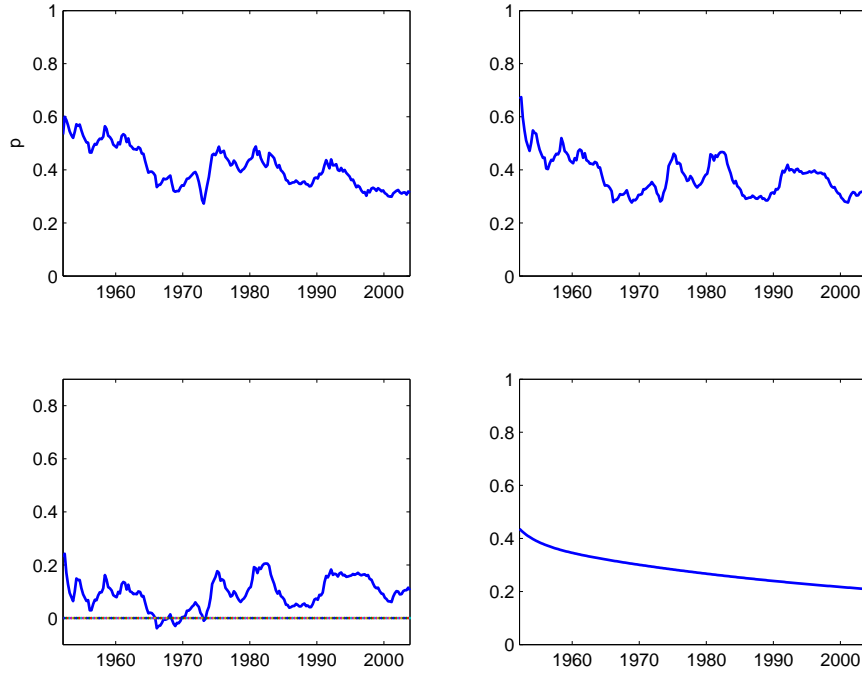
Figure 4: Contributions to slanting of model weights. Top left panel: $\tilde{p}_t - \check{p}_t$. Top right panel: $U[0, \check{\zeta}_t(0), \Sigma(0] - U[\iota, \check{\zeta}_t(1), \Sigma(1)]$. Lower left panel: $\lambda(0) \cdot \check{\zeta}_t(0) - \lambda(1) \cdot \check{\zeta}_t(0)$. Lower right panel: $-\frac{1}{2\theta_2}\lambda(0)'\Sigma_t(0)\lambda(0) + \frac{1}{2\theta_2}\lambda(1)'\Sigma_t(1)\lambda(1)$.

5.7 Differences in estimates of mean consumption growth across submodels

Figure 5 displays the representative consumer's estimates of the time invariant contribution to consumption growth, $\mu(\iota) = \zeta_2(\iota)$, conditional on models $\iota = 0$ and $\iota = 1$, respectively. For each model, the worst case estimate of $\zeta_2(\iota)$ is the dotted line. Notice how the worst case estimate is lower than the estimate under each approximating submodel. Figure 6 shows the posterior standard deviation of $\zeta_2(\iota)$ from the ordinary Kalman filter for each submodel. As this standard deviation converges to zero, less distortion of the mean of $\mu(\iota)$ comes from the T^2 operator. Because it is more difficult to estimate the time invariant component of consumption in the presence of the long run risk component, the posterior standard deviation is larger in the long-run risk submodel. For given θ_2, the smaller precision of the mean in the long run risk model accounts for the bigger downward distortion in the worst case unconditional mean in model $\iota = 1$.

