Quantitative Economics 1 (2010), 129–162 1759-7331/20100129

Fragile beliefs and the price of uncertainty

Lars Peter Hansen
University of Chicago

Thomas J. Sargent
New York University and Hoover Institution

A representative consumer uses Bayes' law to learn about parameters of several models and to construct probabilities with which to perform ongoing model averaging. The arrival of signals induces the consumer to alter his posterior distribution over models and parameters. The consumer's specification doubts induce him to slant probabilities pessimistically. The pessimistic probabilities tilt toward a model that puts long-run risks into consumption growth. That contributes a countercyclical history-dependent component to prices of risk.

Keywords. Learning, Bayes' law, robustness, risk sensitivity, pessimism, prices of risk.

JEL classification. C11, C44, C72, E44, G12.

Le doute n'est pas une condition agréable, mais la certitude est absurde.[1]
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 misspecification foster pessimism that puts countercyclical uncertainty premia into risk prices.

1.1 *Doubts promote fragile beliefs*

A representative consumer values consumption streams according to the multiplier preferences that Hansen and Sargent (2001) used to represent model uncertainty.[2] Fol-

Lars Peter Hansen: l-hansen@uchicago.edu
Thomas J. Sargent: ts43@nyu.edu
We thank Gadi Barlevy, Alberto Bisin, Jaroslav Borovicka, Riccardo Colacito, Mark Gertler, Anastasios Karantounias, Ricardo Mayer, Tomasz Piskorski, Grace Tsiang, Gianluca Violante, Amir Yaron, and three referees for helpful comments on earlier drafts of this paper. We thank Francisco Barillas for excellence in executing the computations. We thank the National Science Foundation for research support under separate grants to Hansen and Sargent.

[1] Doubt is not a pleasant condition, but certainty is absurd.

[2] The relationship of the multiplier preferences of Hansen and Sargent (2001) to the max–min expected utility preferences of Gilboa and Schmeidler (1989) was analyzed by Hansen, Sargent, Turmuhambetova,

DOI: 10.3982/QE9

lowing Hansen and Sargent (2007), the iterated application of risk-sensitivity operators focuses a representative consumer's distrust on model selection and on particular parameters within those models.[3] Ex post, the consumer acts "as if" he believes a probability measure that a malevolent alter ego has twisted pessimistically relative to a baseline approximating model. The apparent pessimism is actually the consumer's instrument for constructing valuations that are robust to misspecifications. By "fragile beliefs" we refer to the sensitivity of pessimistic probabilities to the arrival of news, as determined by the state-dependent value functions that define what the consumer is pessimistic *about*.[4] Our representative consumer's reluctance to trust his model adds "model uncertainty premia" to prices of risk. The parameter estimation and model selection problems make these uncertainty prices be time-dependent and state-dependent, in contrast to the constant uncertainty premia found by Hansen, Sargent, and Tallarini (1999) and Anderson, Hansen, and Sargent (2003).

1.2 *Key components*

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

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–2009IV. In one model, consumption growth rates are only weakly serially correlated, while 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). The representative consumer doubts the Bayesian model-mixing probabilities as well as the specification of each model. The consumer copes with model uncertainty by slanting probabilities toward the model associated with the lowest continuation utility. We show how variations over time in the probabilities attached to models and other state variables put volatility into uncertainty premia.

and Williams (2006), Maccheroni, Marinacci, and Rustichini (2006a, 2006b), Cerreia, Maccheroni, Marinacci, and Montrucchio (2008), and Strzalecki (2008).

[3]Sometimes the literature calls this structured uncertainty.

[4]Harrsion and Kreps (1978) and Scheinkman and Xiong (2003) explored other settings in which difficult to detect departures from rational expectations lead to interesting asset price dynamics that cannot occur under rational expectations.

[5]This second risk-sensitivity operator expresses what Klibanoff, Marinacci, and Mukerji (2005, 2009) called smooth ambiguity and what other researchers call structured model uncertainty. Using a different approach to learning in the presence of model ambiguity, Epstein and Schneider (2008) applied 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.

[6]Barillas, Hansen, and Sargent (2009) emphasized that both risk-sensitivity operators can be view as indirect utility functions for the minimization part for max–min expected utility problems. Tallarini (2000) adopted an interpretation in terms of enhanced risk aversion.

In contrast, Bansal and Yaron assumed that the representative consumer assigns probability 1 to the long-run risk model even though sample evidence is indecisive in selecting between them.[7] Our framework explains why a consumer might act *as if* he puts probability (close to) 1 on the long-run risk model even though he knows that it is difficult to discriminate between these models statistically.

1.3 *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 Brownian motions $W_t(\iota)$, $\bar{W}_t(\iota)$, and $\bar{W}_t$ that are implied by these three information structures. Section 4 then uses these three information specifications and the associated $dW_t(\iota)$, $d\bar{W}_t(\iota)$, and $d\bar{W}_t$, respectively, as 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 asset pricing models with 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 *time-dependent* premia and uncertainty about models contributes *state-dependent* premia. 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. The Appendix describes how we use detection error probabilities to calibrate the representative consumer's concerns about model misspecification.

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 intermediate dates implies that $S_{t+\tau}/S_t$ is the τ-period stochastic discount factor for computing asset prices 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

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

[7] Bansal and Yaron (2004) incorporated other features in their specification of consumption dynamics, including stochastic volatility, and they adopted a recursive utility specification with an intertemporal elasticity of substitution greater than 1.

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

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

The vector of (growth-rate) risk prices for horizon τ is given by the price elasticity

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

where we have scaled by the payoff horizon τ for comparability.[8] 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 α_o. This is true more generally when the investment horizon shrinks to zero.[9]

The vector of local risk prices is given by the limit

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

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 elementary 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 τ using the dynamics of the underlying state variables.

We can exploit local normality to obtain a simple characterization of the slope of 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 $\alpha_t^* = \pi_t/|\pi_t|$ with the optimized local mean being

$$\alpha_t^*\cdot\pi_t = \frac{\pi_t\cdot\pi_t}{|\pi_t|} = |\pi_t|. \tag{4}$$

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

In the power utility model,

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

[8]The negative sign reflects that the consumer dislikes risk.

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

where the growth rate of log consumption is $\log C_{t+\tau} - \log C_t$. Here the vector π_t of local risk prices is the vector of exposures of $-d\log S_t = \delta\, dt + \gamma\, d\log C_t$ 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 ι, $Y_t^{t+\tau}$ 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. Lowercase letters denote values that potentially can be realized. In particular, 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 ι. The hidden Markov structure induces probability densities $f(y|\iota, x)$, $g(x|\iota, \mathcal{Y}_t)$, $h(\iota|\mathcal{Y}_t)$, and $\bar{f}(y|\mathcal{Y}_t)$.[10] Evidently,

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

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 can 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

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

This decomposition helps us understand how our paper relates to earlier asset pricing papers including, for example, Detemple (1986), Dothan and Feldman (1986), David (1997), Veronesi (2000), Brennan and Xia (2001), Ai (2006), David (2008), Croce, Lettau, and Ludvigson (2008), and David and Veronesi (2009)[11] that used 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

[10] 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 $[t, t+\tau]$.

[11] 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, Colacito, Hansen, and Sargent (2008) applied the framework of Hansen and Sargent (2007) in a setting where decisions affect future probabilities of hidden states and therefore experimentation is active. The papers just cited price risks un-

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 such learning models depend only on $\bar{f}$ and not on the underlying structure with hidden states that the model builder used to construct 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, and h means that equivalence is not true in our model because it makes asset prices depend on the evolution of the hidden states themselves 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.[12]

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

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

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

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

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

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

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

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

der 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), David (2008), 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.

[12]As emphasized by Hansen (2007), by exploring these misspecifications, our representative consumer in effect refuses to reduce compound lotteries.

where the likelihood ratio $M_t^{t+\tau}$ satisfies $E(M_t^{t+\tau}|\mathcal{Y}_t) = 1$ and can be decomposed as

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

In Section 6, we show how to represent the three relative densities $\frac{\hat{f}}{f}$, $\frac{\hat{g}}{g}$, and $\frac{\hat{h}}{h}$, 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 set forth in Section 3, changes in probability measures can conveniently be 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 $\{M_t^i\}$ 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 both on the signal history and the 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, so $W_{t+\tau}(\iota) - W_t(\iota) \sim \mathcal{N}(0, \tau I)$. For notational simplicity, we suppose that the same Brownian motion drives both models. Under full information, ι 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_τ up to t and any prior information available as of date zero. In what follows, we first condition on $\mathcal{Y}_t$ and ι for each t. We then omit ι and condition only on $\mathcal{Y}_t$.

