The Price of Macroeconomic Uncertainty*

with Tenuous Beliefs

Lars Peter Hansen† Thomas J.Sargent‡

November 27, 2017

Abstract

A dynamic extension of max-min preferences allows a decision maker to consider both a parametric family of what we call *structured* models and *unstructured* alternatives that are statistically close to them. The decision maker suspects that parameter values vary over time in unknown ways that he cannot describe probabilistically. Because he suspects that all of these parametric models are misspecified, he evaluates decisions under alternative probability distributions with much less structure. We characterize equilibrium uncertainty prices by confronting a representative investor with a portfolio choice problem. We offer a quantitative illustration that focuses on the investor's uncertainty about the size and persistence of macroeconomic growth rates. Nonlinearities in marginal valuations induce time variations in market prices of uncertainty. Prices of uncertainty fluctuate because a representative investor especially fears high persistence in bad states and low persistence in good ones.

Keywords— Risk, uncertainty, asset prices, relative entropy, Chernoff entropy, robustness, variational preferences; baseline, structured, and unstructured models

---

*We thank Lloyd Han for his help with estimation, Victor Zhorin for his help calculating Chernoff entropy and Yiran Fan for help in formulating and solving ode's for robust decision problems. We thank Fernando Alvarez, Ben Brooks, Vera Chau, Tim Christensen, Stavros Panageas, Doron Ravid, and Bálint Szőke for critical comments on earlier drafts. We thank an editor and five referees for insightful and helpful comments. Hansen acknowledges support from the Policy Research Program of the MacArthur Foundation under the project "The Price of Policy Uncertainty."

†University of Chicago, E-mail: lhansen@uchicago.edu.

‡New York University, E-mail: thomas.sargent@nyu.edu.

In what circumstances is a minimax solution reasonable? I suggest that it is reasonable if and only if the least favorable initial distribution is reasonable according to your body of beliefs. Good (1952)

1 Introduction

Applied dynamic economic models today typically rely on the rational expectations assumption that agents inside a model and nature share the same probability distribution. This paper instead assumes that agents are uncertain about their model. They may not know values of parameters governing the evolution of pertinent state variables; they may suspect that these parameters vary over time; they may doubt parametric models. We put agents into what they view as a complicated setting in which learning is very difficult and in which valuations and outcomes are sensitive to their subjective beliefs. We draw ideas extensively from literatures on statistical decision theory, robust control theory, and the econometrics of misspecified models to build a tractable model of how decision makers' specification doubts affect equilibrium prices and quantities.

To illustrate our approach, we use a consumption-based asset pricing model as a laboratory for studying how decision makers' specification worries influence "prices of uncertainty." These prices emerge from how decision makers evaluate utility consequences of alternative specifications of state dynamics. We show how these concerns induce variations in asset values and construct a quantitative example that assigns an important role to uncertainty about macroeconomic growth rates. Investors in our model fear growth rate persistence in times of weak growth, but they fear the absence of persistence when macroeconomic growth is high because these have especially adverse consequences for discounted expected utilities.

We propose methods to simplify evaluation and decision making in the face of these specification challenges. We accomplish this by blending ideas from two distinct approaches. We start by assuming that a decision maker considers a parametric family of structured models (with either fixed or time varying parameters) that we represent in terms of a recursive structure suggested by Chen and Epstein (2002) for continuous time models with Brownian motion information flows. Because our decision maker distrusts all of his structured models, he adds unstructured models that reside within a statistical neighborhood of them.[1] Because Chen and Epstein's framework is too confining to include concerns about

[1] By "structured" we don't mean what econometricians in the tradition of either the Cowles commission

such unstructured statistical models, we extend work by Hansen and Sargent (2001) and Hansen et al. (2006) that described a decision maker who expresses distrust of a probability model by surrounding it with an infinite dimensional family of difficult-to-discriminate unstructured models. A Hansen and Sargent decision maker represents alternative models by multiplying baseline probabilities with likelihood ratios whose entropies relative to the baseline model are forced to be small via a penalty parameter. Formally, we accomplish this extension by applying a continuous-time counterpart of the dynamic variational preferences of Maccheroni et al. (2006b). In particular, we generalize what Hansen and Sargent (2001) and Maccheroni et al. (2006a,b) call multiplier preferences.[2]

