

# Fragile Beliefs and the Price of Uncertainty

Lars Peter Hansen\*

Thomas J. Sargent<sup>‡</sup>

March 5, 2009

## Abstract

A representative consumer uses Bayes' law to learn about parameters and to construct probabilities with which to perform ongoing model averaging. The arrival of signals induces the consumer to alter his posterior distribution over parameters and models. The consumer copes with specification doubts by slanting probabilities pessimistically. One of his models puts long-run risks in consumption growth. The pessimistic probabilities slant toward this model and contribute a counter-cyclical and signal-history-dependent component to prices of risk.

KEY WORDS: Learning, Bayes' law, robustness, risk-sensitivity, pessimism, prices of risk.

---

\*University of Chicago. Email: l-hansen@uchicago.edu.

<sup>†</sup>New York University and Hoover Institution. Email: ts43@nyu.edu.

<sup>‡</sup>We thank Gadi Barlevy, Alberto Bisin, Jaroslav Borovicka, Riccardo Colacito, Mark Gertler, Anastasios Karantounias, Ricardo Mayer, Tomasz Piskorski, Grace Tsiang, Gianluca Violante, and Amir Yaron for helpful comments on earlier drafts of this paper. We thank Francisco Barillas, Ricardo Mayer, and Leandro Nascimento for excellence in executing the computations. We thank the National Science Foundation for research support under separate grants to Hansen and Sargent.

Le doute n'est pas une condition agréable, mais la certitude est absurde.<sup>1</sup>  
Voltaire 1767.

# 1 Introduction

A pessimist thinks that good news is temporary and that bad news endures. This paper describes how a representative consumer's model selection problem and fear of model misspecification foster pessimism that puts countercyclical model uncertainty premia into risk prices.

## 1.1 Doubts promote fragile beliefs

Our representative consumer values consumption streams according to the multiplier preferences that Hansen and Sargent (2001) use to represent aversion to model uncertainty.<sup>2</sup> Following Hansen and Sargent (2007), an iterated application of risk-sensitivity operators focuses the representative consumer's ambiguity on particular parameters and model selection.<sup>3</sup> *Ex post*, the consumer acts 'as if' he believes a probability measure that his malevolent alter ego has twisted pessimistically relative to his approximating model. By 'fragile beliefs' we refer to the responsiveness of pessimistic probabilities to the arrival of news, as determined by the state-dependent value functions that define what the consumer is pessimistic *about*.<sup>4</sup> Our representative consumer's reluctance to trust his model adds 'model uncertainty premia' to prices of risk. They are time-dependent and state-dependent, in contrast to the constant uncertainty premia analyzed by Hansen et al. (1999) and Anderson et al. (2003).

---

<sup>1</sup>Doubt is not a pleasant condition, but certainty is absurd.

<sup>2</sup>The relationship of the multiplier preferences of Hansen and Sargent (2001) to the max-min expected utility preferences of Gilboa and Schmeidler (1989) are analyzed by Hansen et al. (2006), Maccheroni et al. (2006a,b), Cerreia et al. (2008), and Strzalecki (2008).

<sup>3</sup>Sometimes the literature calls this 'structured uncertainty'.

<sup>4</sup>Harrison and Kreps (1978) and Scheinkman and Xiong (2003) explore another setting in which difficult to detect departures from rational expectations lead to interesting asset price dynamics that cannot occur under rational expectations.

## 1.2 Fragile expectations as sources of time-varying risk premia

A hidden Markov model for consumption growth confronts a representative consumer with ongoing model selection and parameter estimation problems. Our representative consumer wants to know components of a hidden state vector, some that stand for unknown parameters within a model and others that index models. A probability distribution over that hidden state vector becomes part of the state that enters the representative consumer's value function. Bayes' law describes its motion over time. The representative consumer copes with model uncertainty by slanting probabilities towards a model that has the lowest utility. We show how variations over time in the probabilities attached to models and other state variables put volatility into the model uncertainty premia.

## 1.3 Key components

In addition to a risk sensitivity operator from Hansen et al. (1999) and Tallarini (2000) that adjusts for uncertain dynamics of known states, another one from Hansen and Sargent (2007) adjusts the probability distribution of hidden Markov states for model uncertainty.<sup>5</sup> We interpret both risk-sensitivity operators as capturing the representative consumer's concerns about robustness.<sup>6</sup>

Our representative consumer assigns positive probabilities to two models whose fits make them indistinguishable for our data on per capita U.S. consumption expenditures on nondurables and services from 1948II-2008III. In one model, consumption growth rates are nearly i.i.d., and in the other there is a highly persistent component of consumption growth rate, as in the long-run risk model of Bansal and Yaron (2004). But the consumer doubts the model-mixing probabilities as well as the specification of each model. In contrast, Bansal and Yaron assume that the representative consumer assigns probability one to the long-run risk model even though sample evidence

---

<sup>5</sup>This second risk-sensitivity operator expresses what Klibanoff et al. (2005, 2008) call smooth ambiguity and what other researchers call 'structured' model uncertainty. As an example of a different approach to learning in the presence of model ambiguity, Epstein and Schneider (2008) apply their recursive multiple priors model to study the response of asset prices to signals when investors are uncertain about a noise variance that influences Bayesian updating.

<sup>6</sup>Tallarini adopted an interpretation in terms of enhanced risk aversion. See Barillas et al. (2008) for the relationship between these interpretations.

is indecisive in selecting between them.<sup>7</sup> Our framework explains why a consumer might 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 statistically.

## 1.4 Organization

We proceed as follows. After section 2 sets out a framework for pricing risks expressed in a vector Brownian motion  $w_t$ , section 3 describes a hidden Markov model and three successively smaller information sets (full information, unknown states, and unknown states and unknown model) together with the three innovations (or news) processes given by the increments to  $W_t(\iota)$ ,  $\bar{W}_t(\iota)$  and  $\bar{W}_t$  that are implied by these three information structures. Section 4 then uses these three information specifications and the associated  $dW_t(\iota)$ ,  $d\bar{W}_t(\iota)$ ,  $d\bar{W}_t$ , respectively, as the risks to be priced without model uncertainty. We construct these section 4 risk prices under the information assumptions ordinarily used in finance and macroeconomics. Section 5 proposes a new perspective on Bayesian learning by pricing each of the risks  $dW_t(\iota)$ ,  $d\bar{W}_t(\iota)$  and  $d\bar{W}_t$  under the full information set. Section 6 describes contributions to risk prices coming from uncertainty about distributions under each of our three information structures. Uncertainty about shock distributions with known states contributes a constant uncertainty premium, while uncertainty about unknown states contributes a time-dependent one and uncertainty about models contributes a state-dependent one. Section 7 presents an empirical example designed to highlight the mechanism through which the state-dependent uncertainty premia give rise to countercyclical prices of risk. Appendix A describes how we use detection error probabilities to calibrate the representative consumer's concerns about model misspecification, while appendix B proliferates models as part of a robustness exercise designed to refine our understanding of the forces that produce countercyclical risk prices.

## 2 Stochastic discounting and risks

Let  $\{S_t\}$  be a stochastic discount factor process that, in conjunction with an expectation operator, assigns date 0 risk-adjusted prices to payoffs at date  $t$ . Trading at

---

<sup>7</sup>Bansal and Yaron (2004) incorporate other features in their specification of consumption dynamics, including stochastic volatility, and they adopt a recursive utility specification with an intertemporal elasticity of substitution greater than 1.

intermediate dates implies that  $\frac{S_{t+\tau}}{S_t}$  is the  $\tau$ -period stochastic discount factor for pricing at date  $t$ . Let  $\{W_t\}$  be a vector Brownian motion innovation process where the increment  $dW_t$  represents new information flowing to consumers at date  $t$ . Synthesize a cumulative time  $t$  payoff as

$$\log Q_t(\alpha) = \alpha \cdot (W_t - W_0) - \frac{t}{2}|\alpha|^2.$$

By subtracting  $\frac{t}{2}|\alpha|^2$ , we have constructed the payoff to be a martingale with unit expectation. By altering the vector  $\alpha$ , we change the exposure of the payoff to components of  $W_t$ . At date  $t$ , we price the payoff  $\frac{Q_{t+\tau}(\alpha)}{Q_t(\alpha)}$  as

$$P_{t,\tau}(\alpha) = E \left[ \frac{S_{t+\tau} Q_{t+\tau}(\alpha)}{S_t Q_t(\alpha)} \middle| \mathcal{Y}_t \right]. \quad (1)$$

The vector of (growth-rate) risk prices for horizon  $\tau$  is given by the price “elasticity”

$$\pi_{t,\tau} = -\frac{\partial}{\partial \alpha} \frac{1}{\tau} \log P_{t,\tau}(\alpha) \Big|_{\alpha=\alpha_o}, \quad (2)$$

where we have scaled by the payoff horizon  $\tau$  for comparability.<sup>8</sup> Since we scaled the payoffs to have a unit expected payoff,  $-\frac{1}{\tau} \log P_{t,\tau}$  is the logarithm of an expected return adjusted for the payoff horizon. In log-normal models, this derivative is independent of  $\alpha_o$ . This is true more generally when the investment horizon shrinks to zero.<sup>9</sup>

The vector of local risk prices is given by the limit

$$\pi_t = -\lim_{\tau \downarrow 0} \frac{\partial}{\partial \alpha} \frac{1}{\tau} \log P_{t,\tau}. \quad (3)$$

It gives the local compensation for exposure to shocks expressed as an increase in the conditional mean return. In conjunction with an instantaneous risk-free rate, local risk prices are key building blocks for pricing assets (*e.g.*, Duffie (2001, pp. 111-114)). Local prices can be compounded to construct the asset prices for arbitrary payoff intervals  $\tau$  using the dynamics of the underlying state variables.

We can exploit local normality to obtain a simple characterization of the slope of

---

<sup>8</sup>The negative sign reflects that the consumer dislikes risk.

<sup>9</sup>Here we are following Hansen and Scheinkman (2009) and Hansen (2008) in constructing a term structure of prices of growth-rate risk.

the mean-standard deviation frontier and thereby reproduce a classical result from finance. The slope of the efficient segment of the mean-standard deviation frontier is the optimized value of the objective function

$$\max_{\alpha, \alpha \cdot \alpha = 1} \alpha \cdot \pi_t$$

where the constraint imposes a unit local variance. The solution is  $\alpha_t^* = \frac{\pi_t}{|\pi_t|}$  with the optimized local mean being

$$\alpha_t^* \cdot \pi_t = \frac{\pi_t \cdot \pi_t}{|\pi_t|} = |\pi_t|. \quad (4)$$

In this local normal environment, the Hansen and Jagannathan (1991) analysis simplifies to comparing an observed mean-standard deviation frontier to the magnitude of the risk price vector implied by alternative models.

In the power utility model,

$$\frac{S_{t+\tau}}{S_t} = \exp(-\delta) \exp[-\gamma(\log C_{t+\tau} - \log C_t)],$$

where the growth rate of log consumption is  $\log C_{t+\tau} - \log C_t$ . Here the vector  $\pi_t$  of local risk prices is the vector of exposures of  $-d \log S_t = \gamma d \log C_t$  to the Brownian increment vector  $dW_t$ .

We use models of Bayesian learning to create alternative specifications of  $dW_t$  and information sets with respect to which the mathematical expectation in (1) is evaluated.