3.2.1. *Innovations representation with model known*

First, suppose that ι is known. Application of the Kalman filter yields the innovations representation

$$d\bar{X}_t(\iota) = A(\iota)\bar{X}_t(\iota)\,dt + K_t(\iota)[dY_t - D(\iota)\bar{X}_t(\iota)\,dt], \tag{10}$$

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

$$\begin{aligned} K_t(\iota) &= [B(\iota)G(\iota)' + \Sigma_t(\iota)D(\iota)'][G(\iota)G(\iota)']^{-1}, \\ \frac{d\Sigma_t(\iota)}{dt} &= A(\iota)\Sigma_t(\iota) + \Sigma_t(\iota)A(\iota)' + B(\iota)B(\iota)' \\ &\quad - K_t(\iota)[G(\iota)B(\iota)' + D(\iota)\Sigma_t(\iota)]. \end{aligned} \tag{11}$$

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*

When there are different $G(\iota)G(\iota)'$'s for different models ι, it is statistically trivial to distinguish among models ι with continuous data records. Technically, the reason is that with different $G(\iota)G(\iota)'$'s, the distinct ι models fail to be mutually absolutely continuous over finite intervals, so one model puts positive probability on events that are certain to be observed over a finite interval and on which the other model puts zero probability. Because we want the models to be difficult to distinguish statistically, we assume that $G(\iota)G(\iota)'$ is independent of ι. 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

$$\nu_t \doteq [\bar{\iota}_t D(1)\bar{X}_t(1) + (1-\bar{\iota}_t)D(0)\bar{X}_t(0)]. \tag{12}$$

Then

$$d\bar{\iota}_t = \bar{\iota}_t(1-\bar{\iota}_t)[\bar{X}_t(1)'D(1)' - \bar{X}_t(0)'D(0)'](\bar{G}')^{-1}\,d\bar{W}_t. \tag{13}$$

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, and (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.[13] 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 show how 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.

5. A full-information perspective on agents' learning

In this section, we describe how to link our paper to other papers on learning and asset prices (e.g., Bossaerts (2002, 2004), David (2008), and Cogley and Sargent (2008)). We think of these papers as studying what happens when an econometrician mistakenly assumes that consumers have a larger information set than they actually do. From Hansen

Table 1. The parameter γ 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 $d\log C_t = H'\,dY_t$. When the model is unknown, $G(\iota)G(\iota)'$ is assumed to be independent of ι.

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	$d\bar{W}_t$	$\gamma\bar{G}'H$	$\gamma\sqrt{H'G(\iota)G(\iota)'H}$

[13]To look at Bayesian learning from another angle, in Section 5 we price the three risk vectors under full information.

and Richard (1987), we know that an econometrician who conditions on *less* information than consumers still draws correct inferences about the magnitude of risk prices. Our message here is that an econometrician who mistakenly conditions on *more* information than consumers makes false inferences about the magnitude of risk prices. We regard the consequences of an econometrician's mistaken conditioning on more information than consumers as contributing to the analysis of risk pricing under consumers' Bayesian learning.

To elaborate on the preceding points, 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 include at least 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.[14] It is true that the econometrician who omits information fails to infer correctly the risk components actually confronted by consumers. But 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 want to 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 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

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

[14]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.

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

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

where δ is the subjective rate of discount. Multiplying by -1 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. The second term is the "mistake" in pricing contributed by the econometrician's error in attributing to consumers a larger information set than they actually have.[15]

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

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

[15]Although our illustrative application in Section 7 uses only consumption growth as a signal, our formulation allows multiple signals. Our application does not use asset prices as signals, but it would be interesting to do so. In standard rational expectations models in the tradition of Lucas (1978) (where agents do not glean information from equilibrium prices as in the rational expectations models described by Grossman (1981)), cross-equation restrictions link asset prices to the dynamics governing macroeconomic fundamentals. These cross-equation restrictions typically presume that investors know parameters governing the macrodynamics. To avoid stochastic singularity, econometric specifications introduce hidden states (including hidden information states) or "measurement errors." In such rational expectations models, prices reveal to an econometrician the information used by economic agents. Rational expectations models that incorporate agents learning about states hidden to them, possibly including parameters of the macrodynamics, were constructed and estimated, for example, by David and Veronesi (2009), who also confronted stochastic singularity in the ways just mentioned. With or without learning, the cross-equation restrictions in such models would be altered if agents were forced to struggle with misspecification as they do in the model of Section 7. In that illustrative application, we have not taken the extra steps that would be involved in confronting stochastic singularity.

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

$$\gamma \bar{G}' H + \bar{G}^{-1}[\nu_t - D(\iota) X_t(\iota)], \tag{14}$$

where ν_t is defined in (12).

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 ι 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 ι and the particular sample path. Thus, Cogley and Sargent (2008) imputed a pessimistic prior to the representative consumer so as to generate a slowly evaporating U.S. equity premium after WWII.

In subsequent sections, 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, (1) 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.[16] 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 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.

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

[16] Hansen and Sargent (2008, Chaps. 11–13) discussed the role of the planner's problem in computing and representing prices with which to confront a representative consumer.

TABLE 2. When the model is unknown, $G(\iota)G(\iota)'$ is assumed to be independent of ι. 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.

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	$d\bar{W}_t$	$\bar{G}'H$	$(\tilde{\iota} - \bar{\iota})\bar{G}^{-1}[D(1)\bar{x}(1) - D(0)\bar{X}(0)]$

function

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

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

$$0 = -\delta\lambda(\iota) + D(\iota)'H + A(\iota)'\lambda(\iota), \tag{15}$$

so that

$$\lambda(\iota) = [\delta I - A(\iota)']^{-1}D(\iota)'H. \tag{16}$$

The form of the value function is the same as that of Tallarini (2000) and Barillas, Hansen, and Sargent (2009). 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 the sense that the decision-maker distrusts it 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 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 T^1 operator of Hansen and Sargent (2007)[17]:

[17]Suppose that the decision-maker has instantaneous utility function $u(x)$, positive discount rate δ, 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

PROBLEM 6.1. *A value function $\lambda(\iota)\cdot x+\kappa(\iota)+c$ satisfies the Bellman equation*

$$0=\min_{\mu}-\delta[\lambda(\iota)\cdot x+\kappa(\iota)]+x'D(\iota)'H+\mu'G(\iota)'H+x'A(\iota)'\lambda(\iota)$$
$$+\mu'B(\iota)'\lambda(\iota)+\frac{\theta_1}{2}\mu'\mu. \tag{17}$$

Here the vector $\mu\,dt$ is a drift distortion to the mean of $dW_t(\iota)$ and θ_1 is a positive penalty parameter that characterizes the decision-maker's fear that model ι is misspecified. We impose the same θ_1 for both models. See Hansen et al. (2006) and Anderson, Hansen, and Sargent (2003) for more general treatments and see the Appendix for how we propose to calibrate θ_1 via detection error probabilities. The minimizing drift distortion

$$\tilde{\mu}(\iota)=-\frac{1}{\theta_1}[G(\iota)'H+B(\iota)'\lambda(\iota)] \tag{18}$$

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

$$\kappa(\iota)=-\frac{1}{2\theta_1\delta}|G(\iota)'H+B(\iota)'\lambda(\iota)|^2. \tag{19}$$

Equating coefficients on x in (17) implies that equation (15) continues to hold. Thus, $\lambda(\iota)$ remains the same as in the model without robustness and is given by (16).

PROPOSITION 6.2. *The value function shares the same $\lambda(\iota)$ with the expected utility model [formula (15)] and $\kappa(\iota)$ is given by (19). The associated worst-case distribution for the Brownian increment is normal with covariance matrix $I\,dt$ and drift $\tilde{\mu}(\iota)dt$ given by (18).*

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 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, and evolves according to

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

a multiplier problem satisfies the Bellman equation $\delta V(x)=\min_h(u(x)+\frac{\theta}{2}h'h+[\nu(x)+\sigma(x)h]V_x(x)+\frac{1}{2}\operatorname{trace}[\sigma(x)'V_{xx}\sigma(x)])$. The indirect value function for this problem satisfies the Bellman equation $\delta S(x)=u(x)+\nu(x)S_x(x)+\frac{1}{2}\operatorname{trace}[\sigma(x)'S_{xx}(x)\sigma(x)]-\frac{1}{2\theta}S_x(x)'\sigma(x)\sigma(x)'S_x(x)$, which is an example of the stochastic differential utility model of Duffie and Epstein (1992). See Hansen et al. (2006).