We illustrate our approach by applying it to an environment that includes macroeconomic growth-rate uncertainty. A representative investor who stands for "the market" has specification doubts. We calculate shadow prices that characterize aspects of model specifications that most concern a representative investor. These shadow prices are also uncertainty prices that clear competitive security markets. Multiplying an endogenously determined vector of worst-case drift distortions by minus one gives a vector of prices that compensate the representative investor for bearing model uncertainty.[3] Time variation in uncertainty prices emerges endogenously since the representative investor's concerns about the persistence of macroeconomic growth rates depend on the state of the macroeconomy.

Viewed as a contribution to the consumption-based asset pricing literature, this paper extends earlier inquiries about whether responses to modest amounts of model ambiguity can substitute for the implausibly large risk aversions that are required to explain observed market prices of risk. Viewed as a contribution to the economic literature on robust control theory and ambiguity, this paper introduces a tractable new way of formulating and quantifying a set of models against which a decision maker seeks robust evaluations and decisions.

Section 2 specifies an investor's baseline probability model and martingale perturbations to it, both cast in continuous time for analytical convenience. Section 3 describes discounted relative entropy, a statistical measure of discrepancy between martingales, and uses it to construct a convex set of probability measures that we impute to our decision maker. This martingale representation proves to be a tractable way for us to formulate robust decision

or rational expectations call "structural" models. We simply mean more or less tightly parameterized statistical models.

[2]Applications of multiplier preferences to macroeconomic policy design and dynamic incentive problems include Karantounias (2013), Bhandari (2014), and Miao and Rivera (2016).

[3]This object also played a central role in the analysis of Hansen and Sargent (2010).

problems in sections 4, 5 and 8.

Section 6 describes and compares two statistical distance measures applicable to a set of martingales, relative entropy and Chernoff entropy. We show how to use these measure 1) in the spirit of Good (1952), *ex post* to assess plausibility of worst-case models, and 2) to calibrate the penalization used to represent variational preferences. By extending estimates from Hansen et al. (2008), section 7 calculates key objects in a quantitative version of a baseline model together with worst-case probabilities associated with a convex set of alternative models that concern both a robust investor and a robust planner. Section 8 constructs a recursive representation of a competitive equilibrium of an economy with a representative investor. Then it links the worst-case model that emerges from a robust planning problem to equilibrium compensations that the representative investor receives in competitive markets. Section 9 tells why it is impossible for our decision maker to learn his way out of the types of model ambiguity with which we present him. It also briefly takes up a dynamic consistency issue present in the problem. Section 10 indicates why a procedure for construcing sets of models recommended by Epstein and Schneider (2003) does not work in our setting. Section 11 offers concluding remarks.

2 Models and perturbations

This section describes nonnegative martingales that alter a baseline probability model. Section 3 then describes how we use a family of parametric alternatives to a baseline model to form a convex set of martingales representing unstructured models that in later sections we use to pose robust decision problems.

2.1 Mathematical framework

For concreteness, we use a specific *baseline* model and in section 3 a corresponding family of parametric alternatives that we call *structured* models. A representative investor cares about a stochastic process $X \doteq \{X_t : t \geqslant 0\}$ that he approximates with a baseline model[4]

$$dX_t = \widehat{\mu}(X_t)dt + \sigma(X_t)dW_t, \tag{1}$$

[4]We let X denote a stochastic process, X_t the process at time t, and x a realized value of the process.

where W is a multivariate Brownian motion.[5]

A decision maker cares about plans. A *plan* is a $\{C_t : t \geqslant 0\}$ that is a progressively measurable process with respect to the filtration $\mathcal{F} = \{\mathcal{F}_t : t \geqslant 0\}$ associated with the Brownian motion W augmented by information available at date zero. The date t component C_t is measurable with respect to $\mathcal{F}_t$.