## 2.1 Learning and asset prices

We assume a hidden Markov model in which  $X_t(\iota)$  is a hidden state vector for an unknown model indexed by  $\iota$ ,  $Y_t^{t+\tau}$  is the stochastic process of signals between date  $t$  and  $t + \tau$ , and  $\mathcal{Y}_t$  is a conditioning information set generated by the history of signals up until time  $t$ . Lower case letters denote potential values that can be realized. That is,  $y$  is a possible realized path for the signal process  $Y_t^{t+\tau}$  and  $x$  is a possible realization of the date  $t$  state vector  $X_t(\iota)$  for any model  $\iota$ . The hidden Markov structure induces

probability densities  $f(y|\iota, x)$ ,  $g(x|\iota, \mathcal{Y}_t)$ ,  $h(\iota|\mathcal{Y}_t)$ , and  $\bar{f}(y|\mathcal{Y}_t)$ .<sup>10</sup> Evidently,

$$\bar{f}(y|\mathcal{Y}_t) = \int \left[ \int f(y|\iota, x) g(x|\iota, \mathcal{Y}_t) dx \right] h(\iota|\mathcal{Y}_t) d\iota. \quad (5)$$

For convenience, let

$$Z_{t+\tau}(\alpha) = \frac{S_{t+\tau} Q_{t+\tau}(\alpha)}{S_t Q_t(\alpha)}.$$

In our construction under limited information,  $Z_{t+\tau}(\alpha)$  can be expressed as a function of  $Y_t^{t+\tau}$  and hence we may express the asset price

$$P_{t,\tau}(\alpha) = E[Z_{t+\tau}(\alpha)|\mathcal{Y}_t]$$

as an integral against the density  $\bar{f}$ .

To express the price in another way that will be useful to us, we first use density  $f$  to construct  $E[Z_{t+\tau}(\alpha)|X_t(\iota) = x, \iota]$  and then write

$$P_{t,\tau}(\alpha) = \int \int E[Z_{t+\tau}(\alpha)|X_t(\iota) = x, \iota] \underbrace{g(x|\iota, \mathcal{Y}_t) dx}_{\substack{\uparrow \\ \text{unknown} \\ \text{state}}} \underbrace{h(\iota|\mathcal{Y}_t) d\iota}_{\substack{\uparrow \\ \text{unknown} \\ \text{model}}}.$$

This decomposition helps us understand how our paper relates to earlier asset pricing papers including, for example, Detemple (1986), David (1997), Veronesi (2000), Brennan and Xia (2001), Ai (2006), and Croce et al. (2008),<sup>11</sup> that use learning about a hidden state to generate an exogenous process for distributions of future signals conditional on past signals as an input into a consumption based asset pricing model. After constructing  $\bar{f}(y|\mathcal{Y}_t)$ , decision making and asset pricing proceeds as in standard asset pricing models without learning. Therefore, the asset pricing implications of

<sup>10</sup>Densities are always expressed relative to a reference measure. In the case of  $Y_t^{t+\tau}$ , the reference measure is a measure over the space of continuous functions defined on the interval  $[t, t + \tau]$ .

<sup>11</sup>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. (2008) apply the framework of Hansen and Sargent (2007) in a setting where decisions affect future probabilities of hidden states and experimentation is active. The papers just cited price risks under the same information structure that is used to generate the risks being priced. In section 5, we offer an interpretation of some other papers (e.g., Bossaerts (2002, 2004) and Cogley and Sargent (2008)) that study the effects of agents' Bayesian learning on pricing risks generated by limited information sets from the point of view of an outside econometrician who has a *larger* information set.

such learning models depend only on  $\bar{f}$  and not on the underlying structure with hidden states that the model builder used to deduce that conditional distribution. The only thing that learning contributes is a justification for a particular specification of  $\bar{f}$ . We would get equivalent asset pricing implications by just *assuming*  $\bar{f}$  from the start.

## 2.2 Robust learning and asset pricing

Our application of distinct risk-sensitivity operators to twist the component distributions  $f, g, h$  means that that equivalence is not true in our model because it makes asset prices depend on the evolution of the hidden states and not simply on the distribution of future signals conditioned on signal histories. Following Hansen and Sargent (2007), this occurs because the representative consumer explores potential misspecifications of the distributions of hidden Markov states and of future signals conditioned on those hidden Markov states.<sup>12</sup>

Our representative consumer copes with model misspecification by replacing the  $f, g, h$  conditional densities, respectively, with worst-case densities  $\hat{f}, \hat{g}, \hat{h}$ . With a robust representative consumer, we can use the implied ( $\hat{\cdot}$ ) version of density  $\bar{f}$  distribution to represent the asset price as

$$\hat{P}_{t,\tau}(\alpha) = \hat{E} \left[ Z_{t+\tau}(\alpha) \middle| \mathcal{Y}_t \right]. \quad (6)$$

Using the density  $\hat{f}$  to account for unknown dynamics, we now construct  $\hat{E}[Z_{t+\tau}(\alpha)|X_t(\iota) = x, \iota]$ . Our information decomposition of the asset price with a robust representative consumer becomes

$$\hat{P}_{t,\tau}(\alpha) = \int \int \hat{E}[Z_{t+\tau}(\alpha)|X_t(\iota) = x, \iota] \underbrace{\hat{g}(x|\iota, \mathcal{Y}_t)dx}_{\substack{\uparrow \\ \text{unknown} \\ \text{state}}} \underbrace{\hat{h}(\iota|\mathcal{Y}_t)d\iota}_{\substack{\uparrow \\ \text{unknown} \\ \text{model}}}.$$

---

<sup>12</sup>As emphasized by Hansen (2007), by exploring these misspecifications in effect the our representative consumer refuses to reduce compound lotteries.

We can also represent the price in terms of the original undistorted distribution

$$\hat{P}_{t,\tau}(\alpha) = E \left[ Z_{t+\tau}(\alpha) \left( \frac{\hat{f}[Y_t^{t+\tau}|\iota, X_t(\iota)]}{f[Y_t^{t+\tau}|\iota, X_t(\iota)]} \right) \left( \frac{\hat{g}[X_t(\iota)|\iota, \mathcal{Y}_t]}{g[X_t(\iota)|\iota, \mathcal{Y}_t]} \right) \left( \frac{\hat{h}[\iota|\mathcal{Y}_t]}{h[\iota|\mathcal{Y}_t]} \right) \middle| \mathcal{Y}_t \right] \quad (7)$$

where we have substituted in the random unobserved state vector and the random future signals. Equivalently, the price with a robust representative consumer can be represented as

$$\hat{P}_{t,\tau}(\alpha) = E \left( \frac{M_t^{t+\tau}}{M_t} Z_{t+\tau}(\alpha) \middle| \mathcal{Y}_t \right)$$

where

$$M_t^{t+\tau} = \underbrace{\frac{\hat{f}[Y_t^{t+\tau}|\iota, X_t(\iota)]}{f[Y_t^{t+\tau}|\iota, X_t(\iota)]}}_{\substack{\uparrow \\ \text{distorted} \\ \text{dynamics}}} \underbrace{\frac{\hat{g}[X_t(\iota)|\iota, \mathcal{Y}_t]}{g[X_t(\iota)|\iota, \mathcal{Y}_t]}}_{\substack{\uparrow \\ \text{distorted} \\ \text{state estimation}}} \underbrace{\frac{\hat{h}[\iota|\mathcal{Y}_t]}{h[\iota|\mathcal{Y}_t]}}_{\substack{\uparrow \\ \text{distorted} \\ \text{model probabilities}}} \quad (8)$$

satisfies  $E(M_t^{t+\tau} | \mathcal{Y}_t) = 1$ .

In section 6, we show how to represent the three relative densities  $\frac{\hat{f}}{f}, \frac{\hat{g}}{g}, \frac{\hat{h}}{h}$ , respectively, that emerge from applying risk-sensitivity operators to conditional value functions. These operators adjust separately for misspecification of  $f, g$ , and  $h$ . Continuation utilities will be key determinants of how our representative consumer uses signal histories to learn about hidden Markov states, an ingredient absent from those earlier applications of Bayesian learning that reduced the representative consumer's information prior to asset pricing. In the continuous-time setting to be laid out in section 3, we show how changes in probability measure can be conveniently depicted as martingales. As we will see, there is a martingale associated with each of the channels highlighted by (8). For the “distorted” dynamics, in section 6.2 we construct a martingale  $\{M_t^f\}$  that alters the hidden state dynamics, including the link between future signals and the current state reflected in the density ratio  $\frac{\hat{f}}{f}$ . The martingale is constructed relative to a sequence of information sets that includes the hidden state histories and knowledge of the model. In section 6.3, we construct a second martingale  $\{M_t^g\}$  by including an additional distortion to state estimation conditioned on a model as reflected in the density ratio  $\frac{\hat{g}}{g}$ . This martingale is relative to a sequence

of information sets that condition on the signal history and model, but not on the history of hidden states. Finally, in section 6.4 we produce a martingale  $\{M_t^u\}$  that alters the probabilities over models and is constructed relative to a sequence of conditioning information sets that includes only the signal history and is reflected in the density ratio  $\frac{\hat{h}}{h}$ .

### 3 Three information structures

We use a hidden Markov model and two filtering problems to construct three information sets that define risks to be priced with and without concerns about robustness to model misspecification.

#### 3.1 State evolution

Two models  $\iota = 0, 1$  take the state-space forms

$$\begin{aligned} dX_t(\iota) &= A(\iota)X_t(\iota)dt + B(\iota)dW_t(\iota) \\ dY_t &= D(\iota)X_t(\iota)dt + G(\iota)dW_t(\iota) \end{aligned} \tag{9}$$

where  $X_t(\iota)$  is the *state*,  $Y_t$  is the (cumulated) *signal*, and  $\{W_t(\iota) : t \geq 0\}$  is a multivariate standard Brownian motion. For notational simplicity, we suppose that the same Brownian motion drives both models. Under full information,  $\iota$  is observed and the vector  $dW_t(\iota)$  gives the new information available to the consumer at date  $t$ .

#### 3.2 Filtering problems

To generate two alternative information structures, we solve two types of filtering problems. Let  $\mathcal{Y}_t$  be generated by the history of the signal  $dY_\tau$  up to  $t$  and any other prior information available as of date zero. In what follows, we first condition on  $\mathcal{Y}_t$  and  $\iota$  for each  $t$ . We then omit  $\iota$  and condition on only  $\mathcal{Y}_t$ .

### 3.2.1 Innovations representation with model known

First, suppose that  $\iota$  is known. Application of the Kalman filter yields the innovations representation

$$d\bar{X}_t(\iota) = A(\iota)\bar{X}_t(\iota)dt + K_t(\iota)[dY_t - D(\iota)\bar{X}_t(\iota)dt]$$

where  $\bar{X}_t(\iota) = E[X_t(\iota)|\mathcal{Y}_t, \iota]$  and

$$K_t(\iota) = [B(\iota)G(\iota)' + \Sigma_t(\iota)D(\iota)'] [G(\iota)G(\iota)']^{-1}$$

$$\begin{aligned} \frac{d\Sigma_t(\iota)}{dt} = & A(\iota)\Sigma_t(\iota) + \Sigma_t A(\iota)' + B(\iota)B(\iota)' \\ & - K_t(\iota)[G(\iota)B(\iota)' + D(\iota)\Sigma_t(\iota)]. \end{aligned} \quad (10)$$

We allow more shocks than signals, so  $G(\iota)$  can have more columns than rows. This possibility leads us to construct a nonsingular matrix  $\bar{G}(\iota)$  where  $G(\iota)G(\iota)' = \bar{G}(\iota)\bar{G}(\iota)'$ . The *innovation process* is

$$d\bar{W}_t(\iota) = [\bar{G}(\iota)]^{-1} [dY_t - D(\iota)\bar{X}_t(\iota)dt],$$

where the innovation  $d\bar{W}_t(\iota)$  comprises the new information revealed by the signal history.