The vector of local shock price is again the negative of the exposure of the stochastic discount factor to the respective shocks. With robustness, the shock price vector $G(\iota)'H$ under full information is augmented by an uncertainty price:

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

Notice the presence of the forward-looking term $\lambda(\iota)$ from (16) in the term that we have labeled "uncertain dynamics." Neither the risk contribution nor the uncertainty contribution to the shock prices is 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 ι) is known but that the state $X_t(\iota)$ is not. We want $\frac{\hat{g}}{g}$ in formula (8) and proceed by seeking a martingale $\{M_t^i\}$ to use under this information structure.

Without concerns about misspecification, the estimates $\bar{x}$ of the state and the covariance matrix Σ used to construct ψ at a given point in time will typically depend on the model ι. The laws of motion for $\bar{x}(\iota)$ and $\Sigma(\iota)$ are (10) and (11), respectively.

Following Hansen and Sargent (2007), we introduce a positive penalty parameter θ_2 and construct a robust estimate of the hidden state $X_t(\iota)$ by solving the following problem cast in terms of objects constructed in Section 3.2.1:[18]

Problem 6.3. *The robust relative density for the hidden state solves*

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

where $\psi(x|\bar{x},\Sigma)$ is the normal density with mean $\bar{x}$ and covariance matrix Σ, and the minimization on the first line is subject to $\int \phi(x)\psi(x|\bar{x},\Sigma)\,dx = 1$.

In the first line of Problem 6.3, ϕ is a density (relative to a normal) that distorts the density ψ for the hidden state and θ_2 is a positive penalty parameter that penalizes ϕ's with large relative entropy (the expected value of $\phi \log \phi$). The second line of Problem 6.3 exploits the outcome that with a linear value function, 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 estimate $\tilde{x}$ solves

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

[18]In the Appendix, we describe how to use statistical detection error probabilities to calibrate the penalty parameter θ_1 as well as another penalty parameter θ_2 to be introduced below.

PROPOSITION 6.4. *The robust value function is*

$$U[\iota, \bar{x}, \Sigma] = \lambda(\iota) \cdot \bar{x} + \kappa(\iota) - \frac{1}{2\theta_2}\lambda(\iota)'\Sigma\lambda(\iota) \tag{20}$$

with the same $\lambda(\iota)$ as in the expected utility model [formula (15)] and the same $\kappa(\iota)$ as in the robust planner's problem with full information [formula (19)]. The worst-case state estimate is

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

The indirect value function on the right side of (20) defines an instance of the 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 that comes from the potential misspecification of the state dynamics as reflected in the drift distortion in the Brownian motion, and the other in the second row of Table 2 that comes from the filtering problem as reflected in a distortion to the estimated mean of the hidden state vector:

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

The state estimation adds time dependence to the uncertainty prices through the evolution of the covariance matrix $\Sigma_t(\iota)$ governed by (11), but adds nothing 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 adjusts for not trusting the benchmark distribution over unknown models, thus constructing $\frac{\hat{h}}{h}$ in formula (8). We do this by twisting the model probability $E(\iota|\mathcal{Y}_t) = \bar{\iota}_t$.

PROBLEM 6.5. *We twist the model probability by sowing*

$$\min_{0\le\iota\le 1} \iota U[1,\bar{x}(1),\Sigma(1)] + (1-\iota)U[0,\bar{x}(0),\Sigma(0)] \\ + \theta_2\iota[\log\iota - \log\bar{\iota}] + \theta_2(1-\iota)[\log(1-\iota) - \log(1-\bar{\iota})].$$

PROPOSITION 6.6. *The indirect value function for Problem 6.5 is the robust value function*[19]

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

The worst-case model probabilities satisfy

$$(1-\tilde{\iota}) \propto (1-\bar{\iota})\exp\left(-\frac{U[0,\bar{x}(0),\Sigma(0)]}{\theta_2}\right), \tag{21}$$

$$\tilde{\iota} \propto \bar{\iota}\exp\left(-\frac{U[1,\bar{x}(1),\Sigma(1)]}{\theta_2}\right), \tag{22}$$

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 that we derived earlier 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

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

and the stochastic discount factor is governed by

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

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 that comes from concerns about model misspecification:

$$\tilde{\iota}\bar{G}^{-1}\left[\frac{1}{\theta_1}G(1)G(1)'H + \frac{1}{\theta_1}G(1)B(1)'\lambda(1)\right] \\ + (1-\tilde{\iota})\bar{G}^{-1}\left[\frac{1}{\theta_1}G(0)G(0)'H + \frac{1}{\theta_1}G(0)B(0)'\lambda(0)\right] \tag{23}$$

[19]This is evidently another application of the T^2 operator of Hansen and Sargent (2007).

$$+\tilde{\iota}\bar{G}^{-1}\left[\frac{1}{\theta_2}D(1)\Sigma(1)\lambda(1)\right]+(1-\tilde{\iota})\bar{G}^{-1}\left[\frac{1}{\theta_2}D(0)\Sigma(0)\lambda(0)\right]$$
$$+(\bar{\iota}-\tilde{\iota})\bar{G}^{-1}[D(1)\bar{x}(1)-D(0)\bar{x}(0)].$$

As summarized in Table 2, the term on the first line reflects uncertainty in *state dynamics* associated with each of the two models. Hansen, Sargent, and Tallarini (1999) featured a similar term. It is forward looking by virtue of the appearance of $\lambda(\iota)$ determined in (16). The term on the second line reflects uncertainty about *hidden states* in each of the two models. Notice that both of these terms depend on $\tilde{\iota}$, so the probability distortion across models impacts their construction. In the limiting case that $\tilde{\iota}=1$, the term on the first line is constant over time and the term on the second line depends on time through $\Sigma(1)$ but not on the signal history. In our application, this limiting case obtains approximately when the penalty parameter θ_2 is sufficiently small. The term on the third line reflects uncertainty about the *models* and depends on the signal history even when $\tilde{\iota}=1$. The component $\bar{G}^{-1}[D(1)\bar{x}(1)-D(0)\bar{x}(0)]$ also drives the evolution of model probabilities given in (13) 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 from new information in the signals to the updated probability assigned to model $\iota=1$. The signal realizations over the next instant alter the decision-maker's posterior probability $\bar{\iota}$ on model 1 as well as his worst-case probability $\tilde{\iota}$, and this is reflected in the equilibrium uncertainty prices. This response vector will recurrently change signs so that new information will not always move $\bar{\iota}$ in the same direction. In the term on the third line of (23), this response vector is scaled by the difference between the current model probabilities under the benchmark and worst-case models. Formulas (21) and (22) indicate how the consumer slants probabilities toward the model with the lower utility. This probability slanting induces additional history dependence because $\bar{\iota}_t$ depends on the signal history.

7. Illustrating the mechanism

To highlight forces that govern the contributions of our three components of model uncertainty to shock prices in formula (23), we create two models $\iota=0,1$, with model $\iota=1$ being a long-run risk model with a predictable growth rate along the lines of Bansal and Yaron (2004) and Hansen, Heaton, and Li (2008). Our models share the form $d\log C_t=dY_t$ and

$$\begin{aligned}
dX_{1t}&=a(\iota)X_{1t}(\iota)\,dt+b_1(\iota)\,dW_{1t}(\iota),\\
dX_{2t}&=0,\\
dY_t&=X_{1t}\,dt+X_{2t}\,dt+g_2(\iota)\,dW_{2t}(\iota),
\end{aligned}\tag{24}$$

where $X_{1t}(\iota)$ and $X_{2t}(\iota)$ are scalars, $W_{1t}(\iota)$ and $W_{2t}(\iota)$ are scalar components of the vector Brownian motion $W_t(\iota)$, $X_{20}(\iota)=\mu_y(\iota)$ is the unconditional mean of consumption growth for model ι, and $a(\iota)\equiv\rho(\iota)-1$. We set $\tau=1$ in the following discrete-time ap-

proximation to the state-space system (9):

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

Additionally we set

$$A(\iota) = \begin{bmatrix} \rho(\iota) - 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad B(\iota) = \begin{bmatrix} b_1(\iota) & 0 \\ 0 & 0 \end{bmatrix},$$
$$D(\iota) = [1 \quad 1], \qquad G(\iota) = [0 \quad g_2(\iota)].$$

A small negative $a(\iota)$ (i.e., an autoregressive coefficient $\rho(\iota)$ close to unity) coupled with a small $b_1(\iota)$ captures long-run risks in consumption growth. Bansal and Yaron (2004) justified such a specification with the argument that it fits consumption growth approximately as well as, and is therefore difficult to distinguish from, an independent and identically distributed (iid) consumption growth model, which we know fits the aggregate per capita U.S. consumption data well. In the spirit of their argument, we form two models with the *same* values of the signal noise $b_2(\iota)$ but that, with *different* values of $b_1(\iota)$, $\rho(\iota) \equiv a(\iota) + 1$, and $\mu_y(\iota) = X_{20}(\iota)$, give *identical* values for the likelihood function. In particular, with our setting of the initial model probability $\bar{\iota}_0$ at .5, the terminal value of $\bar{\iota}_t$ also equals .5, so the two models are indistinguishable statistically over our complete sample. This is our way of making precise the Bansal and Yaron (2004) observation that the long-run risk with a model with high serial correlation in consumption growth and a model with low serial correlation in consumption growth is difficult to distinguish empirically. We impose $\rho(1) = .99$ as our long-run risk model, while the equally good fitting $\iota = 0$ model with low serial correlation in consumption growth has $\rho(0) = .4993$.[20]