Because he does not fully trust baseline model (1), the decision maker explores utility consequences of other probability models that he obtains by multiplying probabilities associated with (1) by likelihood ratios. Following Hansen et al. (2006), we represent a likelihood ratio by a positive martingale M^U with respect to the baseline model (1) that satisfies[6]

$$dM_t^U = M_t^U U_t \cdot dW_t \tag{2}$$

or

$$d \log M_t^U = U_t \cdot dW_t - \frac{1}{2}|U_t|^2 dt, \tag{3}$$

where U is progressively measurable with respect to the filtration $\mathcal{F}$. We adopt the convention that M_t^U is zero when $\int_0^t |U_\tau|^2 d\tau$ is infinite, which happens with positive probability. In the event that

$$\int_0^t |U_\tau|^2 d\tau < \infty \tag{4}$$

with probability one, the stochastic integral $\int_0^t U_\tau \cdot dW_\tau$ is an appropriate probability limit. Imposing the initial condition $M_0^U = 1$, we express the solution of stochastic differential equation (2) as the stochastic exponential

$$M_t^U = \exp\left(\int_0^t U_\tau \cdot dW_\tau - \frac{1}{2}\int_0^t |U_\tau|^2 d\tau\right). \tag{5}$$

As specified so far, M_t^U is a local martingale, but not necessarily a martingale.[7]

Definition 2.1. $\mathcal{M}$ *denotes the set of all martingales* M^U *constructed as stochastic exponentials via representation* (5) *with a* U *that satisfies* (4) *and is progressively measurable with respect to* $\mathcal{F} = \{\mathcal{F}_t : t \geqslant 0\}$.

[5] Applications typically use Markov specifications, but a Markov formulation is not essential. It could be generalized to allow other stochastic processes that can be constructed as functions of a Brownian motion information structure.

[6] James (1992), Chen and Epstein (2002), and Hansen et al. (2006) used this representation.

[7] It is not convenient here to impose sufficient conditions for the stochastic exponential to be a martingale like Kazamaki's or Novikov's. Instead we will verify that an extremum of a pertinent optimization problem does indeed result in a martingale.

Associated with U are probabilities defined by

$$E^U\left[B_t|\mathcal{F}_0\right] = E\left[M_t^U B_t|\mathcal{F}_0\right]$$

for any $t \geqslant 0$ and any bounded $\mathcal{F}_t$-measurable random variable B_t, so the positive random variable M_t^U acts as a Radon-Nikodym derivative for the date t conditional expectation operator $E^U\left[\,\cdot\,|X_0\right]$. The martingale property of the process M^U ensures that conditional expectations operators satisfy a Law of Iterated Expectations.

Under baseline model (1), W is a standard Brownian motion, but under the alternative U model, it has increments

$$dW_t = U_t dt + dW_t^U, \tag{6}$$

where W^U is now a standard Brownian motion. Furthermore, under the M^U probability measure, $\int_0^t |U_\tau|^2 d\tau$ is finite with probability one for each t. While (3) expresses the evolution of $\log M^U$ in terms of increment dW, the evolution in terms of dW^U is:

$$d\log M_t^U = U_t \cdot dW_t^U - \frac{1}{2}|U_t|^2 dt. \tag{7}$$

In light of (7), we can write model (1) as:

$$dX_t = \widehat{\mu}(X_t)dt + \sigma(X_t)\cdot U_t dt + \sigma(X_t)dW_t^U.$$

3 Measuring statistical discrepancies

We use entropy relative to the baseline probability to restrict martingales that represent alternative probabilities. The process $M^U \log M^U$ evolves as an Ito process with date t drift (also called a local mean) $\frac{1}{2}M_t^U|U_t|^2$. Write the conditional mean of $M^U \log M^U$ in terms of a history of local means[8]

$$E\left[M_t^U \log M_t^U|\mathcal{F}_0\right] = \frac{1}{2}E\left(\int_0^t M_\tau^U |U_\tau|^2 d\tau|\mathcal{F}_0\right).$$