5.8 Time series of the conditional stochastic discount factor

The bottom panel of figure 7 shows (32) along the sample path $\{s_t\}$ for the calibrated submodels with $\theta_1 = 20, \theta_2 = .2$. For our logarithmic preference specification, namely, $\beta \exp(s_{t+1})$, the top panel of figure 7 shows the stochastic discount factor without a concern for robustness. By using the same scales on the vertical axes, the top and bottom panels of
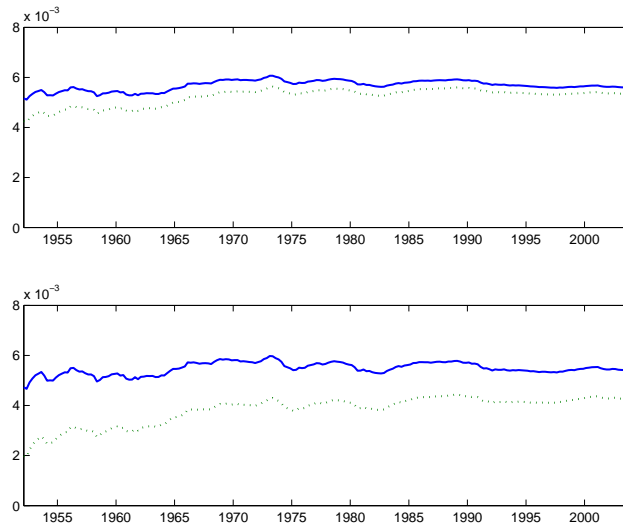
Figure 5: Estimates of $\zeta(0)$ (top panel) and of the second components of $\zeta(1)$ (bottom panel) for $\theta_1 = 20, \theta_2 = .2$. The worst case means are dotted.
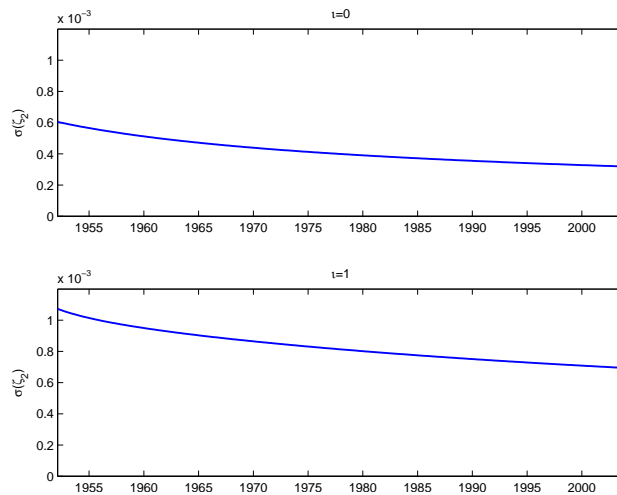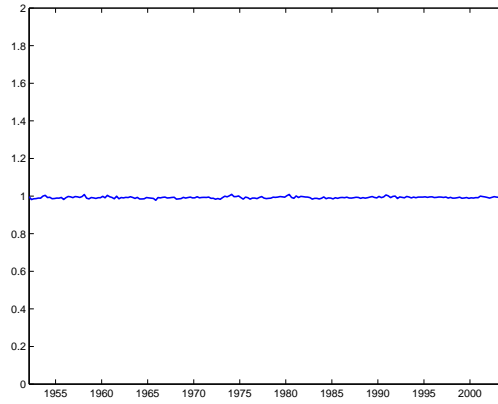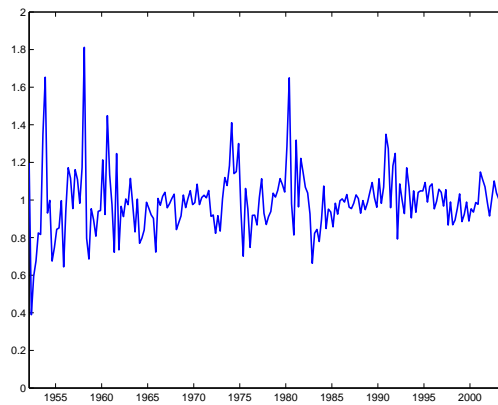

Figure 6: Standard deviation of $\zeta(0)$ (top panel) and of $\zeta_2(1)$ (bottom panel).

(a) $\beta c_t / c_{t+1}$

(b) $g(s_{t+1}|s^t)$

Figure 7: Top panel: the ordinary stochastic discount factor. Bottom panel: the multiplicative adjustment to the stochastic discount factor contributed by robustness when $\theta_1 = 20, \theta_2 = .2$.

figure 7 portray the additional volatility in the stochastic discount factor that is contributed by robustness when $\theta_1 = 20, \theta_2 = .2$. To appreciate why the enhanced volatility contributed by $g(s_{t+1}|s^t)$ is interesting, recall how Hansen and Jagannathan (1991) characterized the return heterogeneity puzzle as a shortfall of the volatility of the stochastic discount factor computed by plugging U.S. data into the formula for the s.d.f. implied by a time-separable CRRA preference specification and a reasonable discount factor when compared with the stochastic discount factor revealed by asset market data.

5.8.1 Time varying risk (or model uncertainty) premia

The continuous time limit in appendix B rationalizes the following approximation to risk-premia when $G(0) = G(1)$:[19]

$$\frac{\check{\mu}_t - \tilde{\mu}_t}{\sqrt{\Delta}} + \sqrt{\Delta}, \tag{37}$$

where Δ is a common value of $G(\iota)G(\iota)'$ for $\iota = 0, 1$, $\check{\mu}_t$ is the conditional mean of consumption growth under the approximating model, and $\tilde{\mu}$ is the conditional mean of consumption growth under the worst case model. In (37) the term $\sqrt{\Delta}$ is the standard contribution to the market price of risk coming from risk aversion with logarithmic preferences, while the term $\frac{\check{\mu}_t - \tilde{\mu}_t}{\sqrt{\Delta}}$ comes from our model uncertainty adjustments. In the continuous time limit, for both the distorted and benchmark models, the growth rate of consumption is conditionally normally distributed rather than being the mixture of normals that we obtained above. By a 'risk price', we mean the required mean compensation needed to compensate for the exposure to a one-standard deviation normal shock. More generally, in discrete time it is the ratio of the conditional standard deviation of the stochastic discount factor to its mean (or the maximum Sharpe ratio; see Hansen and Jagannathan (1991)).