7.1 *Calibrating θ_1 and θ_2*

In the Appendix we describe how we first calibrated θ_1 to drive the average detection error probability over the two ι models with observed states to be .4 and then, with θ_1 thereby fixed, set θ_2 to get a detection error probability of .2 for the signal distribution of the mixture model. This is one of a frontier of configurations of θ's that imply the same detection error probability of .02. We use this particular combination for illustration and explore another combination below. We regard the overall value of the

[20]The sample for real consumption of services and nondurables runs over the period 1948II–2009IV. To fit model $\iota = 1$, we fixed $\rho = .99$ and estimated $b_1 = .00075$, $g_2 = .00468$, and $\mu_y = .0054$. Fixing $g_2 = .00468$, we then found a values of $\rho = .4993$, and associated values $b_1 = .00231$ and $\mu_y = .00468$ 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 g_2 at the above value, the maximum likelihood estimates are $\rho = .8311$, $b_1 = .00149$, and $\mu_y = .00465$. The data for consumption comes from the St. Louis Federal Reserve data set (FRED). They are taken from their latest vintage (02/26/2010) with the identifiers PCNDGC96_20100223 (real consumption on nondurable goods) and PCESVC96_20100223 (real consumption on services). The population series is from the Bureau of Labor Statistics, Series ID LNU00000000. This is civilian noninstitutional population 16 years and over in thousands. The raw data are monthly. We averaged to compute quarterly series.

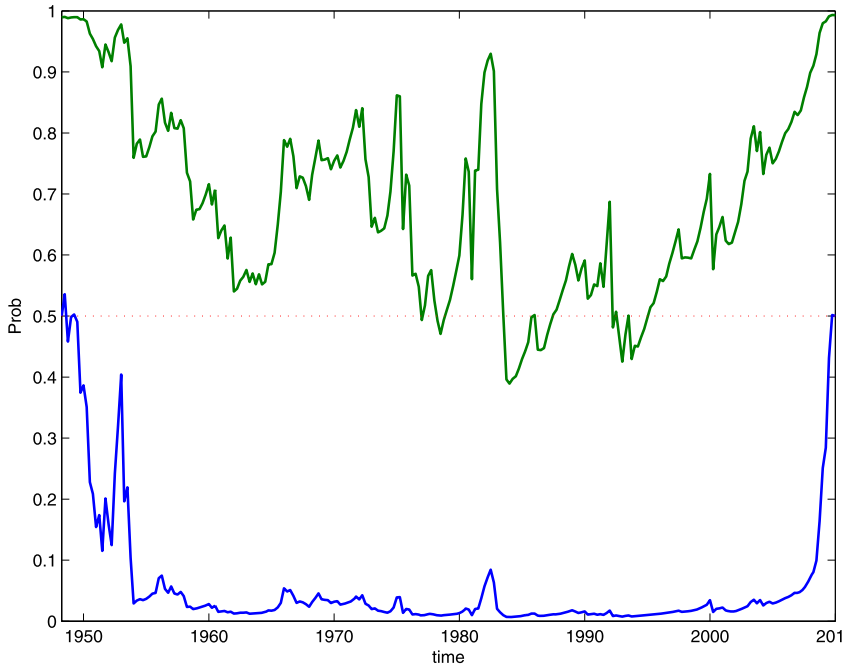

FIGURE 1. Bayesian model probability $\bar{\iota}_t$ (solid line) and worst-case model probability $\tilde{\iota}_t$ (dashed line).

detection error probability as being associated with plausible amounts of model uncertainty.[21] For these values of θ_1 and θ_2,[22] Figure 1 plots values of the Bayesian model mixing probability $\bar{\iota}$ along with the worst-case probability $\tilde{\iota}$. Figure 1 indicates how the worst-case $\tilde{\iota}$ twists toward the long-run risk $\iota = 1$ model. This probability twisting contributes countercyclical movements to the complete set of uncertainty contributions to the shock price (23) that we plot in Figure 2.[23]

Figure 3 decomposes the uncertainty contribution to the shock prices into components that come from the three lines of expression (23), namely, those associated with state dynamics under a known model, unknown states within 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.

[21]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 247 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" values that we took to be the variance of the unconditional stationary distribution of $X_{1t}(\iota)$ and a value for the variance of $X_2(\iota)$ of $.01^2$, which is orders of magnitude larger than the maximum likelihood estimates of μ_y for our entire sample. We set a preliminary state covariance between $X_{1t}(\iota)$ and $X_2(\iota)$ equal to zero. We put these preliminary values into the Kalman filter, ran it for a sample length of 247, and took the terminal covariance matrix as our starting value for the covariance matrix of the hidden state for model ι.

[22]The calibrated values are $\theta_1^{-1} = 7$ and $\theta_2^{-1} = .64$.

[23]The figure plots all components of (23) except the ordinary risk price $\bar{G}H'$.

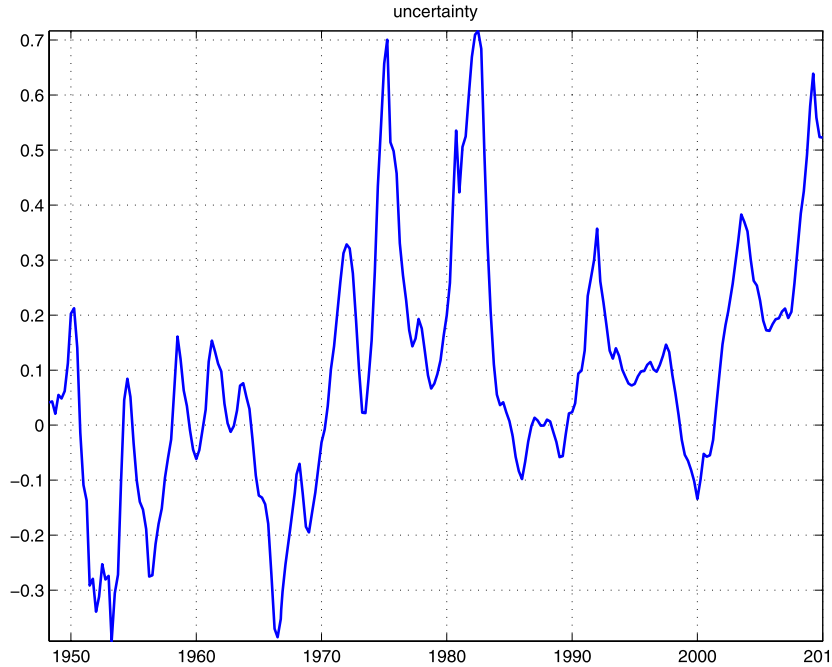

FIGURE 2. Contributions to uncertainty prices from all sources of model uncertainty.

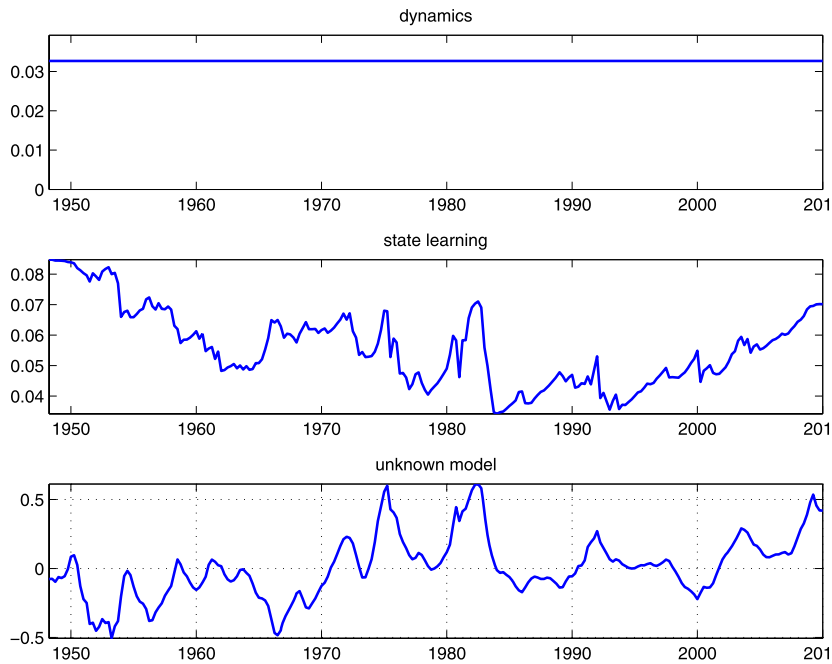

FIGURE 3. Contributions to uncertainty prices coming from separate components on the three lines of (23): from state dynamics (top panel), learning hidden state when the model is known (middle panel), and unknown model (bottom panel).