### 3.2.2 Innovations representation with model unknown

Assume that  $G(\iota)G(\iota)'$  is independent of  $\iota$ . Without this assumption,  $\iota$  is revealed immediately. Let  $\bar{\iota}_t = E(\iota|\mathcal{Y}_t)$  and

$$d\bar{W}_t = \bar{G}^{-1} (dY_t - \nu_t dt) = \bar{\iota}_t d\bar{W}_t(1) + (1 - \bar{\iota}_t) d\bar{W}_t(0)$$

where

$$\nu_t \doteq [\bar{\iota}_t D(1)\bar{X}_t(1) + (1 - \bar{\iota}_t)D(0)\bar{X}_t(0)]. \quad (11)$$

Then

$$d\bar{\iota}_t = \bar{\iota}_t(1 - \bar{\iota}_t)[\bar{X}_t(1)'D(1)' - \bar{X}_t(0)'D(0)'] (\bar{G}')^{-1} d\bar{W}_t. \quad (12)$$

The new information pertinent to consumers is now  $d\bar{W}_t$ .

## 4 Risk prices

Section 3 described three information structures: i) full information, ii) hidden states with a known model, iii) hidden states with an unknown model. We use the associated Brownian motions  $W(\iota)$ ,  $\bar{W}_t(\iota)$ , and  $\bar{W}_t$  as risks to be priced under the same information structure that generated them. (But in section 5 we shall price the three risk vectors under full information in order to look at Bayesian learning from another angle.) The forms of the risk prices are identical for all three information structures and are familiar from Breeden (1979). The local normality of the diffusion model makes the risk prices be equal to the exposures of the log marginal utility to the underlying risks. Let the increment of the logarithm of consumption be given by  $d \log C_t = H' dY_t$ , implying that consumption growth rates are revealed by the increment in the signal vector. Each of our different information sets implies a risk price vector, as reported in Table 1.

Because different risks are being priced, the risk prices differ across information structures. However, the *magnitudes* of the risk price vectors are identical across information structures. As we saw in (4), the magnitude of the risk price vector is the slope of the instantaneous mean-standard deviation frontier. In section 6, we shall show how a concern about model misspecification alters risk prices by adding compensations for bearing model uncertainty. But first we want to look at Bayesian learning and risk prices from a different perspective.

information	local risk	risk price	slope
full	$dW_t(\iota)$	$\gamma G(\iota)' H$	$\gamma \sqrt{H' G(\iota) G(\iota)' H}$
unknown state	$d\bar{W}_t(\iota)$	$\gamma \bar{G}(\iota)' H$	$\gamma \sqrt{H' G(\iota) G(\iota)' H}$
unknown model	$d\bar{W}_t$	$\gamma \bar{G}' H$	$\gamma \sqrt{H' G(\iota) G(\iota)' H}$

Table 1: When the model is unknown,  $G(\iota)G(\iota)'$  is assumed to be independent of  $\iota$ . The parameter  $\gamma$  is the coefficient of relative risk aversion in a power utility model. The entries in the “slope” column are the implied slopes of the mean-standard deviation frontier. The consumption growth rate is  $d \log C_t = H' dY_t$ .

## 5 A full-information perspective on agents' learning

In this section, we study what happens when an econometrician mistakenly assumes that consumers have a larger information set than they actually do. We know that an econometrician who conditions on *less* information than consumers still draws correct inferences about the magnitude of risk prices. But we shall see that an econometrician who mistakenly conditions on *more* information than consumers makes false inferences about that magnitude. We regard the consequences of an econometrician's mistakenly conditioning on more information than consumers as contributing to the analysis of risk pricing under consumers' Bayesian learning.

Hansen and Richard (1987) systematically studied the consequences for risk prices of an econometrician's conditioning on less information than consumers. Given a correctly specified stochastic discount factor process, if economic agents use *more* information than an econometrician, the consequences for the econometrician's inferences about risk prices can be innocuous. In constructing conditional moment restrictions for asset prices, all that is required is that the econometrician at least includes prices in his information set. By application of the law of iterated expectations, the product of a cumulative return and a stochastic discount factor remains a martingale when some of the information available to consumers is omitted from the econometrician's information set.<sup>13</sup> While the econometrician who omits information fails correctly to infer the risk components actually confronted by consumers, that mistake does not prevent him from correctly inferring the slope of the mean-standard deviation frontier, as indicated in the third column of table 1 in section 3.

We now consider the reverse situation when economic agents use *less* information than an econometrician. We use the full-information structure but price risks generated by less informative information structures, in particular,  $d\bar{W}_t(\iota)$  and  $d\bar{W}_t$ . In pricing  $d\bar{W}_t(\iota)$  and  $d\bar{W}_t$  under full information, we use pricing formulas that take the mistaken Olympian perspective (often used in macroeconomics) that consumers know the full-information probability distribution of signals. This mistake by the

---

<sup>13</sup>This observation extends an insight of Shiller (1972), who, in the context of a rational expectations model of the term structure of interest rates, pointed out that when an econometrician omits conditioning information used by agents, there emerges an error term that is uncorrelated with information used by the econometrician. Hansen and Sargent (1980) studied the econometric implications of such "Shiller errors" in a class of linear rational expectations models.

econometrician induces a pricing error relative to the prices that actually confront the consumer because the econometrician has misspecified the information available to the consumer. The price discrepancies capture effects of a representative agent's learning that Bossaerts (2002, 2004) and Cogley and Sargent (2008) featured.

## 5.1 Hidden states but known model

Consider first the case in which the model is known. Represent the innovation process as

$$d\bar{W}_t(\iota) = [\bar{G}(\iota)]^{-1} \left( D(\iota) [X_t(\iota) - \bar{X}_t(\iota)] dt + G(\iota) dW_t(\iota) \right).$$

This expression reveals that  $d\bar{W}_t(\iota)$  bundles two risks:  $X_t(\iota) - \bar{X}_t(\iota)$  and  $dW_t(\iota)$ . An innovation under the reduced information structure ceases to be an innovation in the original full information structure. The “risk”  $X_t(\iota) - \bar{X}_t(\iota)$  under the limited information structure ceases to be a risk under the full information structure.

Consider the small time interval limit of

$$\frac{\bar{Q}_{t+\tau}(\bar{\alpha})}{Q_t(\bar{\alpha})} = \exp \left( \bar{\alpha}' [\bar{W}_{t+\tau}(\iota) - \bar{W}_t(\iota)] - \frac{|\bar{\alpha}|^2 \tau}{2} \right).$$

This has unit expectation under the partial information structure. The local price computed under the full information structure is

$$-\delta - \gamma H X_t(\iota) + \bar{\alpha}' [\bar{G}(\iota)]^{-1} D(\iota) [X_t(\iota) - \bar{X}_t(\iota)] + \frac{1}{2} \left| -\gamma H' G(\iota) + \bar{\alpha}' [\bar{G}(\iota)]^{-1} G(\iota) \right|^2 - \frac{|\bar{\alpha}|^2}{2}$$

where  $\delta$  is the subjective rate of discount. Multiplying by minus one and differentiating with respect to  $\bar{\alpha}$  gives the local price

$$\gamma \bar{G}(\iota)' H + [\bar{G}(\iota)]^{-1} D(\iota) [\bar{X}_t(\iota) - X_t(\iota)].$$

The first term is the risk price under partial information (see Table 1), while the second term is the part of the forecast error in the signal under the reduced information set that can be forecast perfectly under the full information set.

## 5.2 States and model both unknown

Consider next what happens when the model is unknown. Suppose that  $\iota = 1$  and represent  $d\bar{W}_t$  as

$$d\bar{W}_t = \bar{G}^{-1} [G(1)dW_t(1) + D(1)X_t(1)dt] - \bar{G}^{-1} [\bar{\iota}_t D(1)\bar{X}_t(1)dt + (1 - \bar{\iota}_t)D(0)\bar{X}_t(0)dt].$$

There is an analogous calculation for  $\iota = 0$ . When we compute local prices under full information, we obtain

$$\gamma\bar{G}'H + \bar{G}^{-1} [\nu_t - D(\iota)X_t(\iota)], \quad (13)$$

where  $\nu_t$  is defined in (11).

The term  $\gamma\bar{G}'H$  is the risk price under reduced information when the model is unknown (see Table 1). The term  $\bar{G}^{-1} [\nu_t - D(\iota)X_t(\iota)]$  is a contribution to the risk price measured by the econometrician coming from the effects of the consumer's learning on the basis of his more limited information set. With respect to the probability distribution used by the consumer, this term averages out to zero. Since  $\iota$  is unknown, the average includes a contribution from the prior. For some sample paths, this term can have negative entries for substantial lengths of time, indicating that the prices under the reduced information exceed those computed under full information. Other trajectories could display the opposite phenomenon. It is thus possible that the term  $\bar{G}^{-1} [\nu_t - D(\iota)X_t(\iota)]$  contributes apparent pessimism or optimism, depending on the prior over  $\iota$  and the particular sample path. In what follows, we use concerns about robustness to motivate priors that are *necessarily* pessimistic. Our notion of pessimism is endogenous in the sense that it depends on the consumer's continuation values. That endogeneity makes pessimism time-dependent and state-dependent in ways that we explore below.

## 6 Price effects of concerns about robustness

When prices reflect a representative consumer's fears of model misspecification, (2) must be replaced by (6) or equivalently (7). To compute distorted densities under our alternative information structures, we must find value functions for a planner

who fears model misspecification.<sup>14</sup> In section 4, we constructed what we called “risk prices” that assign prices to exposures to shocks. We now construct somewhat analogous prices, but because they will include contributions from fears of model misspecification, we shall refer to them as “shock prices”. We construct components of these prices for our three information structures and display them in the last column of Table 2. Specifically, this column gives the contribution to the shock prices from each type of model uncertainty.

information	local risk	risk price	uncertainty price
full	$dW_t(\iota)$	$G(\iota)'H$	$\frac{1}{\theta_1}[B(\iota)'\lambda(\iota) + G(\iota)'H]$
unknown state	$d\bar{W}_t(\iota)$	$\bar{G}(\iota)'H$	$\frac{1}{\theta_2}\bar{G}(\iota)^{-1}D(\iota)\Sigma_t(\iota)\lambda(\iota)$
unknown model	$d\bar{W}_t$	$\bar{G}'H$	$(\tilde{\iota} - \bar{\iota})\bar{G}^{-1}[D(1)\bar{x}(1) - D(0)\bar{X}(0)]$

Table 2: When the model is unknown,  $G(\iota)G(\iota)'$  is assumed to be independent of  $\iota$ . The consumption growth rate is  $d \log C_t = H'dY_t$ . Please *cumulate* contributions to uncertainty prices as you move down the last column.