In our model, the first component of the 'risk price' is more properly regarded as a 'model uncertainty price'. The second component $\sqrt{\Delta}$ is very small. Figure 8 shows that variations in the amount by which the model Bayesian model weights are twisted, displayed in the top left panel of figure 4, induce substantial fluctuations in the uncertainty premia. The model uncertainty premia in figure 8 show secular movements and short-run variations. There are time periods in which the long run risk model predicts that future discounted consumption growth will be relatively low. To illustrate this dependence, we depict the time series for $\lambda(1) \cdot \check{\zeta}_t(1) - \lambda(0) \cdot \check{\zeta}_t(0)$ in the bottom panel. Since the subjective discount factor is very close to unity, $\lambda(1) \cdot \check{\zeta}_t(1)$ is dominated by the estimate of the time invariant component $\zeta_2(1)$. Nevertheless, the presence of the long run risk component is important because it makes it challenging to disentangle the time invariant component $\zeta_2(1)$ from the long run risk component $\zeta_{1t}(1)$.

Figure 9 reports the uncertainty prices for our baseline specification $\theta_1 = 20, \theta_2 = .2$ as well as the two other specifications $\theta_1 = 20, \theta_2 = .1$ and $\theta_1 = .2, \theta_2 = .2$. The $\theta_1 = 20, \theta_2 = .1$ specification makes the distorted probabilities concentrate almost completely on model $\iota = 1$ (so that $\tilde{p}_t \approx 1$) Nevertheless the time series variation in the risk prices remains because the

---

[19]In the continuous time limit, for both the distorted and benchmark models, the growth rate of consumption is conditionally normally distributed rather than being the mixture of normals that we obtained above. Model averaging is used in computing the conditional means for the respective normal distributions.

nontrivial (and now enhanced) differences between the distorted and the original probabilities assigned to the each of the models. In particular, notice from formula (33) that when $\tilde{p}_t \approx 1$,

$$\check{\mu}_s - \tilde{\mu}_s \approx (\check{p}_t - 1)[D(1)\check{\zeta}_t(1) - D(0)\check{\zeta}_t(0)] - [G(1)w^*(1) + D(1)u^*(1)].$$

This formula tells us that even when the representative consumer assigns probability one to the long-run risk model (just as the investors in the model of Bansal and Yaron do), the representative consumer's doubts about the model misspecification still contribute variations in the market price of risk coming from variations over time in the (non-robust) Bayesian model weight $\check{p}_t$ and the estimates of $\zeta(0)$ and $\zeta(1)$ under the approximating model.

The $\theta_1 = .2, \theta_2 = .2$ specification focuses equally on all three forms of potential model misspecification and can be cast in terms of a Radon-Nikodym derivative that perturbs the density of $s_{t+1}|\mathcal{S}_t$, the entropy of that perturbation, and a single risk-sensitivity operator with parameter $\theta = .2$. The model then has a value function like that of model 1 of section 2.2, except that the state variable is now the one that comes from the innovation representation for the signal.[20]

Figure 10 again shows the market price of model uncertainty for our baseline $\theta_1 = 20, \theta_2 = .2$ model, as well as two smooth curves that depict the price of model uncertainty in the model that emerge from setting $\check{p}_0 = 1$ and $\theta_1 = 20, \theta_2 = .2$ and $\theta_1 = .2, \theta_2 = .2$. Setting $\check{p}_0 = 1$ means that the representative consumer commits to the $\iota = 1$ model. That renders our second application of the T^2 operator irrelevant, though the first remains. The first application of T^2 gives rise to the smoothly declining market prices of model uncertainty in these $\check{p}_t = 1$ models.

Figure 11 uses (33) to decompose the market price of model uncertainty for the special case $\theta_1 = 20, \theta_2 = .1$, implying that $\tilde{p}_t \approx 1$. Given the small value of θ_2, the slanted probabilities are essentially one: $\tilde{p}_t \approx 1$. Although the continuation values fluctuate, slanted probabilities remain concentrated on the long run risk model. This means that the model uncertainty premia are approximately

$$\frac{1}{\sqrt{GG'}}\Big((\check{p}_t - 1)[D(1)\check{\zeta}_t(1) - D(0)\check{\zeta}_t(0)] - [G(1)w^*(1) + D(1)u^*_t(1)]\Big),$$

which contain two distinct components: (i) a time varying backward-looking component that is induced by the learning from the signal history; and (ii) a smooth forward-looking component[21] induced by distortions $w^*(\iota)$ and $u^*(\iota)$ in the mean vectors for the underlying shocks and for the state estimates, respectively. Fluctuations of the first component reflect variation in the model probabilities calculated under the approximating model using Bayes law. They respond to newly arrived signals, as do differences in the conditional means for consumption growth across models. The bottom two panels of figure 11 depict the two components. They are highly correlated because the conditional means enter directly into the construction of the likelihood function.

---

[20]See Hansen and Sargent (2006a) for a presentation of how a model with $\theta_1 = \theta_2$ can be represented in terms of single risk-sensitivity operator.