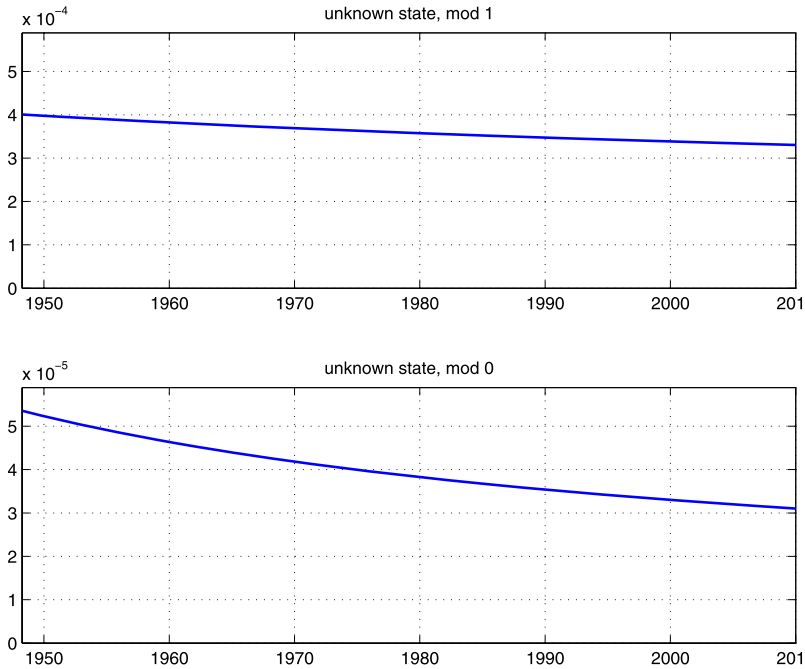

FIGURE 4. $D(\iota)\Sigma(\iota)\lambda(\iota)$ for $\iota = 1$ (top panel) and $\iota = 0$ (bottom panel).

The contribution on the first line of (23) 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.[24] However, a forward-looking component still contributes to the other two components of the uncertainty prices because continuation values influence the worst-case distortions to model probabilities and filtered estimates of the hidden states.

The contribution on the second line of (23) 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 in model $\iota = 1$ than in model $\iota = 0$, because a very persistent hidden state slows convergence of $\Sigma_t(1)$. In particular, the variance of the estimated unconditional mean of consumption growth, $\Sigma_t(\iota)_{22}$, converges more slowly to zero for the long-run risk model $\iota = 1$ than for model $\iota = 0$. The second contribution varies over time through variation in the twisted model probability $\tilde{\iota}$.

[24]In Section 7.4, we consider an example that activates this forward-looking component by specifying that $G(\iota)B(\iota)'$ is not zero.

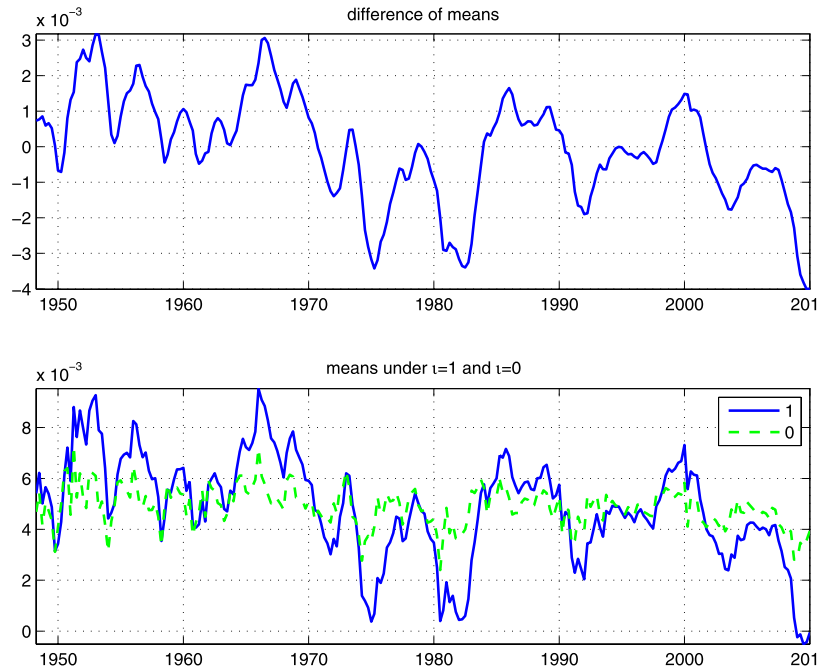

FIGURE 5. Difference in means and means themselves from models $\iota = 1$ and $\iota = 0$.

The contribution on the third line of (23) generally fluctuates 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. While pessimism arising from a concern for robustness necessarily increases the uncertainty prices via the terms on the first two lines of (23), it may either lower or raise it through the term on the third line. The slope of the mean-standard deviation frontier—the maximum Sharpe ratio—is the absolute value of the shock price. Therefore, sizable shock prices of *either* sign imply large maximum Sharpe ratios. Negative shock prices for some signal histories indicate that the representative consumer sometimes fears *positive* consumption innovations because of how they affect probabilities that he attaches to alternative models ι. How concerns about model uncertainty affect uncertainty premia that are embedded in prices of particular risky assets ultimately depends on how their returns are correlated with consumption shocks.

7.2 *Explanation for countercyclical uncertainty prices*

The intertemporal behavior of robustness-induced probability slanting accounts for how learning in the presence of uncertainty about models induces time variation in uncertainty prices. Our representative consumer attaches positive probabilities to a model with statistically subtle high persistence in consumption growth, namely, the $\iota = 1$ long-run risk model, and also to model $\iota = 0$ that asserts much less persistent consumption growth rates. The asymmetrical response of model uncertainty prices to consumption

growth shocks comes from (i) how the representative consumer's concern about misspecification of the probabilities that he attaches to the two models causes him to calculate worst-case probabilities that depend on value functions and (ii) how the value functions for the two models become closer together after positive consumption growth shocks and farther apart after negative shocks. The long-run risk model confronts the consumer with a long-lived shock to consumption growth. That affects the set of possible misspecifications that he worries about and gets reflected in a more negative value of $\kappa(\iota) - (1/2\theta_2)\lambda(\iota)'\Sigma(\iota)\lambda(\iota)$ in formula (20) for the continuation value.[25] The resulting difference in constant terms (terms that depend on calendar time but not on the predicted states) in the value functions for the models with ($\iota = 1$) and without ($\iota = 0$) 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 more persistent consumption growth $\iota = 1$ model relative to the probability that he attaches to the $\iota = 0$ model. Although the long-run risk model has a more negative constant term, when a string of higher than average consumption growths occurs, 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 the exponential twisting formulas (21) and (22), continuation values determine the worst-case probability $\tilde{\iota}$ that the representative consumer attaches to the long-run risk $\iota = 1$ model. Thus our cautious consumer slants probability more toward the 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.

7.3 *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 θ_1 and θ_2, we refocus the representative consumer's concern about misspecification on a different mixture of state 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 θ_2 to set the detection error probability again to .2 (here $\theta_2^{-1} = -1.72$). Notice that now the uncertainty contribution to shock prices remains positive over time. Evidently, in this economy, the representative consumer's fear of *good* consumption news is much less prevalent.

7.4 *State-dependent contributions from unknown dynamics*

The fact that our specification (24) implies that $G(\iota)B(\iota)' = 0$ for $\iota = 0, 1$ disables a potentially interesting component of uncertainty contributions in formula (23). To activate this component, we briefly study a specification in which $G(\iota)B(\iota)' \neq 0$ and in which its

[25] Over our sample, the $[\kappa(1) - (1/2\theta_2)\lambda(1)'\Sigma(1)\lambda(1)] - [\kappa(0) - (1/2\theta_2)\lambda(0)'\Sigma(0)\lambda(0)]$ rises monotonically from -7.46 to -7.25.