## 6.1 Value function without robustness

We study a consumer with unitary elasticity of intertemporal substitution and so start with the value function for discounted expected utility using a logarithm period utility function

$$\begin{aligned}
V(x, c, \iota) &= \delta E \left[ \int_0^\infty \exp(-\delta\tau) \log C_{t+\tau} | X_t(\iota) = x, \log C_t = c, \iota \right] \\
&= \delta E \left[ \int_0^\infty \exp(-\delta\tau) (\log C_{t+\tau} - \log C_t) | X_t(\iota) = x, \log C_t = c, \iota \right] + c \\
&= \lambda(\iota) \cdot x + c
\end{aligned}$$

where the vector  $\lambda(\iota)$  satisfies

$$0 = -\delta\lambda(\iota) + D(\iota)'H + A(\iota)'\lambda(\iota), \quad (14)$$

<sup>14</sup>Hansen and Sargent (2008, chs.11-13) discuss the role of the planner’s problem in computing and representing prices with which to confront a representative consumer.

so that

$$\lambda(\iota) = [\delta I - A(\iota)']^{-1} D(\iota)' H. \quad (15)$$

The form of the value function is the same as that of Tallarini (2000) and Barillas et al. (2008). The value function under limited information simply replaces  $x$  with the best forecast  $\bar{x}$  of the state vector given past information on signals.

## 6.2 Full information

Consider first the full information environment in which states are observed and the model is known. The model, however, now becomes a benchmark in a way that we formalize mathematically. Specifically, a concern for robustness under full information gives us a way to construct  $\frac{\hat{f}}{f}$  in (8) via a martingale  $\{M_t^f(\iota)\}$  with respect to the benchmark probability model. The relative density  $\frac{\hat{f}}{f}$  distorts the future signals conditioned on the current state and model by distorting both the state and signal dynamics. In a diffusion setting, a concern about robustness induces the consumer to consider distortions that append a drift  $\mu_t dt$  to the Brownian increment and to impose a quadratic penalty to this distortion. This leads to a minimization problem whose indirect value function yields the  $\mathbb{T}^1$  operator of Hansen and Sargent (2007).<sup>15</sup>

### Problem 6.1.

$$\begin{aligned} 0 = \min_{\mu} & -\delta[\lambda(\iota) \cdot x + \kappa(\iota)] + x' D(\iota)' H + \mu' G(\iota)' H + x' A(\iota)' \lambda(\iota) \\ & + \mu' B(\iota)' \lambda(\iota) + \frac{\theta_1}{2} \mu' \mu. \end{aligned} \quad (16)$$

where we conjecture a value function of the form  $\lambda(\iota) \cdot x + \kappa(\iota) + c$ .

Here  $\theta_1$  is a positive penalty parameter that characterizes the decision maker's fear that model  $\iota$  is misspecified. We impose the same  $\theta_1$  for both models. See Hansen et al. (2006) and Anderson et al. (2003) for more general treatments and see appendix

---

<sup>15</sup>Suppose that the decision maker has instantaneous utility function  $u(x)$ , positive discount rate  $\delta$ , and that the state follows the diffusion  $dx_t = \nu(x_t)dt + \sigma(x_t)dW_t$ . The value function  $V(x)$  associated with a multiplier problem satisfies the Bellman equation  $\delta V(x) = \min_h (u(x) + \frac{\theta}{2} h' h + [\nu(x) + \sigma(x)h]V_x(x) + \frac{1}{2} \text{trace}[\sigma(x)' V_{xx} \sigma(x)])$ . The indirect value function for this problem satisfies the Bellman equation  $\delta S(x) = u(x) + \nu(x)S_x(x) + \frac{1}{2} \text{trace}[\sigma(x)' S_{xx}(x)\sigma(x)] - \frac{1}{2\theta} S_x(x)' \sigma(x)\sigma(x)' S_x(x)$ , which is an example of the stochastic differential utility specification of Duffie and Epstein (1992). See Hansen et al. (2006).

As for how we propose to calibrate  $\theta_1$ . The minimizing drift distortion

$$\tilde{\mu}(\iota) = -\frac{1}{\theta_1} [G(\iota)'H + B(\iota)'\lambda(\iota)] \quad (17)$$

is independent of the state vector  $X(\iota)$ . As a result,

$$\kappa(\iota) = -\frac{1}{2\theta_1\delta} |G(\iota)'H + B(\iota)'\lambda(\iota)|^2. \quad (18)$$

Equating coefficients on  $x$  in (16) implies that equation (14) continues to hold. Thus,  $\lambda(\iota)$  remains the same as in the model without robustness and is given by (15).

**Proposition 6.2.** *The value function shares the same  $\lambda(\iota)$  with the expected utility model [formula (14)] and  $\kappa(\iota)$  is given by (18). The associated worst case distribution for the Brownian increment is normal with covariance matrix  $Idt$  and drift  $\tilde{\mu}(\iota)dt$  given by (17).*

Under full information, the likelihood of the worst-case model relative to that of the benchmark model is a martingale  $\{M_t^f(\iota)\}$  with local evolution

$$d \log M_t^f(\iota) = \tilde{\mu}(\iota)'dW_t(\iota) - \frac{1}{2}|\tilde{\mu}(\iota)|^2 dt,$$

so that the mean of  $M_t^f(\iota)$  is evidently unity. The stochastic discount factor (relative to the benchmark model) includes contributions both from the consumption dynamics and from the martingale, so that

$$d \log S_t^f = d \log M_t^f(\iota) - \delta dt - d \log C_t.$$

The vector of local shock price is once again the negative of the exposure of the stochastic discount factor to the respective shocks. With robustness, the shock price vector under full information is augmented by an uncertainty price:

$$\underbrace{G(\iota)'H}_{\substack{\uparrow \\ \text{risk}}} + \underbrace{\frac{1}{\theta_1} [G(\iota)'H + B(\iota)'\lambda(\iota)]}_{\substack{\uparrow \\ \text{uncertainty}}}$$

Neither the risk contribution nor the uncertainty contribution to the shock prices is

either state dependent or time dependent. We have completed the first row of Table 2.

### 6.3 Unknown states

Now suppose that the model (the value of  $\iota$ ) is known but that the state  $X_t(\iota)$  is not. We proceed to construct  $\frac{\hat{q}}{g}$  in formula (8).

We seek a martingale  $\{M_t^i\}$  to use under this information structure. Following Hansen and Sargent (2007), we introduce a positive penalty parameter  $\theta_2$  and construct a robust estimate of the hidden state  $X_t(\iota)$  by solving:

**Problem 6.3.**

$$\begin{aligned} \min_{\phi, f} \int_{\phi(x)\psi(x|\bar{x}, \Sigma)dx=1} [\lambda(\iota) \cdot x + \kappa(\iota) + \theta_2 \log \phi(x)] \phi(x)\psi(x|\bar{x}, \Sigma)dx \\ = \min_x \lambda(\iota) \cdot x + \kappa(\iota) + \frac{\theta_2}{2}(x - \bar{x})'(\Sigma)^{-1}(x - \bar{x}) \end{aligned}$$

where  $\psi(x|\bar{x}, \Sigma)$  is the normal density with mean  $\bar{x}$  and covariance matrix  $\Sigma$ .

The estimates of the state and the covariance matrix used to construct  $\psi$  at a given point in time will typically depend on the model  $\iota$ .

In the first line of Problem 6.3,  $\phi$  is a density (relative to a normal) that distorts the density  $\psi$  for the hidden state and  $\theta_2$  is a positive penalty parameter that penalizes  $\phi$ 's with large values of relative entropy (the expected value of  $\phi \log \phi$ ). The second line of Problem 6.3 exploits the fact that the worst-case density is necessarily normal with a distortion  $\tilde{x}$  to the mean of the state. This structure makes it straightforward to compute the integral and therefore simplifies the minimization problem. In particular, the worst-case state estimate  $\tilde{x}$  solves

$$0 = \lambda(\iota) + \theta_2(\Sigma)^{-1}(\tilde{x} - \bar{x}).$$

**Proposition 6.4.** *The robust value function is*

$$U[\iota, \bar{x}, \Sigma] = \lambda(\iota) \cdot \bar{x} + \kappa(\iota) - \frac{1}{2\theta_2} \lambda(\iota)' \Sigma \lambda(\iota) + c \quad (19)$$

with the same  $\lambda(\iota)$  as in the expected utility model [formula (14)] and the same  $\kappa(\iota)$  as in the robust planner's problem with full information [formula (18)]. The worst-case

state estimate is

$$\tilde{x} = \bar{x} - \frac{1}{\theta_2} \Sigma \lambda(\iota).$$

The indirect value function on the right side of (19) defines an instance of the  $\mathbb{T}^2$  operator of Hansen and Sargent (2007). Under the distorted evolution,  $dY_t$  has drift

$$\tilde{\xi}_t(\iota) dt = D(\iota) \tilde{X}_t(\iota) dt + G(\iota) \tilde{\mu}(\iota) dt,$$

while under the benchmark evolution it has drift

$$\bar{\xi}_t(\iota) dt = D(\iota) \bar{X}_t dt.$$

The corresponding likelihood ratio for our limited information setup is a martingale  $M_t^i(\iota)$  that evolves as

$$d \log M_t^i(\iota) = [\tilde{\xi}_t(\iota) - \bar{\xi}_t(\iota)]' [\bar{G}(\iota)]^{-1} d\bar{W}_t(\iota) - \frac{1}{2} [\bar{G}(\iota)]^{-1} [\tilde{\xi}_t(\iota) - \bar{\xi}_t(\iota)]^2 dt,$$

and therefore the stochastic discount factor evolves as

$$d \log S_t^i = d \log M_t^i(\iota) - \delta dt - d \log C_t.$$

There are now two contributions to the uncertainty price, the one in the last column of the first row of table 2 coming from the potential misspecification of the state dynamics as reflected in the drift distortion to the Brownian motion and the other in the second row of table 2 coming from the filtering problem as reflected in a distortion in the estimated mean of hidden state vector:

$$\underbrace{\bar{G}(\iota)' H}_{\uparrow \text{ risk}} + \underbrace{\frac{1}{\theta_1} [\bar{G}(\iota)]^{-1} G(\iota) [G(\iota)' H + B(\iota)' \lambda(\iota)]}_{\uparrow \text{ model uncertainty}} + \underbrace{\frac{1}{\theta_2} [\bar{G}(\iota)]^{-1} D(\iota) \Sigma_t(\iota) \lambda(\iota)}_{\uparrow \text{ estimation uncertainty}}.$$

The state estimation adds time dependence to the uncertainty prices through the evolution of the covariance matrix  $\Sigma_t(\iota)$  governed by (10), but not through the observed history of signals. We have completed the second row of Table 2.

## 6.4 Model unknown

Finally, we obtain a martingale  $\{M_t^u\}$  that represents a robust adjustment for an unknown model. Thus, we now construct  $\frac{\hat{\iota}}{h}$  in formula (8). We do this by twisting the model probability  $\bar{\iota}_t$  by solving:

**Problem 6.5.**

$$\begin{aligned} \min_{0 \leq \iota \leq 1} \quad & \iota U[1, \bar{x}(1), \Sigma(1)] + (1 - \iota) U[0, \bar{x}(0), \Sigma(0)] \\ & + \theta_2 \iota [\log \iota - \log \bar{\iota}] + \theta_2 (1 - \iota) [\log(1 - \iota) - \log(1 - \bar{\iota})] \end{aligned}$$

**Proposition 6.6.** *The indirect value function for this problem is the robust value function<sup>16</sup>*

$$-\theta_2 \log \left[ \bar{\iota} \exp \left( -\frac{1}{\theta_2} U[1, \bar{x}(1), \Sigma(1)] \right) + (1 - \bar{\iota}) \exp \left( -\frac{1}{\theta_2} U[0, \bar{x}(0), \Sigma(0)] \right) \right].$$

The worst-case model probabilities satisfy

$$(1 - \tilde{\iota}) \propto (1 - \bar{\iota}) \exp \left( -\frac{U[0, \bar{x}(0), \Sigma(0)]}{\theta_2} \right) \quad (20)$$

$$\tilde{\iota} \propto \bar{\iota} \exp \left( -\frac{U[1, \bar{x}(1), \Sigma(1)]}{\theta_2} \right) \quad (21)$$

where the constant of proportionality is the same for both expressions.