[21]See the formulas for $w^*(\iota), \lambda(\iota), u^*_t(\iota)$ given in formulas (23),(25), (26), in particular, formula (25) for $\lambda(\iota)$, to understand why we say 'forward looking'.
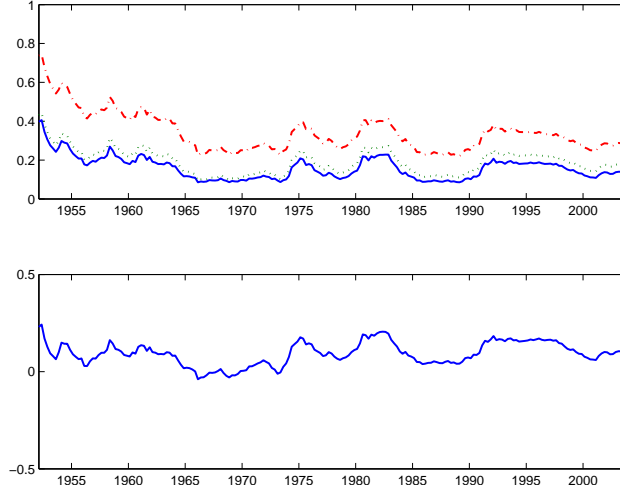
Figure 8: Model uncertainty premia. Top panel: $\frac{\check{\mu}_t - \tilde{\mu}_t}{\sqrt{\Delta}}$; the solid line is for $\theta_1 = 20, \theta_2 = .2$; the dashed-dotted line is for $\theta_1 = 20, \theta_2 = .1$, and the dotted line is for $\theta_1 = .2, \theta_2 = .2$. Bottom panel: $\lambda(1) \cdot \check{\zeta}_t(1) - \lambda(0) \cdot \check{\zeta}_t(0)$.
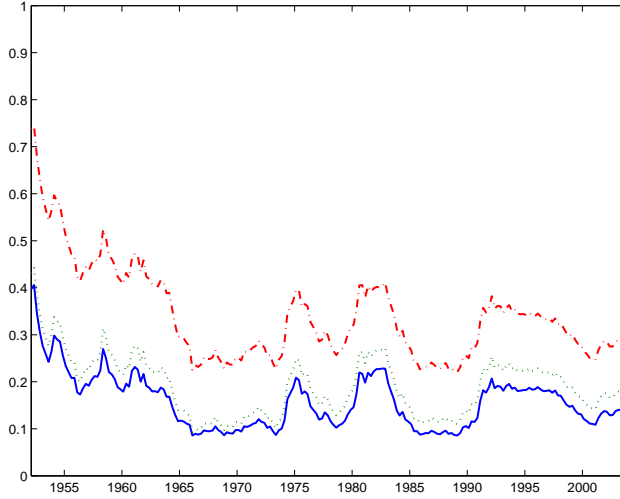

Figure 9: Model uncertainty premia for different configurations. The solid line is the $\theta_1 = 20$, $\theta_2 = .2$, our benchmark specification. The dotted line slightly above the line for the benchmark model assumes that $\theta_1 = \theta_2 = .2$. The dash-dotted line assumes that $\theta_1 = 20, \theta_2 = .1$; for this setting $\tilde{p}_t \approx 1$.
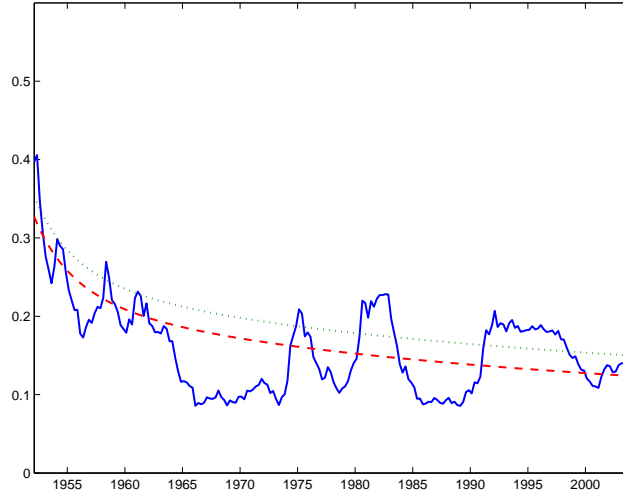
Figure 10: Model uncertainty premia for different configurations. The solid line is for our benchmark specification with $\theta_1 = 20$, $\theta_2 = .2$. The dashed line is for $\check{p}_0 = 1$, so that the representative consumer commits to model $\iota = 1$, and $\theta_1 = 20$, $\theta_2 = .2$. The dotted line is for $\check{p}_0 = 1$, so that the representative consumer commits to model $\iota = 1$, and $\theta_1 = \theta_2 = .2$.
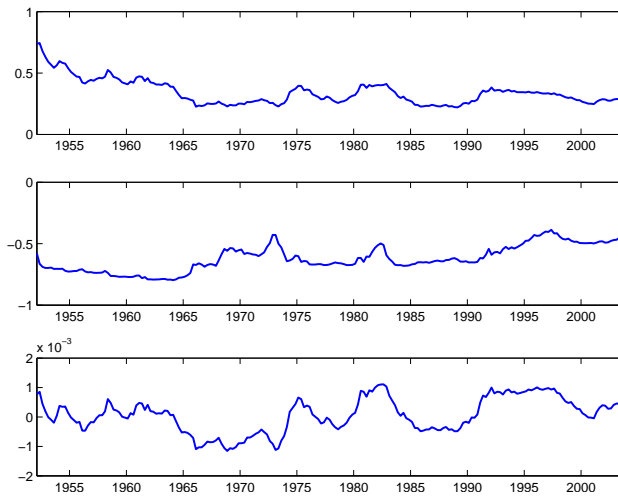