To formulate a decision problem that chooses probabilities to minimize expected utility, we will use this representation without imposing that M^U is a martingale and then verify that

[8]There exists a variety of sufficient conditions that justify this equality.

the solution is indeed a martingale. Hansen et al. (2006) justify this approach.[9]

To construct entropy relative to a probability specification affiliated with a martingale M^S defined by a drift distortion process S that is measurable with respect to $\mathcal{F}$, rather than a log likelihood ratio $\log M_t^U$ with respect to the baseline model, we use a log likelihood ratio $\log M_t^U - \log M_t^S$ with respect to the M_t^S model to arrive at:

$$E\left[M_t^U\left(\log M_t^U - \log M_t^S\right)|\mathcal{F}_0\right] = \frac{1}{2}E\left(\int_0^t M_\tau^U |U_\tau - S_\tau|^2 d\tau \middle| \mathcal{F}_0\right).$$

When the following limits exist, a notion of relative entropy appropriate for stochastic processes is:

$$\begin{aligned}\lim_{t\to\infty}\frac{1}{t}E\left[M_t^U\left(\log M_t^U - \log M_t^S\right)\middle|\mathcal{F}_0\right] &= \lim_{t\to\infty}\frac{1}{2t}E\left(\int_0^t M_\tau^U|U_\tau - S_\tau|^2 d\tau\middle|\mathcal{F}_0\right)\\ &= \lim_{\delta\downarrow 0}\frac{\delta}{2}E\left(\int_0^\infty \exp(-\delta\tau)M_\tau^U|U_\tau - S_\tau|^2 d\tau\middle|\mathcal{F}_0\right).\end{aligned}$$

The second line is the limit of Abel integral averages, where scaling by δ makes the weights $\delta\exp(-\delta\tau)$ integrate to one. Rather than using undiscounted relative entropy, we shall use Abel averages with a discount rate equaling the subjective rate that discounts expected utility flows. With that in mind, we define a discrepancy between two martingales M^U and M^S as:

$$\Delta\left(M^U; M^S|\mathcal{F}_0\right) = \frac{\delta}{2}\int_0^\infty \exp(-\delta t)E\left(M_t^U \mid U_t - S_t\mid^2 \middle|\mathcal{F}_0\right)dt.$$

Hansen and Sargent (2001) and Hansen et al. (2006) set $S_t \equiv 0$ to construct discounted relative entropy neighborhoods of a baseline model:

$$\Delta(M^U; 1|\mathcal{F}_0) = \frac{\delta}{2}\int_0^\infty \exp(-\delta t)E\left(M_t^U|U_t|^2\middle|\mathcal{F}_0\right)dt \geqslant 0, \tag{8}$$

where baseline probabilities are represented here by a degenerate $S_t \equiv 0$ drift distortion affiliated with a martingale that is identically one. Formula (8) quantifies how a martingale M^U distorts baseline model probabilities.

Hansen and Sargent (2001) and Hansen et al. (2006) start from a unique baseline model. Instead, we start from a convex set $M^S \in \mathcal{M}^o$ of *structured* models that we represent by

[9]See their Claims 6.1 and 6.2.

martingales with respect to one such baseline model. We describe how to form $\mathcal{M}^o$ in subsection 3.1. Structured models $\mathcal{M}^o$ are parametric alternatives to the baseline model that particularly concern the decision maker. For scalar $\theta > 0$, define a scaled discrepancy of martingale M^U from a set of martingales $\mathcal{M}^o$ as

$$\Theta(M^U|\mathcal{F}_0) = \theta \inf_{M^S \in \mathcal{M}^o} \Delta\left(M^U; M^S|\mathcal{F}_0\right). \tag{9}$$

Scaled discrepancy $\Theta(M^U|\mathcal{F}_0)$ equals zero for M^U in $\mathcal{M}^o$ and is positive for M^U not in $\mathcal{M}^o$. We use discrepancy $\Theta(M^U|\mathcal{F}_0)$ to define a set of unstructured models near $\mathcal{M}^o$ for which a decision maker wants to investigate utility consequences. The scaling parameter θ measures how heavily an expected utility maxmizing decision maker penalizes an expected utility minimizing agent for distorting probabilities relative to models in $\mathcal{M}^o$.