Under the distorted probabilities, the signal increment  $dY_t$  has a drift

$$\tilde{\kappa}_t dt = [\tilde{\iota}_t \tilde{\xi}_t(1) + (1 - \tilde{\iota}_t) \tilde{\xi}_t(0)] dt,$$

which we compare to the drift derived previously under the benchmark probabilities:

$$\bar{\kappa}_t dt = [\bar{\iota}_t \bar{\xi}_t(1) + (1 - \bar{\iota}_t) \bar{\xi}_t(0)] dt.$$

The associated martingale constructed from the relative likelihoods has evolution

$$d \log M_t^u = (\tilde{\kappa}_t - \bar{\kappa}_t)' (\bar{G}')^{-1} d\bar{W}_t - \frac{1}{2} |\bar{G}^{-1} (\tilde{\kappa}_t - \bar{\kappa}_t)|^2 dt$$

---

<sup>16</sup>This is evidently another application of the  $T^2$  operator of Hansen and Sargent (2007).

and the stochastic discount factor is

$$d \log S_t = d \log M_t^u - \delta dt - d \log C_t.$$

The resulting shock price vector equals the exposure of  $d \log S_t$  to  $d\bar{W}_t$  and is the ordinary risk price  $\bar{G}'H$  plus the following contribution coming from concerns about model misspecification:

$$\begin{aligned} & \tilde{\iota} \bar{G}^{-1} \left[ \frac{1}{\theta_1} G(1)G(1)'H + \frac{1}{\theta_1} G(1)B(1)'\lambda(1) \right] + (1 - \tilde{\iota}) \bar{G}^{-1} \left[ \frac{1}{\theta_1} G(0)G(0)'H + \frac{1}{\theta_1} G(0)B(0)'\lambda(0) \right] \\ & + \tilde{\iota} \bar{G}^{-1} \left[ \frac{1}{\theta_2} D(1)\Sigma(1)\lambda(1) \right] + (1 - \tilde{\iota}) \bar{G}^{-1} \left[ \frac{1}{\theta_2} D(0)\Sigma(0)\lambda(0) \right] \\ & + (\bar{\iota} - \tilde{\iota}) \bar{G}^{-1} [D(1)\bar{x}(1) - D(0)\bar{x}(0)]. \end{aligned} \quad (22)$$

As summarized in Table 2, the first term reflects uncertainty in *state dynamics* associated with each of the two models. Hansen et al. (1999) feature a similar term. It is forward looking by virtue of the appearance of  $\lambda(\iota)$  determined in (15). The next term reflects uncertainty about *hidden states* in each of the respective models. When  $\tilde{\iota} < 1$ , it depends partly on the evolution of  $\tilde{\iota}$ . In the limiting case in which  $\tilde{\iota} = 1$ , the first term is constant over time and the next one depends on time but not on the signal history. In our application, this limiting case approximately obtains when  $\theta_2$  is sufficiently small. The third term reflects uncertainty about the *models* and depends on the signal history even when  $\tilde{\iota} = 1$ . The term that is scaled by  $\bar{\iota} - \tilde{\iota}$  is also central to the evolution of model probabilities given in (12) and dictates how new information contained in the signals induces changes in the model probabilities under the benchmark specification. In effect,  $\bar{G}^{-1} [D(1)\bar{x}(1) - D(0)\bar{x}(0)]$ , appropriately scaled, is the response vector of new information in the signals to the updating of the probability assigned to models  $\iota$ . The signal realizations over the next instant push the decision-maker's posterior towards one of the two models and this is reflected in the equilibrium uncertainty prices. In addition to responding to the signal history, this response vector will recurrently change signs within an observed sample whenever discriminating between models is challenging. In such cases, we expect new information not always to move probabilities in the same direction. In the third term of (22), this response vector is scaled by the difference between the current model

probabilities under the benchmark and worst case models. Formulas (20) and (21) indicate how the consumer slants the model probabilities towards the model with the worse utility consequences. This probability slanting induces additional history dependence through  $\bar{\iota}_t$ , which depends on the signal history.

## 7 Illustrating the mechanism

To highlight the forces that govern the component contributions of model uncertainty to shock prices in formula (22), we create a long-run risk model with a predictable growth rate along the lines of Bansal and Yaron (2004) and Hansen et al. (2008a). Our models share the form  $d \log C_t = dY_t$  and

$$\begin{aligned} dX_{1t} &= a(\iota)X_{1t}(\iota)dt + \sigma_1(\iota)dW_{1t}(\iota) \\ dX_{2t} &= 0 \\ dY_t &= X_{1t}dt + X_{2t}dt + \sigma_2(\iota)dW_{2t}(\iota) \end{aligned} \tag{23}$$

where  $X_{1t}(\iota), X_{2t}(\iota)$  are scalars and  $W_{1t}(\iota), W_{2t}(\iota)$  are scalar components of the vector Brownian motion  $W_t(\iota)$  and where  $X_{20}(\iota) = \mu_y(\iota)$  is the unconditional mean of consumption growth for model  $\iota$ . We use the following discrete-time approximation to the state space system (9):

$$\begin{aligned} X_{t+\tau}(\iota) - X_t(\iota) &= \tau A(\iota)X_t(\iota) + B(\iota)[W_{t+\tau}(\iota) - W_t(\iota)] \\ Y_{t+\tau} - Y_t &= \tau D(\iota)X_t(\iota) + G(\iota)[W_{t+\tau}(\iota) - W_t(\iota)]. \end{aligned}$$

We set  $\tau = 1$ .

A small negative  $a(\iota)$  coupled with a small  $\sigma_1(\iota)$  captures long-run risks in consumption growth. Bansal and Yaron (2004) justify such a specification with the argument that it fits consumption growth approximately as well as, and is therefore difficult to distinguish from, an iid consumption growth model, which we know fits the aggregate per capital U.S. consumption data well. We respect this argument by forming two models with the same values of the signal noise  $\sigma_2(\iota)$  but that with different values of  $\sigma_1(\iota), \rho(\iota) = a(\iota) + 1$ , and  $\mu_y(\iota) = X_{20}(\iota)$ , give *identical* values of the likelihood. We impose  $\rho(1) = .99$  to capture a long-run risk model, while the equally

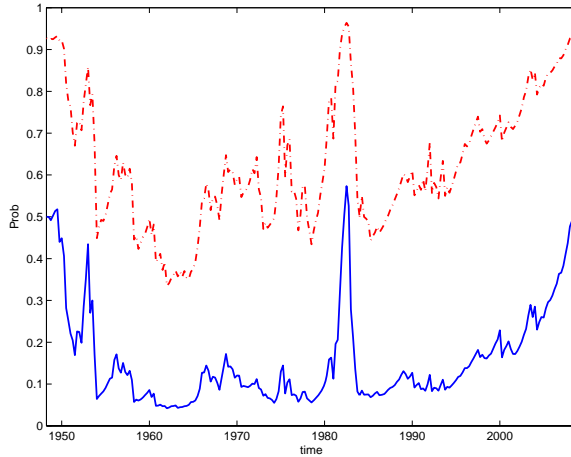


Figure 1: Bayesian model probability  $\bar{\iota}_t$  (solid line) and worst-case model probability  $\tilde{\iota}_t$  (dashed line).

good fitting  $\iota = 0$  model has  $\rho = .36$ .<sup>17</sup> Thus, we have constructed two models that over our sample are indistinguishable statistically. This is our way of making precise the Bansal and Yaron (2004) observation that the long-run risk and iid consumption growth models are difficult to distinguish empirically.

In appendix A we describe how we first calibrated  $\theta_1$  to drive the average detection error probability over the two  $\iota$  models with observed states to be .4 and then, with  $\theta_1$  thereby fixed, set  $\theta_2$  to get a detection error probability of .2 for the signal distribution of the mixture model. We regard these values of detection error probabilities as being associated with moderate amounts of model uncertainty.<sup>18</sup> For these values

<sup>17</sup>The sample for real consumption of services and durables runs over the period 1948II-2008III. To fit model  $\iota = 1$ , we fixed  $\rho = .99$  and estimated  $\sigma_1 = .00043, \sigma_2 = .0048, \mu_y = .0045$ . Fixing  $\sigma_2$  equal to .0048, we then found a values of  $\rho = .36$  and associated values  $\sigma_1 = .00204, \mu_y = .00478$  that give virtually the same value of the likelihood. In this way, we construct two good fitting models that are difficult to distinguish, with model  $\iota = 1$  being the long-run risk model and model  $\iota = 0$  much more closely approximating an iid growth model. Freezing the value of  $\sigma_2$  at the above value, the maximum likelihood estimates are  $\rho = .8179, \sigma_1 = .00132, \mu_y = .0047$ . The data for consumption comes from the St. Louis Fed data set (FRED). They are taken from their latest vintage (11/25/2008) with the following identifiers PCNDGC96\_20081125 (real consumption on nondurable goods), PCESVC96\_20081125 (real consumption on services). The population series is from the BLS, Series ID: LNS10000000. This is civilian noninstitutional population 16 years and over in thousands. The raw data are monthly. We averaged it to compute a quarterly series.

<sup>18</sup>We initiate the Bayesian probability  $\bar{\iota}_0 = .5$  and set the covariance matrices  $\Sigma_0(\iota)$  over hidden states at values that approximate what would prevail for a Bayesian who had previously observed a sample of the length 242 that we have in our actual sample period. In particular, we calibrated the initial state covariance matrices for both models as follows. First, we set preliminary ‘uninformative’

of  $\theta_1, \theta_2$ ,<sup>19</sup> figure 1 plots values of the Bayesian model mixing probability  $\bar{\iota}$  along with the worst-case probability  $\tilde{\iota}$ . As described in the previous paragraph, we have constructed our two models so that with our setting of the initial model probability  $\bar{\iota}_0$  at .5, the terminal value of  $\bar{\iota}_t$  approximates .5. An interesting thing about figure 1 is to watch how the worst-case  $\tilde{\iota}_t$  twists toward the long-run risk  $\iota = 1$  model. This probability twisting contributes to the countercyclical movements to the uncertainty contributions to the shock price (22) that we plot in figure 2.<sup>20</sup>

Figure 3 decomposes the uncertainty contribution to the shock prices into components coming from the three lines of expression (22), namely, those associated with state dynamics with a known model, unknown states with a known model, and an unknown model, respectively. As anticipated, the first two contributions are positive, the first being constant while the second varies over time. The third contribution, due to uncertainty about the model, alternates in sign.