Figure 11: Decomposition of uncertainty premium when $\tilde{p}_t \approx 1$. Top panel: uncertainty premium, with $\theta_1 = 20$, $\theta_2 = .1$. Middle panel: $(\check{p}_t - 1)$. Bottom panel: $[D(0)\check{\zeta}_t(0) - D(1)\check{\zeta}_t(1)]$.
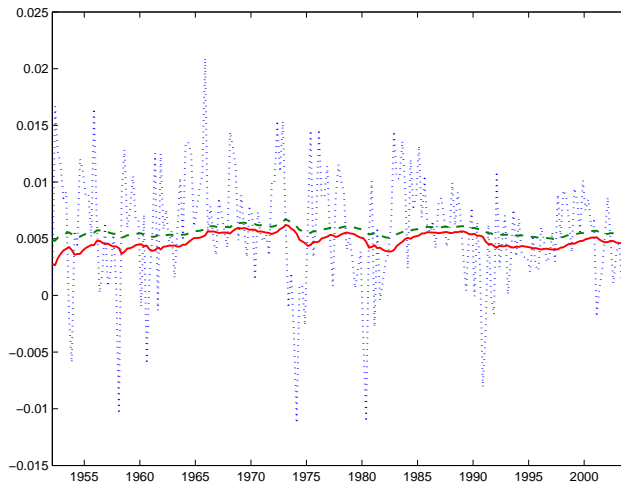
Figure 12: Consumption growth signal and alternative conditional means. The dotted line is the consumption growth rate, the dashed line is the model weighted conditional mean, and the solid line is the distorted weighted conditional mean. The distorted weighted conditional mean is computed using $\theta_1 = 20$ and $\theta_2 = .2$

Our model uncertainty premia are given by differences between the Bayesian model average mean and a distorted counterpart, divided by the conditional volatility of consumption. It is interesting to compare the differences in means relative to fluctuations in the signal, namely, consumption growth. As emphasized by Anderson et al. (2003), difficulties in estimating local means can cause statistical ambiguity that will make the representative consumer reluctant to reject our worst case calculations. Anderson et al. (2003) use formal statistical discrepancy measures as a guide as to how much concern about model misspecification we should be impute to the representative consumer. Their analysis abstracts from learning, however. While we introduce explicit forms of learning, a concern about model misspecification remains central to our analysis. For instance, through our θ_2 parameter, we capture the extent to which our representative consumer is not fully confident about his posterior probabilities. Thus, our representative consumer's beliefs are fragile. While our choices of θ's in this paper are merely illustrative, figure 12 reveals that they imply distortions in the conditional mean averaged across models that are small when compared to the movements in the signal.

6   Concluding remarks

The asymmetric responses of uncertainty premia to good and back shocks to consumption growth rates comes from the way that our representative consumer copes with uncertainty about what probabilities to attach to two competing candidate submodels for consumption growth. He attaches positive probability to a submodel with statistically subtle persistence in consumption growth, but also believes that the data might be generated by a submodel asserting iid consumption growth rates. The asymmetrical response of model uncertainty premia to consumption growth shocks depends on (1) how the consumer's wish to be robust

30

with respect to the probability that he attaches to submodels causes him to calculate worst case probabilities that depend on value functions for both submodels, and (2) how the value functions conditioned on the respective submodels respond to shocks in ways that bring the value functions closer together after positive consumption growth shocks and push them farther apart after adverse consumption growth shocks. Thus, that the submodel with persistent growth confronts the consumer with more overall risk in future consumption gets encoded in a smaller constant term in our formula for the consumer's value function. This difference in constant terms in the value functions for the submodels with and without long-run consumption risk sets the stage for an asymmetric response of uncertainty premia to consumption growth shocks. Consecutive periods of higher than average consumption growth raise the probability the consumer attaches to the submodel with persistent consumption growth relative to that of the iid consumption growth model. Although the long-run risk submodel implies more overall consumption risk, persistence of consumption growth under this submodel implies that consumption growth is expected to remain higher than average for many future periods. This pushes the continuation values associated with the two submodels closer together than they are when consumption growth rates have recently been lower than average. These continuation values determine the worst-case probabilities that the representative consumer attaches to the submodels. That the continuation values for the two models become farther apart after a string of negative consumption growth shocks implies that our cautious consumer slants probability more towards the pessimistic long-run risk model when recent observations of consumption growth have been lower than average than when these observed growth rates are higher than average. This behavior of probability slanting over time is how learning in the presence of model uncertainty induces time variation in uncertainty premia.