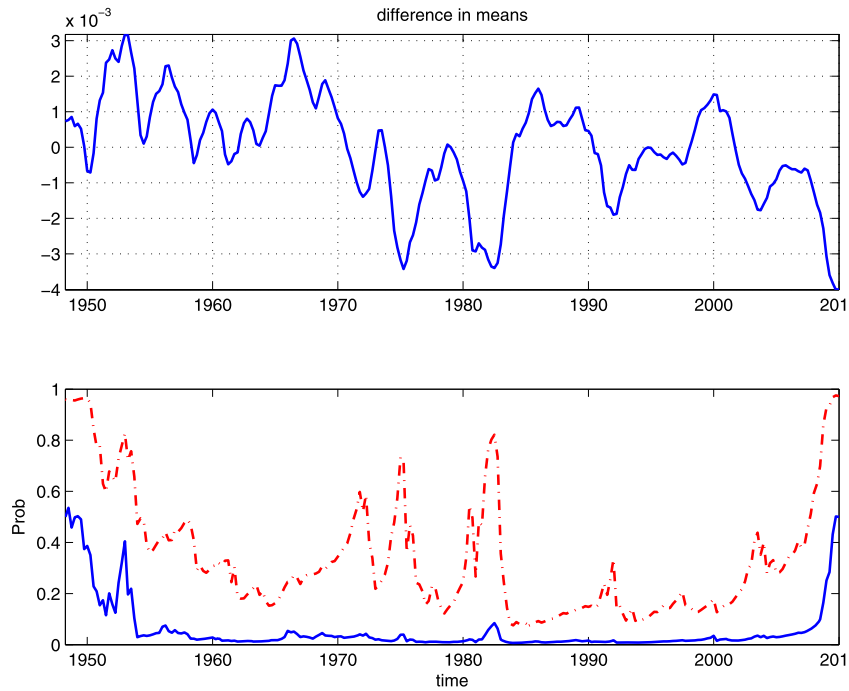

FIGURE 6. Difference in means (top panel) and Bayesian model probability $\bar{\iota}_t$ (solid line) and worst-case model probability $\tilde{\iota}_t$ (dashed line) (bottom panel). Here θ_1 is set to $+\infty$ and θ_2 is set to give a detection error probability of .2.

difference across the two models contributes in interesting ways. In particular, we modify (24) to the single-shock specification

$$\begin{aligned} dX_{1t}(\iota) &= a(\iota)X_{1t}(\iota)\,dt + \hat{b}_1(\iota)\,dW_t(\iota), \\ dX_{2t}(\iota) &= 0, \\ dY_t &= X_{1t}(\iota)\,dt + X_{2t}(\iota)\,dt + \hat{g}_1(\iota)\,dW_t(\iota), \end{aligned} \tag{25}$$

where $X_{1t}(\iota)$ and $X_{2t}(\iota)$ are again scalars, and $W_t(\iota)$ is now a scalar Brownian motion. We construct this one-noise system by simply taking the time-invariant innovations representation for the two-noise, one-signal system (24). We also assume that the representative consumer observes both states for both models $\iota = 0, 1$. Thus, the model is structured so that with ι known, the consumer faces no filtering problem. Therefore, the second source of uncertainty contribution vanishes and (23) simplifies to

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

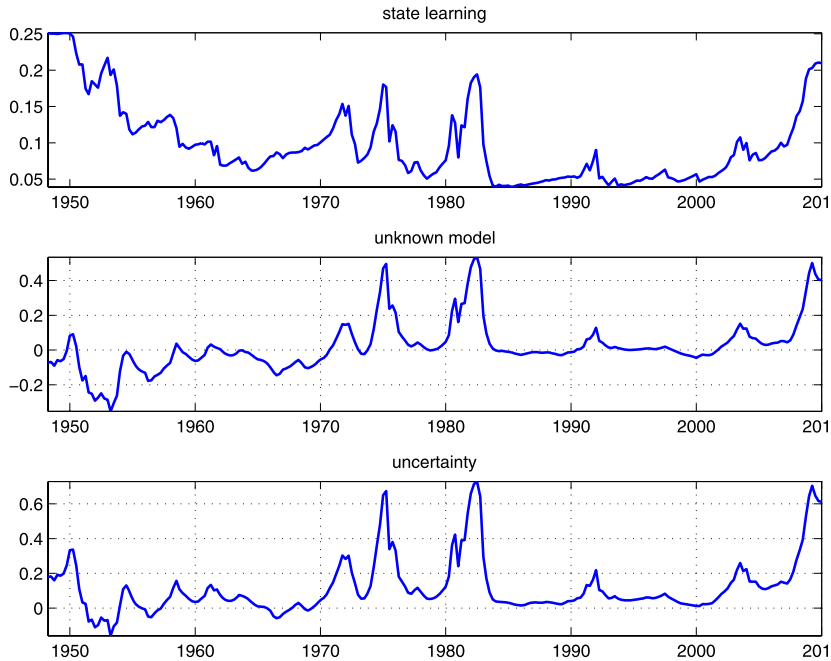

FIGURE 7. Contributions to uncertainty prices from learning hidden state (top panel), models known, unknown model (middle panel), and all sources (bottom panel). Here θ_1 is set to $+\infty$ and θ_2 is set to give a detection error probability of .2. Because $\theta_1 = +\infty$, the contribution from unknown dynamics is identically zero.

Although the representative consumer observes the states, he (or she) does not know which model is correct and constructs the model probability $\tilde{\iota}$ in a robust way.

Figures 8 and 9 illustrate outcomes when we set $\theta_1^{-1} = 1.97$, which we calibrated as described in the Appendix to deliver a detection error probability of .3, and $\theta_2^{-1} = 1.06$, which delivers an overall detection error probability of .2 for our *one-shock* model (25). The term $\tilde{\mu}(\iota) = -\theta_1^{-1}[G(\iota)'H + B(\iota)'\lambda(\iota)]$ is now $-.0460$ for $\iota = 0$ and $-.4454$ for model $\iota = 1$. The contribution of unknown state dynamics reported in the top panel of Figure 9 now varies over time. This variation reflects the difference in $(1/\theta_1)G(\iota)B(\iota)'\lambda(\iota)$ across the two models as well as the fluctuating value of $\tilde{\iota}$. Notice that while the overall uncertainty component to the shock price varies, this variation is much less than in our previous calculations. So while our one-shock model gives rise to time variation in the contribution from a concern about misspecified dynamics, by ignoring robust state estimation, this model excludes some of the interesting variation in the uncertainty exposure prices in our original two-shock model. The prices of exposure to consumption uncertainty are predominately positive, implying that the consumer typically does not fear positive consumption shocks.

7.5 *Reinterpretation of Bansal and Yaron*

If we were to lower θ_2 enough to imply $\tilde{\iota} = 1$, then the representative consumer would ex post act as if he puts probability 1 on the long-run risk model, as assumed by Bansal

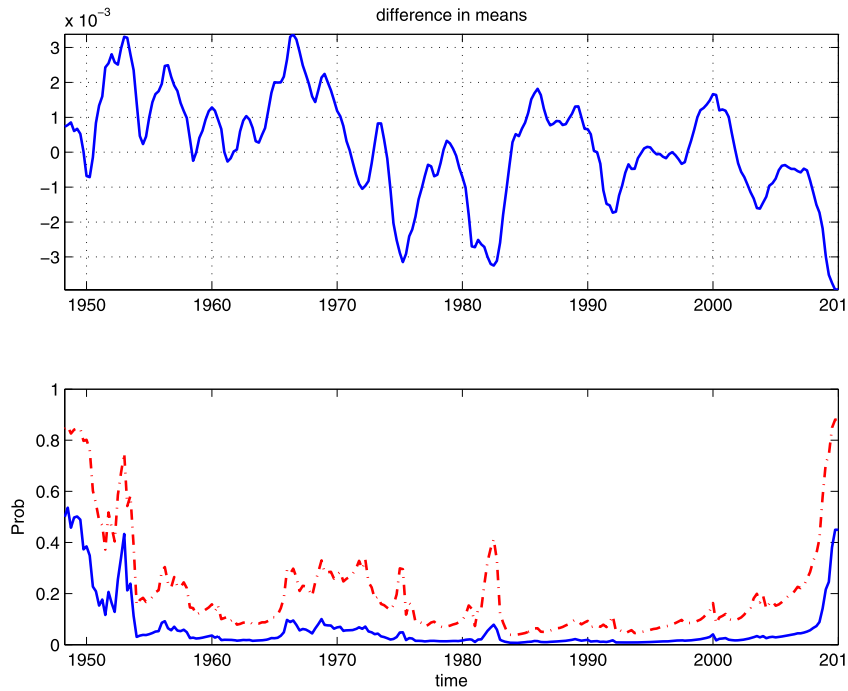

FIGURE 8. The one-noise system. Difference in means (top panel) and Bayesian model probability $\bar{\iota}_t$ (solid line) and worst-case model probability $\tilde{\iota}_t$ (dashed line) (bottom panel).

and Yaron (2004). Then (26) simplifies to

$$\begin{aligned}&\bar{G}^{-1}\left[\frac{1}{\theta_1}G(1)G(1)'H+\frac{1}{\theta_1}G(1)B(1)'\lambda(1)\right]\\&\quad+(\bar{\iota}-1)\bar{G}^{-1}[D(1)\bar{x}(1)-D(0)'\bar{x}(0)].\end{aligned}\tag{27}$$

The first term that captures unknown dynamics becomes constant, while the effects of not knowing the model contribute time variation to the second term. Figure 10 reports the two lines of (27) for the one-noise model calibrated with θ_1 as before and $\tilde{\iota}_t$ set identically to 1 by brute force. The first term of (27) is present in the Bansal and Yaron (2004) approach that has the consumer assign probability 1 without doubt to the long-run risk model, but not the second term accounting for the consumer's doubt about the correct model in our expression (27). So our ex post "as if" interpretation goes only part way toward rationalizing the Bansal and Yaron approach, but it also adds a new ingredient.