The first contribution is constant and relatively small in magnitude. We have specified our models so that  $G(\iota)B(\iota)' = 0$  and thus

$$\left[ \frac{1}{\theta_1} G(\iota)G(\iota)'H + \frac{1}{\theta_1} G(\iota)B(\iota)'\lambda(\iota) \right] = \frac{1}{\theta_1} \bar{G}\bar{G}'H,$$

which is the same for both models. While the forward-looking component to shock prices reflected in  $\frac{1}{\theta_1}B(\iota)'\lambda(\iota)$  is present in the model with full information, it is absent in our specification with limited information. Later we consider a specification in which this forward-looking component is activated.<sup>21</sup> The second contribution features state estimation. Figure 4 shows the  $D(\iota)\Sigma(\iota)\lambda(\iota)$  components that are important elements of state uncertainty. This figure reveals how hidden states are more difficult to learn about in model  $\iota = 1$  than in model  $\iota = 0$ , because the presence of the very persistent hidden state slows convergence of  $\Sigma_t(1)$ . In particular, for model  $\iota = 1$ , the variance of the estimated unconditional mean of consumption growth,

---

values that we took to be the variance of the unconditional stationary distribution of  $X_{1t}(\iota)$  and a value for the variance of  $X_{2t}(\iota)$  of .01<sup>2</sup>, which is orders of magnitude larger than the maximum likelihood estimates of  $\mu_y$  for our entire sample. We set a preliminary state covariance between  $X_{1t}(\iota)$  and  $X_{2t}(\iota)$  equal to zero. We put these preliminary values into the Kalman filter, ran it for a sample length of 242, and took the terminal covariance matrix as our starting value for the covariance matrix of the hidden state for model  $\iota$ .

<sup>19</sup>The calibrated values are  $\theta_1^{-1} = 7, \theta_2^{-1} = 1$ .

<sup>20</sup>The figure plots all components of (22) except the ordinary risk price  $\bar{G}H'$ .

<sup>21</sup>See section 7.4, where we specify an example in which  $G(\iota)B(\iota)'$  is not zero.

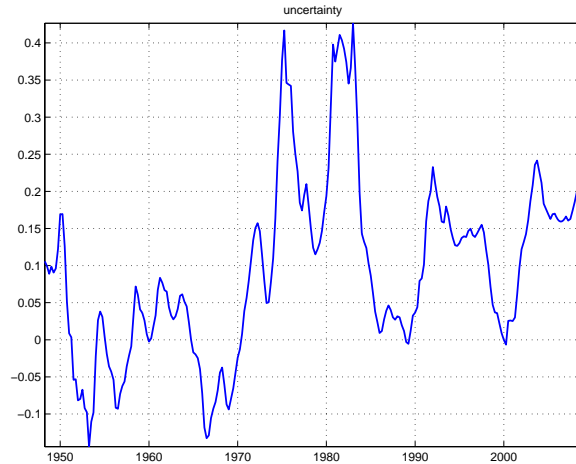


Figure 2: Contributions to uncertainty prices from all sources of model uncertainty.

$\Sigma_t(1)_{22}$ , converges more slowly to zero. The third contribution will generally fluctuate over time in ways that depend on the evolution of the discrepancy between the estimated means  $D(\iota)\bar{x}(\iota)$  under the two models, depicted in figure 5. Thus, while pessimism arising from a concern for robustness necessarily increases the uncertainty prices via the first two terms, it may either lower or raise it through the third term. Recall that the slope of the mean-standard deviation frontier, the maximum Sharpe ratio, is the absolute value of the shock price vector. Sizable shock prices of either sign thus imply large maximum Sharpe ratios. Negative shock prices for some signal histories indicate that positive consumption innovations are sometimes feared because of what they imply about the plausibility of alternative models. The magnitude of these prices determines the familiar risk-return tradeoffs from finance. How concerns about model uncertainty affect uncertainty premia on prices of risky assets will ultimately depend on how returns are correlated with consumption shocks.

## 7.1 Explanation for countercyclical uncertainty premia

The intertemporal behavior of robustness-induced probability slanting accounts for how learning in the presence of uncertainty about models induces time variation in uncertainty premia. Our representative consumer attaches positive probabilities to a model with statistically subtle persistence in consumption growth, namely, the long-run risk model of Bansal and Yaron (2004), and also to another model asserting close

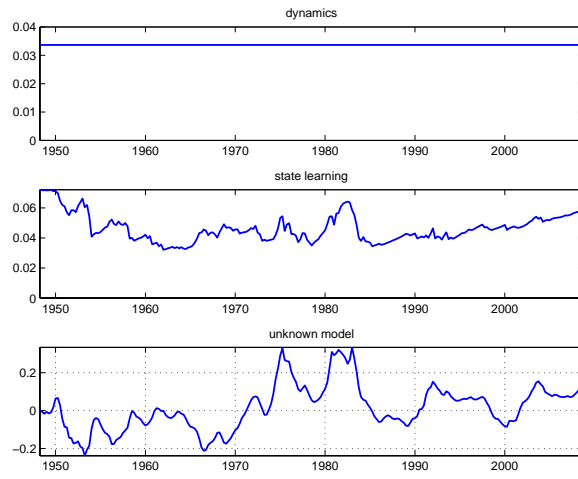


Figure 3: Contributions to uncertainty prices from state dynamics (top panel) and learning hidden state when the model is known (middle panel), and unknown model (bottom panel)

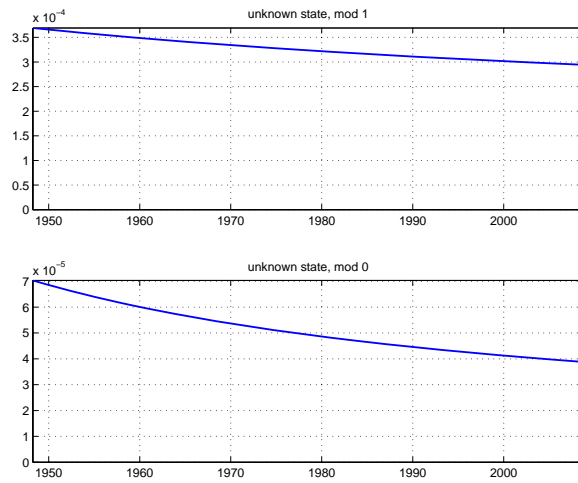


Figure 4:  $D(t)\Sigma(t)\lambda(t)$  for  $\iota = 1$  (top panel) and  $\iota = 0$  (bottom panel).

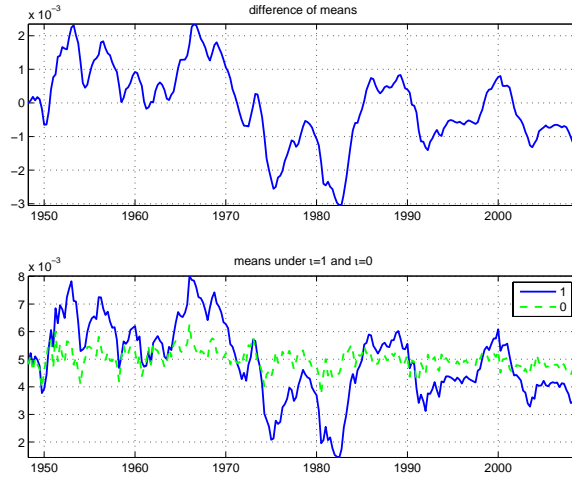


Figure 5: Difference in means and means themselves from models  $\iota = 1$  and  $\iota = 0$ .

to iid consumption growth rates.<sup>22</sup> The asymmetrical response of model uncertainty premia to consumption growth shocks comes from (i) how the representative consumer's concern about possible misspecification of the probabilities that he attaches to models causes him to calculate worst case probabilities that depend on value functions, and (ii) how the value functions for the two models respond to shocks in ways that bring them closer together after positive consumption growth shocks and push them farther apart after negative shocks. The long-run risk model with very persistent consumption growth confronts the consumer with a long-lived shock to consumption growth. That affects the set of possible model misspecifications that he worries about. The representative consumer's concerns about these misspecifications are reflected in a more negative value of  $\kappa(\iota) - \frac{1}{2\theta_2} \lambda(\iota)' \Sigma(\iota) \lambda(\iota)$  in formula (19) for the continuation value. Over our sample, the difference across models rises monotonically from  $-2.55$  to  $-2.42$ . The resulting difference in constant terms in the value functions for the models 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 that the consumer attaches to the  $\iota = 1$  model with persistent consumption growth relative to that of the approximately iid consumption growth  $\iota = 0$  model. Although the long-run risk model has a more negative constant term, when a string of higher than average con-

<sup>22</sup>Appendix B reports a sensitivity analysis aimed to add insight about the source of countercyclical shock prices.

sumption growths occur, persistence of consumption growth under this model means that consumption growth can be expected to remain higher than average for many future periods. This pushes the continuation values associated with the two models closer together than they are when consumption growth rates have recently been lower than average. Via exponential twisting formulas, continuation values determine the worst-case probabilities that the representative consumer attaches to the models. 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 have been higher than average.

## 7.2 Roles of different types of uncertainty

The decomposition of uncertainty contributions to shock prices depicted in figure 3 helps us to think about how these contributions would change if, by changing  $\theta_1$  and  $\theta_2$ , we refocus the representative consumer's concern about misspecification on a different mixture of dynamics, hidden states, and unknown model. Figures 6 and 7 show the consequences of turning off fear of unknown dynamics by setting  $\theta_1 = +\infty$  while lowering  $\theta_2$  to set the detection error probability again to .2 (here  $\theta_2^{-1} = -1.95$ ). Notice that now the uncertainty contribution to shock prices remains positive over time. Apparently, the consumers in this economy no longer fear good news about consumption.

## 7.3 Effects of learning under rational expectations

It is interesting to contrast the kind of pessimism coming from robustness with the kind featured in Cogley and Sargent (2008) that is induced by a pessimistic prior joined with ordinary Bayesian learning. Figure 8 shows the contributions to shock prices  $\gamma \bar{G}' H + \bar{G}^{-1} [\nu_t - D(\iota)x_t]$  given in expression (13) when we assume that the true model used to price risks under full information is model  $\iota = 0$  with parameters set at values estimated at the end of our sample. Notice how the learning contribution to the shock price fluctuates between positive and negative values. These fluctuations can be interpreted in terms of alternating spells of Bayesian-learning-induced optimism and pessimism relative to what we have assumed are the true hidden state variables

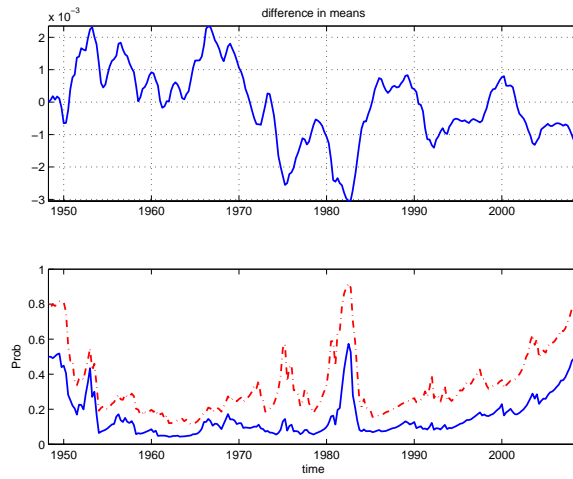


Figure 6: Difference in means (top panel) and Bayesian model probability  $\bar{\iota}_t$  (solid line) and worst-case model probability  $\tilde{\iota}_t$  (dashed line) (bottom panel). Here  $\theta_1$  is set to  $+\infty$  and  $\theta_2$  is set to give a detection error probability of .2.