The model uncertainty premia that emerge from our model combine (1) the same constant forward-looking contribution w^* that was featured in earlier work without learning by Tallarini (2000) and Bansal and Yaron (2004), (2) an additional smoothly decreasing components $u_t^*(\iota)$ that comes from learning about uncertain parameter values within submodels, and (3) the potentially volatile time varying contribution that we have just discussed and that is caused by robust learning about the distribution over submodels.[22]

Our mechanism for producing time varying risk premia differs from other approaches. For instance, Campbell and Cochrane (1999) induce secular movements in risk premia that are backward looking because a social externality depends on current and past average consumption. To generate variation in risk premia, Bansal and Yaron (2004) assume stochastic volatility in consumption.[23]

Our analysis features the role of learning on local or one-period uncertainty prices. A study of the consequences robust learning and model selection for multi-period uncertainty prices is a natural next step. This will prompt us to explore two changes in our environment: a) the introduction of stochastically growing financial cash flows with alternative exposures to uncertainty, and b) the relaxation of the unitary elasticity of substitution in preferences. While a unitary elasticity of substitution simplifies our calculations, it implies that the

---

[22]The implied risk free rate volatility is modest. The average risk free rate is about 2.5 percent with a standard deviation of about .5.

[23]Our interest in learning and time series variation in the uncertainty premium differentiates us from Weitzman (2005) and Jobert et al. (2006), who focus on long run averages.

ratio of consumption to wealth is constant. Consumption claims, however, have no obvious counterpart in financial data leading to the study of a). The unitary elasticity referred to in b) can be relaxed at the cost of using additional numerical techniques in approximation.

Our environment features two competing models. Implications like ours could also arise in setups in which a hidden state moves very infrequently, as in the regime shift models of advocated by Wonham (1964), Hamilton (1989), and others. Indeed, our analysis of model averaging can be viewed as a limiting case of a stochastic regime model with two states. While our example is highly stylized, we can imagine a variety of environments in which learning about low frequency phenomenon is challenging, especially when investors are not fully confident in their probability assessments. In this regard, Hansen et al. (2006) show that while long run risk components have important quantitative impacts on low frequency implications of stochastic discount factors, it is also statistically challenging to measure those components.

We have indicated how our results depend on the new parameter that we have introduced, θ_2. Among several important things that remain to be done with this framework, it will be useful to think about how to calibrate θ_2. We can do that by modifying and extending the logic of the detection error probabilities used by Anderson et al. (2003).

A An alternative recursion

To understand the relationship of our work to the model of Croce et al. (2006), we note that when $\theta_1 = \theta_2$, the mean distortion for the signal depends directly on the covariance of the signal with the hidden state, $G(\iota)C(\iota)' + G(\iota)\Sigma_t(\iota)A(\iota)'$, and on the covariance matrix of the signal, $G(\iota)G(\iota)' + D(\iota)\Sigma_t(\iota)D(\iota)'$, both conditioned on the signal history. But when θ_1 and θ_2 are distinct, as they are in most of this paper, the decomposition of these objects into the contributions from the underlying shock vector ϵ_{t+1} and $\zeta_t(\iota) - \check{\zeta}_t(\iota)$ matters.

Hansen and Sargent (2006a) consider another formulation that in the present context would translate into a value function $Q(q(z))$ that satisfies the following recursion that uses the composite operator $\mathsf{T}^2 \circ \mathsf{T}^1$:

$$Q(q, c) = (1 - \beta)c + \mathsf{T}^2\mathsf{T}^1[\beta Q(q^*, c^*)] \tag{38}$$

where q consists of the posterior density of $\zeta(0)$, the density of $\zeta(1)$ and the probabilities assigned to each value of ι and q^* is the counterpart for the next time period and is based on an additional observation of the signal. Here the continuation value $Q(q, c)$ depends on the information available to the consumer, but the sequential application of the two operators still decomposes the transition into uncertainty over tomorrow's signal into uncertainty conditioned on the current hidden state and uncertainty over beliefs about that state.

If we endow the representative consumer with knowledge of ι, the value function $Q(q, c)$ coincides with $U[\iota, \check{\zeta}(\iota), \Sigma(\iota)] + c$, except it has a different constant term. Since the constant does not affect them, the worst-case models remain the same as the ones computed in section 4.3.1, as would the asset pricing implications. These outcomes are special to the log-linear preferences and laws of motion and the absence of uncertainty about ι. If there is uncertainty about ι, the value function (38) would have to be solved numerically and would differ from (27) in more ways than just the constant term. The asset pricing implications would also differ.