8. CONCLUDING REMARKS

The perspective of Bansal and Yaron (2004) is that while (a) there are subtle but recursive-utility-relevant stochastic features of consumption and dividend processes that are difficult to detect from statistical analysis of those series alone, nevertheless (b) data on asset prices together with cross-equation restrictions in the rational expectations style of Hansen and Sargent (1980) substantially tighten parameter estimates of the

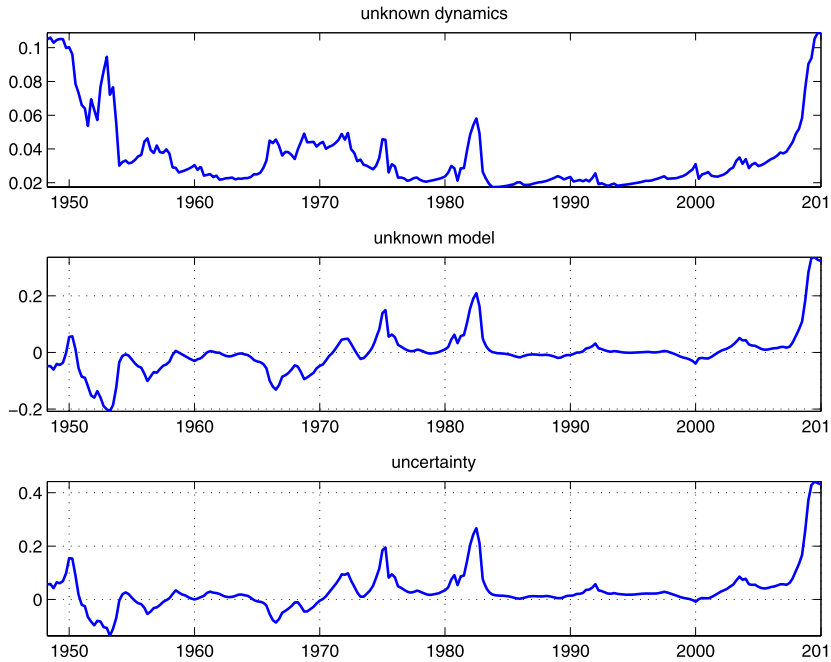

FIGURE 9. The one-noise system. Contributions to uncertainty prices from unknown dynamics (top panel), unknown model (middle panel), and both sources (bottom panel). Because the state is observed, there is no contribution from robust learning about the hidden states $X_{jt}(\iota)$.

joint consumption, dividend processes that agents believe with confidence when they price assets. Thus, although agents' beliefs about the "fundamental" joint consumption, dividend process are difficult to infer from observations on that process *alone*, adding asset prices and the full *confidence* in a stochastic specification that is implicit in the rational expectations hypothesis lets us discover those beliefs.

Our response to point (a) differs from Bansal and Yaron's. Instead of being completely confident in a single stochastic specification, our representative agent is suspicious of that specification and struggles to learn while acknowledging his specification doubts. This leads us to modify Bayes' law in ways that introduce new sources of uncertainty prices. We find contributions of model uncertainty to shock prices that 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, Sargent, and Tallarini (1999) and Anderson, Hansen, and Sargent (2003), (ii) additional components $-\theta_2^{-1}\Sigma(\iota)\lambda(\iota)$ that smoothly decrease in time and that come from learning about parameter values within models, and (iii) the potentially volatile time-varying contribution highlighted in Section 7.2 that reflects the consumer's robust learning about the probability distribution over models.

Our shock prices are counterparts to what are interpreted as risk prices in much of the asset pricing literature, but for us they include both risk and model uncertainty components. Our mechanism for producing time-varying shock prices differs from popular

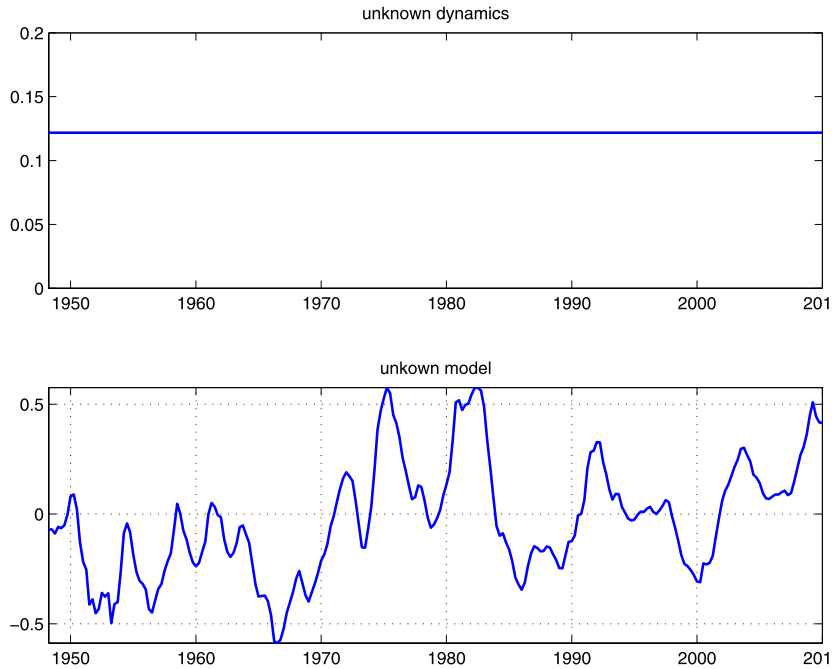

FIGURE 10. The one-noise system with worst-case model probability $\tilde{\iota} \equiv 1$. Unknown dynamics (top panel) and unknown model (bottom panel).

approaches in the existing literature. For instance, Campbell and Cochrane (1999) induced secular movements in risk premia that are backward looking because a social externality depends on current and past average consumption. To generate variations in risk premia, Bansal and Yaron (2004) assumed stochastic volatility in consumption.[26]

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 *multiperiod* uncertainty prices is a natural next step. Multiperiod valuation requires compounding local prices. When the prices are time-varying, this compounding can have nontrivial consequences.

To obtain convenient formulas for prices, we imposed a unitary elasticity of substitution, which implies a constant ratio of consumption to wealth. Measuring the consumption–wealth ratio properly is a difficult task, but we agree that it is probably worthwhile eventually to pay the costs in terms of the computational tractability that would be required to extend our model to allow a variable consumption–wealth ratio.[27]

[26]Our interest in learning and time-series variation in the uncertainty premium differentiates us from Weitzman (2005) and Jobert, Platania, and Rogers (2006), who focused on long-run averages.

[27]We have doubts about the frequently used empirical procedure of using dividend to price ratios to approximate consumption to wealth ratios. 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 can behave very differently from the ratio of wealth to consumption.

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, Heaton, and Li (2008) showed 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.

Appendix. Detection error probabilities

By adapting procedures developed by Hansen, Sargent, and Wang (2002) and Anderson, Hansen, and Sargent (2003) in ways described by Hansen, Mayer, and Sargent (forthcoming), 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 (θ_1, θ_2). Compute the fraction of simulations for which $L_t^w/L_T^a > 1$ 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 $L_T^a/L_T^w > 1$ (call it r_w), where the simulations are generated under the worst-case model. As in Hansen, Sargent, and Wang (2002) and Anderson, Hansen, and Sargent (2003), define the overall detection error probability to be

$$p(\theta_1, \theta_2) = \frac{1}{2}(r_a + r_w). \tag{28}$$

Because in this paper we use what Hansen, Mayer, and Sargent (forthcoming) call Game I, we use the following sequential procedure to calibrate θ_1 first, then θ_2. First, we pretend that $x_t(\iota)$ is observable for $\iota = 0, 1$ and calibrate θ_1 by calculating detection error probabilities for a system with an observed state vector using the approach of Hansen, Sargent, and Wang (2002) and Hansen and Sargent (2008, Chap. 9). Then having pinned down θ_1, we use formula (28) to calibrate θ_2. This procedure takes the point of view that θ_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 θ_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 ι and the model ι state $x_t(\iota)$ being observed, we computed the detection error probability as a function of θ_1 for models $\iota = 0, 1$. (ii) Using a prior probability of $\pi = .5$, we averaged the two curves described in point (i) and plotted the average against θ_1. We calibrated θ_1 to yield an average detection error probability of .4 and used this value of θ_1 in the next step. (iii) With θ_1 locked at the value just set, we then calculated and plotted the detection error for the mixture model against θ_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 θ_2 to fix the overall detection error equal to .2.