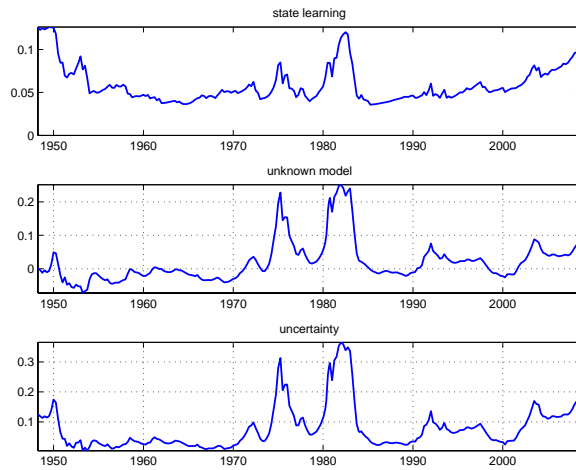


Figure 7: Contributions to uncertainty prices from learning hidden state (top panel), models known; unknown model (middle panel), and all sources (bottom panel). Here  $\theta_1$  is set to  $+\infty$  and  $\theta_2$  is set to give a detection error probability of .2. Because  $\theta_1 = +\infty$ , the contribution from unknown dynamics is identically zero.

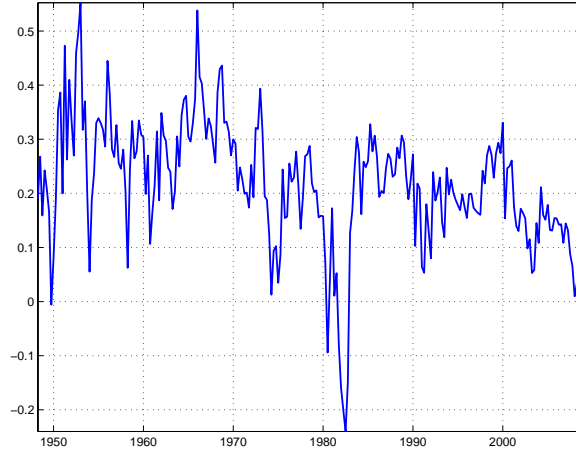


Figure 8: Contribution of learning to risk price.

with the true model.<sup>23</sup>

## 7.4 A specification with state-dependent contributions from unknown dynamics

The fact that our specification (23) implies that  $G(\iota)B(\iota)' = 0$  for  $\iota = 0, 1$  disables a potentially interesting component of uncertainty contributions in formula (22). To activate this effect, we briefly study a specification in which  $G(\iota)B(\iota)' \neq 0$  and in which its difference across the two models contributes in interesting ways. In particular, we modify (23) to the single-shock specification

$$\begin{aligned}
 dX_{1t}(\iota) &= a(\iota)X_{1t}(\iota)dt + \sigma_1(\iota)dW_t(\iota) \\
 dX_{2t}(\iota) &= 0 \\
 dY_t &= X_{1t}(\iota)dt + X_{2t}(\iota)dt + \sigma_2(\iota)dW_t(\iota)
 \end{aligned} \tag{24}$$

---

<sup>23</sup>Suppose that the state vector processes  $\{X_t(\iota)\}$  are stationary and ergodic and the associated stationary distributions are used as the prior for the limited information structures. In this case, learning is about perpetually moving targets. In long samples, the entries of  $\{X_t(\iota) - \bar{X}_t(\iota)\}$  will change signs so that on average they agree. In contrast, if one entry of  $X_t(\iota)$  is truly invariant but unknown *a priori*, then a systematic bias can emerge in a sample trajectory analogous to the one depicted in figure 8 even as the impact of the prior decays over time. For finite  $t$ 's, the expectation of  $\bar{X}_t(\iota)$  conditioned on the invariant parameter will be biased, as is standard in Bayesian analysis. This bias disappears only when we average across such trajectories using the prior over the invariant parameter.

where  $X_{1t}(\iota)$ ,  $X_{2t}(\iota)$  are again scalars and  $W_t(\iota)$  is now a scalar Brownian motion. We construct this system from the time-invariant innovations representation for system (23). This makes the second component of the state, namely, the unconditional mean of consumption growth, be known at time 0 because the (2,2) component of the steady state covariance matrix of the hidden state is zero. The model is structured so that the signal reveals the first component of the state, so with  $\iota$  known, the consumer faces no filtering problem. Therefore, the second source of uncertainty contribution vanishes and (22) simplifies to

$$\begin{aligned} & \tilde{\iota} \bar{G}^{-1} \left[ \frac{1}{\theta_1} G(1)G(1)'H + \frac{1}{\theta_1} G(1)B(1)'\lambda(1) \right] + (1 - \tilde{\iota}) \bar{G}^{-1} \left[ \frac{1}{\theta_1} G(0)G(0)'H + \frac{1}{\theta_1} G(0)B(0)'\lambda(0) \right] \\ & + (\bar{\iota} - \tilde{\iota}) \bar{G}^{-1} [D(1)\bar{x}(1) - D(0)\bar{x}(0)]. \end{aligned} \quad (25)$$

Figures 9 and 10 illustrate outcomes when we set  $\theta_1^{-1} = 3.8$ , which delivers a detection error probability of .44, and  $\theta_2^{-1} = 2$ , which delivers an overall detection error probability of .137. We chose these values to illustrate the key forces at work.<sup>24</sup> The contribution of unknown state dynamics in the top panel of figure 10 now varies over time. This reflects the difference in  $\frac{1}{\theta_1} G(\iota)B(\iota)'\lambda(\iota)$  across the two models as well as the fluctuating value of  $\tilde{\iota}$ . Notice that while the overall shock price varies, this variation is much smaller than in our previous calculations. While the current example increases the contribution from a concern about misspecified dynamics, it is also true that by ignoring robust state estimation, we have excluded much of the interesting variation in shock prices.

If we were to lower  $\theta_2$  enough to imply  $\tilde{\iota} = 1$ , then the representative consumer would act as if he puts probability one on the long-run risk model, as assumed by Bansal and Yaron (2004). Then (25) simplifies to

$$\begin{aligned} & \bar{G}^{-1} \left[ \frac{1}{\theta_1} G(1)G(1)'H + \frac{1}{\theta_1} G(1)B(1)'\lambda(1) \right] \\ & + (\bar{\iota} - 1) \bar{G}^{-1} [D(1)\bar{x}(1) - D(0)\bar{x}(0)]. \end{aligned} \quad (26)$$

The first term becomes constant, and the effect of not knowing the model contributes time-variation to the second term. The first term is captured under the Bansal and Yaron (2004) approach that has the consumer assign probability 1 to the long-run

---

<sup>24</sup>The term  $\tilde{\mu}(\iota) = -\theta_1^{-1} [G(\iota)'H + B(\iota)'\lambda(\iota)]$  is now -0.0231 for  $\iota = 0$  and -0.146 for model  $\iota = 1$ .

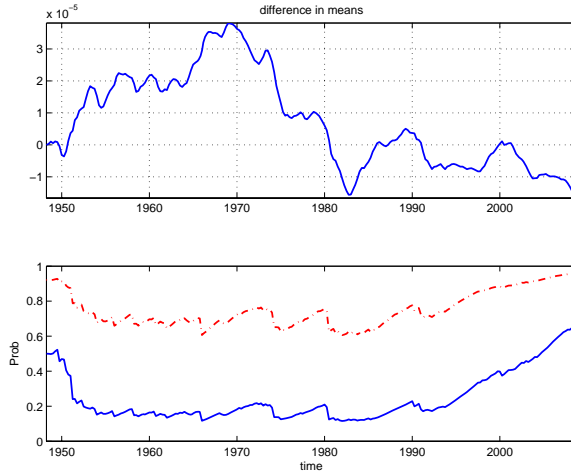


Figure 9: Difference in means (top panel) and Bayesian model probability  $\bar{\iota}_t$  (solid line) and worst-case model probability  $\tilde{\iota}_t$  (dashed line) (bottom panel). Here  $\theta_1^{-1}$  is set to give a detection error probability of .44 and  $\theta_2^{-1}$  is set to give a detection error probability of .137.

risk model, but not the second term.

## 8 Concluding remarks

The contributions of model uncertainty to shock prices combine (i) the same constant forward-looking contribution  $\tilde{\mu}(\iota) = -\theta_1^{-1}[G(\iota)'H + B(\iota)'\lambda(\iota)]$  that was featured in earlier work without learning by Hansen et al. (1999) and Anderson et al. (2003), (ii) additional smoothly decreasing in time components  $-\theta_2^{-1}\Sigma(\iota)\lambda(\iota)$  that come from learning about parameter values within models, and (iii) the potentially volatile time varying contribution highlighted in section 7.1 that reflects the consumer's robust learning about the probability distribution over models.

Our shock prices include risk and model uncertainty components, but they are the counterparts to risk prices in much of the asset pricing literature. Our mechanism for producing time-varying shock prices differs from other approaches, however. 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)

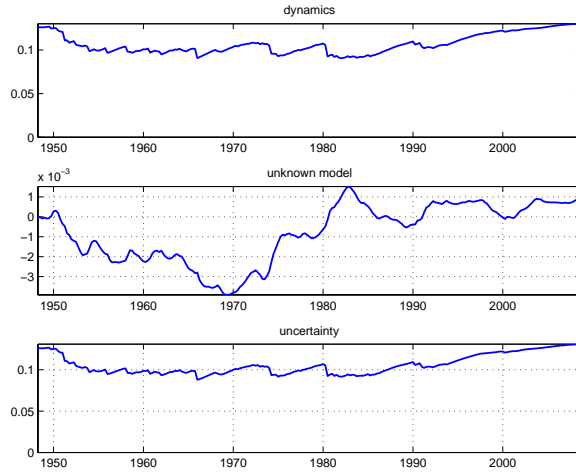


Figure 10: Contributions to uncertainty prices from unknown dynamics (top panel); unknown model (middle panel), and both sources (bottom panel). Here  $\theta_1^{-1}$  is set to give a detection error probability of .44 and  $\theta_2^{-1}$  is set to give a detection error probability of .137.

assume stochastic volatility in consumption.<sup>25</sup>

Our analysis features the effects of robust learning on local prices of exposure to uncertainty. Studying the consequences of robust learning and model selection for multi-period uncertainty prices is a natural next step. Multi-period valuation requires the compounding of local prices, and when the prices are time-varying this compounding can have nontrivial consequences.

In order to obtain convenient formulas for prices, we imposed a unitary elasticity of substitution, which implies that the ratio of consumption to wealth is constant. Consumption claims have no obvious counterpart in financial data. Dividends on aggregate measures of equity differ from aggregate consumption in important ways and the aggregate values measured in equity markets omit important components of wealth. Thus, aggregate dividend-price ratios may behave very differently from the ratio of wealth to consumption. In spite of measurement challenges, it is interesting to move beyond a stylized environment in which the consumption wealth-ratio is implied to be constant. We can accomplish this by relaxing the unitary specification for elasticity of substitution. We suspect that doing that might be important more generally for valuing durable claims with observable counterparts.

---

<sup>25</sup>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.

While our example economy is highly stylized, we can imagine a variety of other environments in which learning about low-frequency phenomena is especially challenging when consumers are not fully confident about their probability assessments. Hansen et al. (2008a) show that while long-run risk components have important quantitative impacts on low frequency implications of stochastic discount factors and cash flows, it is statistically challenging to measure those components. Belief fragility emanating from model uncertainty promises to be a potent source of fluctuations in the prices of long-lived assets.