B Continuous time version

Consider the following continuous time model:

$$
\begin{aligned}
d\zeta_t(\iota) &= A(\iota)\zeta_t(\iota)dt + C(\iota)dB_t \\
ds_t &= D(\iota)\zeta_t dt + G(\iota)dB_t
\end{aligned}
$$

where B_t is a multivariate standard Brownian motion. Assume that $G(\iota)G(\iota)' = \Delta$ is independent of ι, a good approximation for our application. Application of the Kalman filter implies:

$$d\check{\zeta}_t(\iota) = A(\iota)\check{\zeta}_t(\iota) + K_t(\iota)[ds_t - D(\iota)\check{\zeta}_t(\iota)]$$

where

$$K_t(\iota) = [C(\iota)G(\iota)' + A(\iota)\Sigma_t(\iota)D(\iota)']\,[D(\iota)\Sigma_t(\iota)D(\iota)' + \Delta]^{-1}$$

and

$$d\Sigma_t(\iota) = A(\iota)\Sigma_t(\iota)A(\iota)'-$$

$$[C(\iota)G(\iota)' + A(\iota)\Sigma_t(\iota)D(\iota)'][D(\iota)\Sigma_t(\iota)D(\iota)' + \Delta]^{-1}[C(\iota)G(\iota)' + A(\iota)\Sigma_t(\iota)D(\iota)']'.$$

Construct the innovation process:

$$d\breve{B}_t = ds_t - \breve{\mu}_t dt$$

where $\breve{p}_t = E(\iota|\mathcal{S}_t)$ and

$$\breve{\mu}_t \doteq [\breve{p}_t D(1)\breve{\zeta}_t(1) + (1-\breve{p}_t)D(0)\breve{\zeta}_t(0)].$$

To determine the evolution of $\breve{p}_t$, first construct

$$L_T(\iota) = \exp\left[\int_0^T \breve{\zeta}_t(\iota)'D(\iota)'\Delta^{-1}ds_t - \frac{1}{2}\int_0^T D(\iota)\breve{\zeta}_t(\iota)'\Delta^{-1}D(\iota)'\breve{\zeta}_t(\iota)dt\right],$$

which gives the likelihood as a function of ι up to scale. Then

$$
\begin{aligned}
\frac{dL_t(\iota)}{L_t(\iota)} &= \breve{\zeta}_t(\iota)'D(\iota)'\Delta^{-1}\left[\breve{p}_t D(1)\breve{\zeta}_t(1) + (1-\breve{p}_t)D(0)\breve{\zeta}_t(0)\right]dt + \\
&\quad \breve{\zeta}_t(\iota)'D(\iota)'\Delta^{-1}d\breve{B}_t
\end{aligned}
$$

The posterior probabilities of $\iota = 0,1$ are proportional to $(1-\breve{p}_0)L_t(0), \breve{p}_0 L_t(1)$. Each probability is a martingale relative to the filtration $\{\mathcal{S}_t, t \geq 0\}$. In particular,

$$d\breve{p}_t = \breve{p}_t(1-\breve{p}_t)[\breve{\zeta}_t(1)'D(1)' - \breve{\zeta}_t(0)'D(0)']\Delta^{-1}d\breve{B}_t.$$

To compute the distorted distribution, first form:

$$
\begin{aligned}
\lambda(\iota) &= [\delta I - A(\iota)']^{-1}D(\iota)' \\
\kappa(\iota) &= -\frac{1}{2\delta\theta_1}|\lambda(\iota)'C(\iota) + G(\iota)|^2
\end{aligned}
$$

where $\delta = \log\beta$. Under model ι, ds_t has drift:

$$D(\iota)\breve{\zeta}_t(\iota) - \frac{1}{\theta_1}G(\iota)\left[C(\iota)'\lambda(\iota) + G(\iota)'\right] - \frac{1}{\theta_2}D(\iota)\Sigma_t\lambda(\iota)$$

and instantaneous covariance matrix Δ. For each model ι, compute the value functions

$$U[\iota, \breve{\zeta}(\iota), \Sigma(\iota)] = \lambda(\iota) \cdot \breve{\zeta}_t(\iota) + \kappa(\iota) - \frac{1}{2\theta_2}\lambda(\iota)'\Sigma_t(\iota)\lambda(\iota).$$

Then as in discrete time, the distorted probabilities $(1-\tilde{p}_t)$ and $\tilde{p}_t$, respectively, are proportional to:

$$\exp\left(-\frac{U[0, \breve{\zeta}_t(0), \Sigma_t(0)]}{\theta_2}\right)(1-\breve{p}_t), \quad \text{and} \quad \exp\left(-\frac{U[1, \breve{\zeta}_t(1), \Sigma_t(1)]}{\theta_2}\right)\breve{p}_t.$$

Averaging over models, we find that the distorted drift is:

$$\tilde{\mu}_t = \tilde{p}_t\left(D(1)\breve{\zeta}_t(1) - \frac{1}{\theta_1}G(1)\left[C(1)'\lambda(1) + G(1)'\right] - \frac{1}{\theta_2}D(1)\Sigma_t\lambda(1)\right) +$$

$$(1 - \tilde{p}_t) \left(D(0)\check{\zeta}_t(0) - \frac{1}{\theta_1} G(0) \left[C(0)'\lambda(0) + G(0)' \right] - \frac{1}{\theta_2} D(0)\Sigma_t \lambda(0) \right)$$

The drift distortion induced by robustness is $\tilde{\mu}_t - \check{\mu}_t$. For the scalar signal example studied in the paper, the model uncertainty premia implied by this distortion is:

$$\frac{\check{\mu}_t - \tilde{\mu}_t}{\sqrt{\Delta}}.$$

The plots in the paper are given by forming the negative of the discrete mean distortion and dividing by Δ. While the Δ in the discrete time specification is model dependent, the difference is very small.