3.1 A family $\mathcal{M}^o$ of structured models

We construct a family of structured probabilities by forming a set of martingales M^S with respect to a baseline probability associated with model (1). Formally,

$$\mathcal{M}^o = \left\{M^S \in \mathcal{M} \text{ such that } S_t \in \Xi_t \text{ for all } t \geqslant 0\right\} \tag{10}$$

where Ξ is a process of convex sets adapted to the filtration $\mathcal{F}$. Chen and Epstein (2002) also used an instant-by-instant constraint like (10) to construct a set of probability models.[10]

When the set $\mathcal{M}^o$ of probabilities comprises part of the preferences of a max-min decision maker, restriction (10) imposes a recursive structure on those preferences that justifies using dynamic programming. That is a consequence of the fact that these preferences satisfy a dynamic consistency property axiomatized by Epstein and Schneider (2003). Example 3.1 provides a specifications of Ξ in (10) that encompasses the application with uncertainty about macroeconomic growth rates to be featured in section 5. In section 9, we revisit restriction (10) and discuss its implications for applications not explored in this paper. It is important to note here that in contrast to Chen and Epstein (2002), we use constraint (10) only as an intermediate step in constructing a larger set of statistically similar unstructured models whose utility consequences the decision maker wants to know.

[10] Anderson et al. (1998) also explored consequences of a constraint like (10) but without the state dependence in Ξ. Allowing for state dependence is important in the applications featured in this paper.

Example 3.1. *Suppose that S_t^j is a time invariant function of the Markov state X_t for each $j = 1, \ldots, n$. Linear combinations of S_t^j's generate the following set of time-invariant parameter models:*

$$\left\{ M^S \in \mathcal{M} : S_t = \sum_{j=1}^{n} \pi^j S_t^j, \pi \in \Pi \textit{ for all } t \geqslant 0 \right\}. \tag{11}$$

The unknown parameter vector is $\pi = \begin{bmatrix} \pi^1 & \pi^2 & \ldots & \pi^n \end{bmatrix}' \in \Pi$, a closed convex subset of $\mathbb{R}^n$. We can include time-varying parameter models in $\mathcal{M}^o$ by changing (11) *to:*

$$\left\{ M^S \in \mathcal{M} : S_t = \sum_{j=1}^{n} \widetilde{\pi}_t^j S_t^j, \widetilde{\pi}_t \in \Pi \textit{ for all } t \geqslant 0 \right\}, \tag{12}$$

where the time-varying parameter vector $\widetilde{\pi}_t = \begin{bmatrix} \widetilde{\pi}_t^1 & \widetilde{\pi}_t^2 & \ldots & \widetilde{\pi}_t^n \end{bmatrix}'$ has realizations confined to Π, the same convex subset of $\mathbb{R}^n$ that appears in (11)*. The decision maker has an incentive to compute the mathematical expectation of $\widetilde{\pi}_t$ conditional on date t information, which we denote π_t. Since realizations of $\widetilde{\pi}_t$ are restricted to be in Π, conditional expectations π_t of $\tilde{\pi}_t$ also belong to Π, so what now plays the role of Ξ in* (10) *becomes*

$$\Xi_t = \left\{ S_t = \sum_{j=1}^{n} \pi_t^j S_t^j, \;\; \pi_t \in \Pi, \;\; \pi_t \textit{ is } \mathcal{F}_t \textit{ measurable} \right\}. \tag{13}$$

As the quantitative example in section 7 demonstrates, even when structured models are linear in a Markov state, max-min expressions of ambiguity aversion can deliver worst-case models with nonlinearities. An *ex post* assessment of empirical plausibility of the type envisioned by Good (1952) would ask whether such nonlinearities are plausible.