References

Ai, H. (2006), "Incomplete information and equity premium in production economy." Unpublished paper, University of Minnesota and Duke University. [133]

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, 68–123. [130, 142, 156, 158]

Bansal, R. and A. Yaron (2004), "Risks for the long run: A potential resolution of asset pricing puzzles." *Journal of Finance*, LIX, 1481–1509. [130, 131, 146, 147, 154, 155, 156, 157]

Barillas, F., L. P. Hansen, and T. J. Sargent (2009), "Doubts or variability?" *Journal of Economic Theory*, 144, 2388–2418. [130, 141]

Bossaerts, P. (2002), *The Paradox of Asset Pricing*. Princeton University Press, Princeton, New Jersey. [134, 137, 138]

Bossaerts, P. (2004), "Filtering returns for unspecified biases in priors when testing asset pricing theory." *Review of Economic Studies*, 71, 63–86. [134, 137, 138]

Breeden, D. T. (1979), "An intertemporal asset pricing model with stochastic consumption and investment opportunities." *Journal of Financial Economics*, 7, 265–296. [137]

Brennan, M. J. and Y. Xia (2001), "Stock price volatility and equity premium." *Journal of Monetary Economics*, 47, 249–283. [133]

Campbell, J. Y. and J. H. Cochrane (1999), "By force of habit: A consumption-based explanation of aggregate stock market behavior." *Journal of Political Economy*, 107, 205–251. [157]

Cerreia, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2008), "Uncertainty averse preferences." Working paper, Collegio Carlo Alberto. [130]

Cogley, T., R. Colacito, L. P. Hansen, and T. J. Sargent (2008), "Robustness and U.S. monetary policy experimentation." *Journal of Money, Credit and Banking*, 40, 1599–1623. [133]

Cogley, T. and T. 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. [134, 137, 138, 140]

Croce, M. M., M. Lettau, and S. C. Ludvigson (2008), "Investor information, long-run risk, and the duration of risky cash flows." Working Paper 12912, NBER. [133]

David, A. (1997), "Fluctuating confidence in stock markets." *Journal of Financial and Quantitative Analysis*, 32, 457–462. [133]

David, A. (2008), "Heterogeneous beliefs, speculation, and the equity premium." *Journal of Finance*, LXIII, 41–83. [133, 134, 137]

David, A. and P. Veronesi (2009), "What ties return volatilities to price valuations and fundamentals?" Chicago Booth Working Paper 10-05, SSRN eLibrary. [133, 139]

Detemple, J. (1986), "Asset pricing in a production economy with incomplete information." *Journal of Finance*, 41, 383–390. [133]

Dothan, M. U. and D. Feldman (1986), "Equilibrium interest rates and multiperiod bonds in a partially observable economy." *The Journal of Finance*, 41, 369–382. [133]

Duffie, D. (2001), *Dynamic Asset Pricing Theory*, third edition. Princeton University Press, Princeton, New Jersey. [132]

Duffie, D. and L. G. Epstein (1992), "Stochastic differential utility." *Econometrica*, 60, 353–394. [142]

Epstein, L. G. and M. Schneider (2008), "Ambiguity, information quality, and asset pricing." *Journal of Finance*, 63, 197–228. [130]

Gilboa, I. and D. Schmeidler (1989), "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics*, 18, 141–153. [129]

Grossman, S. J. (1981), "An introduction to the theory of rational expectations under asymmetric information." *Review of Economic Studies*, 48, 541–559. [139]

Hansen, L. P. (2007), "Beliefs, doubts and learning: Valuing macroeconomic risk." *American Economic Review*, 97, 1–30. [134]

Hansen, L. P. (2008), "Modeling the long run: Valuation in dynamic stochastic economies." Presented as the Fisher–Schultz Lecture to the Econometric Society. [132]

Hansen, L. P., J. Heaton, and N. Li (2008), "Consumption strikes back: Measuring long run risk." *Journal of Political Economy*, 116, 260–302. [146, 158]

Hansen, L. P. and R. Jagannathan (1991), "Implications of security market data for models of dynamic economies." *Journal of Political Economy*, 99, 225–262. [132]

Hansen, L. P., R. Mayer, and T. J. Sargent (forthcoming), "Robust hidden Markov LQG problems." *Journal of Economics Dynamics and Control*. [158]

Hansen, L. P. and S. F. Richard (1987), "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models." *Econometrica*, 55, 587–614. [137, 138]

Hansen, L. P. and T. J. Sargent (1980), "Formulating and estimating dynamic linear rational expectations models." *Journal of Economic Dynamics and Control*, 2, 7–46. [138, 155]

Hansen, L. P. and T. J. Sargent (2001), "Robust control and model uncertainty." *American Economic Review*, 91, 60–66. [129]

Hansen, L. P. and T. J. Sargent (2007), "Recursive robust estimation and control without commitment." *Journal of Economic Theory*, 136, 1–27. [130, 133, 134, 141, 143, 144, 145]

Hansen, L. P. and T. J. Sargent (2008), *Robustness*. Princeton University Press, Princeton, New Jersey. [140, 158]

Hansen, L. P., T. Sargent, and T. Tallarini (1999), "Robust permanent income and pricing." *Review of Economic Studies*, 66, 873–907. [130, 146, 156]

Hansen, L. P., T. J. Sargent, G. A. Turmuhambetova, and N. Williams (2006), "Robust control, min–max expected utility, and model misspecification." *Journal of Economic Theory*, 128, 45–90. [130, 142]

Hansen, L. P., T. J. Sargent, and N. E. Wang (2002), "Robust permanent income and pricing with filtering." *Macroeconomic Dynamics*, 6, 40–84. [158]

Hansen, L. P. and J. Scheinkman (2009), "Long-term risk: An operator approach." *Econometrica*, 77, 177–234. [132]

Harrsion, J. M. and D. M. Kreps (1978), "Speculative investor behavior in a market with heterogeneous expectations." *Quarterly Journal of Economics*, 92, 323–336. [130]

Jobert, A., A. Platania, and L. C. G. Rogers (2006), "A Bayesian solution to the equity premium puzzle." Unpublished paper, Statistical Laboratory, University of Cambridge. [157]

Klibanoff, P., M. Marinacci, and S. Mukerji (2005), "A smooth model of decision making under ambiguity." *Econometrica*, 73, 1849–1892. [130]

Klibanoff, P., M. Marinacci, and S. Mukerji (2009), "Recursive smooth ambiguity preferences." *Journal of Economic Theory*, 144, 930–976. [130]

Lucas, R. E., Jr. (1978), "Asset prices in an exchange economy." *Econometrica*, 46, 1429–1445. [139]

Maccheroni, F., M. Marinacci, and A. Rustichini (2006a), "Ambiguity aversion, robustness, and the variational representation of preferences." *Econometrica*, 74, 1447–1498. [130]

Maccheroni, F., M. Marinacci, and A. Rustichini (2006b), "Dynamic variational preferences." *Journal of Economic Theory*, 128, 4–44. [130]

Scheinkman, J. A. and W. Xiong (2003), "Overconfidence and speculative bubbles." *Journal of Political Economy*, 111, 1183–1220. [130]

Shiller, R. (1972), *Rational Expectations and the Term Structure of Interest Rates*. Ph.D. thesis. Massachusetts Institute of Technology. [138]

Strzalecki, T. (2008), "Axiomatic foundations of multiplier preferences." Unpublished paper, Department of Economics, Northwestern University. [130]

Tallarini, T. D. (2000), "Risk-sensitive real business cycles." *Journal of Monetary Economics*, 45, 507–532. [130, 141]

Veronesi, P. (2000), "How does information quality affect stock market returns." *Journal of Finance*, 55, 807–837. [133]

Weitzman, M. (2005), "A unified theory of equity puzzles." Unpublished paper, Harvard University. [157]

Submitted August, 2009. Final version accepted May, 2010.