References

Ai, Hengjie. 2006. Incomplete Information and Equity Premium in Production Economy. Unpublished.

Anderson, E., L.P. Hansen, and T.J. Sargent. 2003. A Quartet of Semigroups for Model Specification, Robustness, Prices of Risk, and Model Detection. *Journal of the European Economic Association* 1 (1):68–123.

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

Barillas, Francisco, Lars Peter Hansen, and Thomas J. Sargent. 2006. Reinterpreting a graph (and a parameter) of Tallarini. Unpublished paper, University of Chicago and New York University.

Brennan, M. J. and Y. Zia. 2001. Stock Price Volatility and Equity Premium. *Journal of Monetary Economics* 47:249–283.

Cagetti, Marco, Lars Peter Hansen, Thomas J. Sargent, and Noah Williams. 2002. Robustness and Pricing with Uncertain Growth. *Review of Financial Studies* 15 (2):363–404.

Campbell, John Y. and John H. Cochrane. 1999. By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy* 107 (2):205–251.

Cogley, Timothy, Riccardo Colacito, Lars Hansen, and Thomas Sargent. 2005. Robustness and U.S. Monetary Policy Experimentation. Unpublished.

Croce, Mariano M., Martin Lettau, and Sydney C. Ludvigson. 2006. Investor Information, Long-Run Risk, and the Duration of Risky Cash Flows. Unpublished.

David, A. 1997. Fluctuating Confidence in Stock Markets. *Journal of Financial and Quantitative Analysis* 32:457–462.

Detemple, J. 1986. Asset Pricing in a Production Economy with Incomplete Information. *Journal of Finance* 41:383–390.

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

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

Hamilton, J. D. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57:357–384.

Hansen, L. P., J. Heaton, and N. Li. 2006. Consumption Strikes Back: Measuring Long Run Risk. University of Chicago.

Hansen, Lars Peter. 2007. Beliefs, Doubts and Learning: The Valuation of Macroeconomic Risk. *American Economic Review, Papers and Proceedings* XXX–XXXX.

Hansen, Lars Peter and Ravi Jagannathan. 1991. Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy* 99:225–262.

Hansen, Lars Peter and Thomas Sargent. 1995. Discounted Linear Exponential Quadratic Gaussian Control. *IEEE Transactions on Automatic Control* 40 (5):968–971.

Hansen, Lars Peter and Thomas J. Sargent. 2006a. Recursive Robust Estimation and Control without Commitment. Unpublished.

———. 2006b. Robust Estimation and Control for LQ Gaussian problems without Commitment. Unpublished.

Hansen, Lars Peter, Thomas Sargent, and Thomas Tallarini. 1999. Robust Permanent Income and Pricing. *Review of Economic Studies* 66:873–907.

Hansen, Lars Peter, Thomas J. Sargent, and Neng E. Wang. 2002. Robust Permanent Income and Pricing with Filtering. *Macroeconomic Dynamics* 6:40–84.

Harrsion, J. Michael and David M. Kreps. 1978. Speculative Investor Behavior in a Market with Heterogeneous Expectations. *Quarterly Journal of Economics* 92 (2):323–336.

Jobert, A., A. Platania, and L. C. G. Rogers. 2006. A Bayesian Solution to the Equity Premium Puzzle. Statistical Laboratory, University of Cambridge.

Klibanoff, P., M. Marinacci, and S. Mukerji. 2003. A Smooth Model of Decision Making Under Ambiguity. Northwestern University.

Kreps, D. M. and E. L. Porteus. 1978. Temporal Resolution of Uncertainty and Dynamic Choice. *Econometrica* 46:185–200.

Scheinkman, Jose A. and Wei Xiong. 2003. Overconfidence and Speculative Bubbles. *Journal of Political Economy* 111 (6):1183–1220.

Segal, U. 1990. Two-Stage Lotteries without the Reduction Axiom. *Econometrica* 58:349–377.

Tallarini, Thomas D. 2000. Risk-sensitive real business cycles. *Journal of Monetary Economics* 45 (3):507–532.

Veronesi, P. 2000. How Does Information Quality Affect Stock Market Returns. *Journal of Finance* 55:807–837.

Weitzman, M. 2005. A Unified Theory of Equity Puzzles. Harvard University.

Wonham, W. J. 1964. Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering. *Siam Journal on Control* 2:347–368.