## A Detection error probabilities

By adapting procedures developed by Hansen et al. (2002) and Anderson et al. (2003) in ways described by Hansen et al. (2008b), we can use simulations to approximate a detection error probability. Repeatedly simulate  $\{Y_{t+1} - Y_t\}_{t=1}^T$  under the approximating model. Evaluate the likelihood functions  $L_T^a$  and  $L_T^w$  of the benchmark model and worst-case model for a given  $(\theta_1, \theta_2)$ . Compute the fraction of simulations for which  $\frac{L_t^w}{L_t^a} > 1$  and call it  $r_a$ . This approximates the probability that the likelihood ratio says that the worst-case model generated the data when the approximating model actually generated the data. Do a symmetrical calculation to compute the fraction of simulations for which  $\frac{L_t^a}{L_t^w} > 1$  (call it  $r_w$ ), where the simulations are generated under the worst case model. As in Hansen et al. (2002) and Anderson et al. (2003), define the overall detection error probability to be

$$p(\theta_1, \theta_2) = \frac{1}{2}(r_a + r_w). \quad (27)$$

Because in this paper we use what Hansen et al. (2008b) call Game I, we use the following sequential procedure to calibrate  $\theta_1$  first, then  $\theta_2$ . First, we pretend that  $x_t(\iota)$  is observable for  $\iota = 0, 1$  and calibrate  $\theta_1$  by calculating detection error probabilities for a system with an observed state vector using the approach of Hansen et al. (2002) and Hansen and Sargent (2008, ch. 9). Then having pinned down  $\theta_1$ , we use formula (27) to calibrate  $\theta_2$ . This procedure takes the point of view that  $\theta_1$  measures how difficult it would be to distinguish one model of the partially hidden state from another if we were able to observe the hidden state, while  $\theta_2$  measures how difficult it is to distinguish alternative models of the hidden state. The probability  $p(\theta_1, \theta_2)$  measures both sources of model uncertainty.

We proceeded as follows. (i) Conditional on model  $\iota$  and the model  $\iota$  state  $x_t(\iota)$  being observed, we computed the detection error probability as a function of  $\theta_1$  for models  $\iota = 0, 1$ . (ii) Using a prior probability of  $\pi = .5$ , we averaged the two curves described in point (1) and plotted the average against  $\theta_1$ . We calibrated  $\theta_1$  to yield an average detection error probability of .4 and used this value of  $\theta_1$  in the next step. (iii) With  $\theta_1$  locked at the value just set, we then calculated and plotted the detection error for the mixture model against  $\theta_2$ . To generate data under the approximating mixture model, we sampled sequentially from the conditional density of signals under the mixture model, building up the Bayesian probabilities  $\bar{\iota}_t$  sequentially along a sample path. Similarly, to generate data under the worst case mixture model, we sampled sequentially from the conditional density for the worst-case signal distribution, building up the worst-case model probabilities  $\tilde{\iota}_t$  sequentially. We set  $\theta_2$  to fix the overall detection error equal to .2.

## B Sensitivity analysis

This appendix spotlights the force that produces countercyclical uncertainty contributions to shock prices by introducing a perturbation to our model that attenuates that force. The persistent countercyclical uncertainty contributions to shock prices in figure 8 come from a setting in which the representative consumer entertains two models that are difficult to distinguish. We study how uncertainty contributions to shock prices change when we expand the consumer's universe of models to include ones that fit the data even better than the two models in section 7. In particular, we now endow the model with seven models having the same value of  $\sigma_2$  but now with values of  $\rho = .36, .52, .67, .82, .89, .95, .99$ , with values of  $\sigma_1, \mu_y$  being concentrated out via likelihood function maximization. In terms of the likelihood function for the whole sample, the values at the end values  $.36, .99$  are the poorest fitting ones and the  $\rho = .82$  one is the best fitting. We start the representative consumer with a uniform prior over the seven models and set  $\theta_1^{-1} = 6.85, \theta_2^{-1} = .8$  (these are set to give the same detection error probabilities of  $.4$  for the state dynamics and  $.2$  over all that we used in the text) to obtain the uncertainty contribution to shock prices reported in figure 11. We report Bayesian model probabilities and their worst-case counterparts in figure 12.

Countercyclical shock prices still emerge, but they are moderated relative to those in figure 8 in the text. The reason is to be found in how the presence of models that fit better eventually pushes down the Bayesian model probability on long-run risk model. Pushing that probability down far enough diminishes its influence on uncertainty contributions to shock prices even in the face of the tendency to twist model probabilities toward the long-run risk model. Even after twisting, the worst-case probabilities on that model are much smaller than they were in figure 1.

We find the comparison among competing models that have disparate implications for shock prices as featured in our paper to be interesting. While adding our models with  $\rho$ 's between  $.36$  and  $.99$  that give higher values of the likelihood diminishes the variation in contributions to uncertainty that we have computed, the impact on the rational expectations counterpart can be even more dramatic because those intermediate models fit the data in ways that imply substantially smaller shock prices. For the recursive utility model with full information, the magnitude of the shock price is  $\gamma|[B(\iota)' \lambda(\iota) + G(\iota)' H]|$  where  $\gamma$  is a measure of risk aversion. For the model with the largest likelihood,  $|[B(\iota)' \lambda(\iota) + G(\iota)' H]| = 0.008622$ , while this magnitude is  $0.039627$  for the  $\rho = .99$  model. Thus, a value of  $\gamma$  more than four times larger is required at the maximum likelihood estimate for the magnitude of the shock price to remain the same as for the model with  $\rho = .99$ . To the extent that  $\rho = .99$  is statistically implausible, a rational expectations econometrician either rejects the model or finds that consumers are highly risk averse.

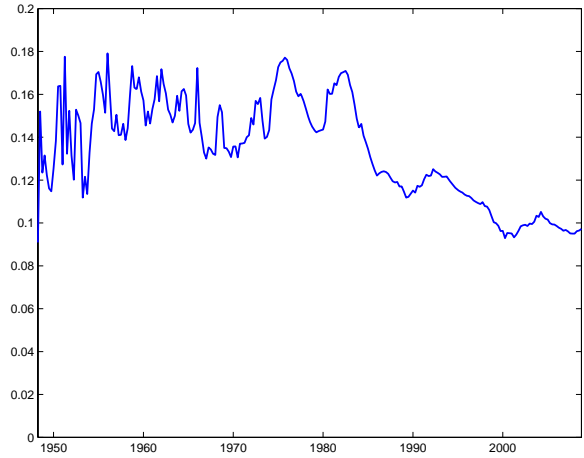


Figure 11: Contribution of model uncertainty to risk price with seven models.

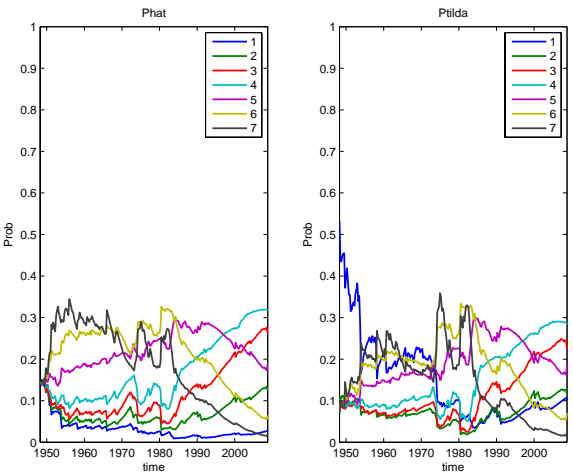


Figure 12: Bayesian and worst case model probabilities with seven models.

## 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. 2008. Doubts or Variability? Forthcoming in *Journal of Economic Theory*.
- Bossaerts, Peter. 2002. *The Paradox of Asset Pricing*. Princeton, NJ: Princeton University Press.
- . 2004. Filtering returns for unspecified biases in priors when testing asset pricing theory. *Review of Economic Studies* 71:6386.
- Breeden, Douglas T. 1979. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics* 7 (3):265 – 296.
- Brennan, M. J. and Y. Xia. 2001. Stock Price Volatility and Equity Premium. *Journal of Monetary Economics* 47:249–283.
- 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.
- Cerreia, Simone, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio. 2008. Uncertainty Averse Preferences. Working paper, Collegio Carlo Alberto.
- Cogley, Timothy and Thomas J. Sargent. 2008. The Market Price of Risk and the Equity Premium: A Legacy of the Great Depression? *Journal of Monetary Economics* 55:454–478.

- Cogley, Timothy, Riccardo Colacito, Lars Peter Hansen, and Thomas J. Sargent. 2008. Robustness and U.S. Monetary Policy Experimentation. *Journal of Money, Credit and Banking* 40 (8):1599–1623.
- Croce, Mariano M., Martin Lettau, and Sydney C. Ludvigson. 2008. Investor Information, Long-Run Risk, and the Duration of Risky Cash Flows. NBER WP 12912.
- 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.
- Duffie, D. and L. G. Epstein. 1992. Stochastic Differential Utility. *Econometrica* 60 (2):353–394.
- Duffie, Darrell. 2001. *Dynamic Asset Pricing Theory, Third Edition*. Princeton, New Jersey: Princeton University Press.
- Epstein, Larry G. and Martin Schneider. 2008. Ambiguity, Information Quality, and Asset Pricing. *Journal of Finance* 63 (1):197–228.
- Gilboa, I. and D. Schmeidler. 1989. Maxmin Expected Utility with Non-Unique Prior. *Journal of Mathematical Economics* 18:141–153.
- Hansen, Lars Peter. 2007. Beliefs, Doubts and Learning: Valuing Macroeconomic Risk. *American Economic Review* 97 (2):1–30.
- . 2008. Modeling the Long Run: Valuation in Dynamic Stochastic Economies. Presented as the Fisher-Schultz Lecture to the Econometric Society.
- 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 Scott F. Richard. 1987. The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica* 55:587–614.

- Hansen, Lars Peter and Thomas J. Sargent. 1980. Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control* 2:7 – 46.
- . 2001. Robust Control and Model Uncertainty. *American Economic Review* 91:60–66.
- . 2007. Recursive robust estimation and control without commitment. *Journal of Economic Theory* 136 (1):1–27.
- . 2008. *Robustness*. Princeton, New Jersey: Princeton University Press.
- Hansen, Lars Peter and Jose Scheinkman. 2009. Long-term Risk: An Operator Approach. *Econometrica* 77 (1):177–234.
- 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.
- Hansen, Lars Peter, Thomas J. Sargent, Gauhar A. Turmuhambetova, and Noah Williams. 2006. Robust Control, Min-Max Expected Utility, and Model Misspecification. *Journal of Economic Theory* 128:45–90.
- Hansen, Lars Peter, John Heaton, and Nan Li. 2008a. Consumption Strikes Back: Measuring Long Run Risk. *Journal of Political Economy* 116 (2):260–302.
- Hansen, Lars Peter, Ricardo Mayer, and Thomas J. Sargent. 2008b. Robust Hidden Markov LQG Problems. University of Chicago and New York University.
- Harrison, 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, Peter, Massimo Marinacci, and Sujoy Mukerji. 2005. A Smooth Model of Decision Making under Ambiguity. *Econometrica* 73 (6):1849–1892.

- . 2008. Recursive Smooth Ambiguity Preferences. Forthcoming in the Journal of Economic Theory.
- Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini. 2006a. Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica* 74 (6):1447–1498.
- . 2006b. Dynamic Variational Preferences. *Journal of Economic Theory* 128:4–44.
- Scheinkman, Jose A. and Wei Xiong. 2003. Overconfidence and Speculative Bubbles. *Journal of Political Economy* 111 (6):1183–1220.
- Shiller, Robert. 1972. *Rational Expectations and the Term Structure of Interest Rates*. Ph.D. thesis, Massachusetts Institute of Technology.
- Strzalecki, Tomasz. 2008. Axiomatic Foundations of Multiplier Preferences. Northwestern University Department of Economics.
- 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.