In section 4.2.2 we describe another construction of Ξ_t that is motivated in part by using relative entropy to restrict alternative models that concern the decision maker. In our application, we use the section 4.2.2 way of constructing Ξ to guide how we choose the set Π of potential parameter values.

3.1.1 Comparison to earlier formulations

Especially in applications to asset pricing, it is useful to compare the role of the baseline model here with its role in the robust decision model formulated by Hansen and Sargent (2001) and Hansen et al. (2006). For example, in Barillas et al. (2009), who build on the two papers just mentioned, the baseline model is both (a) a unique structured model, the distrust of which motivates a decision maker to compute worst-case models to guide evaluations and decisions; and (b) the probability model under which we as outside analysts quantitatively evaluate prices of uncertainty. While the baseline model continues to play role (b) in this paper, it now plays a substantially modified version of role (a). The baseline model is now just one of a set of structured models that the decision maker entertains, albeit a pivotal one that anchors specifications of the remaining members of the set. The decision maker distrusts all models in the set of structured models constructed as we have described.

3.2 Misspecification of structured models

Unlike a decision maker of Epstein and Schneider (2003), our decision maker wants to evaluate the utility consequences not just of the structured models in $\mathcal{M}^o$ but also of unstructured models that statistically are difficult to distinguish from them. For that purpose, he employs the scaled statistical discrepancy measure $\Theta(M^U|\mathcal{F}_0)$ defined in (9).[11] The decision maker uses the scaling parameter $\theta < \infty$ and the relative entropy that it implies to calibrate a set of nearby unstructured models. We pose a minimization problem in which θ serves as a penalty parameter that precludes exploring utility consequences of unstructured probabilities that statistically deviate too much from the structured models. This minimization problem induces a preference ordering within a broader class of dynamic variational preferences that Maccheroni et al. (2006b) showed are dynamically consistent.

To understand how our formulation relates to dynamic variational preferences, notice how structured models represented in terms of their drift distortion processes S_t appear separately on the right side of the statistical discrepancy measure

$$\Delta\left(M^U; M^S|\mathcal{F}_0\right) = \frac{\delta}{2}\int_0^\infty \exp(-\delta t) E\left(M_t^U \mid U_t - S_t \mid^2 \Big| \mathcal{F}_0\right) dt.$$

Specification (9) leads to a conditional discrepancy

$$\xi_t(U_t) = \inf_{S_t \in \Xi_t} |U_t - S_t|^2$$

[11]Watson and Holmes (2016) and Hansen and Marinacci (2016) discuss misspecification challenges confronted by statisticians and economists.

and an associated scaled integrated discounted discrepancy

$$\Theta\left(M^U|\mathcal{F}_0\right) = \frac{\theta\delta}{2}\int_0^\infty \exp(-\delta t) E\left[M_t^U \xi_t(U_t)\middle|\mathcal{F}_0\right] dt. \tag{14}$$

Our decision maker cares also about the utility consequences of statistically close unstructured models that he describes in terms of the discrepancy measure $\Theta\left(M^U|\mathcal{F}_0\right)$. For any hypothetical state- and date-contingent plan – a consumption plan in the example of section 5 – our decision maker follows Hansen and Sargent (2001) by minimizing a discounted expected utility function plus a θ-scaled relative entropy penalty $\Theta(M^U|\mathcal{F}_0)$ over the set of models.

4 Recursive Representation of Preferences

A decision maker uses scalar continuation value stochastic processes to rank alternative consumption plans. Date t continuation values tell a decision maker's date t ranking. Continuation value processes have a recursive structure that makes preferences be dynamically consistent. For Markovian consumption processes, a Hamilton-Jacobi-Bellman (HJB) equation describes the evolution of continuation values.

4.1 Continuation values

For a consumption plan $\{C_t\}$, the continuation value process $\{V_t\}_{t=0}^\infty$ is defined by

$$V_t = \min_{\{U_\tau : t \leqslant \tau < \infty\}} E\left(\int_0^\infty \exp(-\delta\tau)\left(\frac{M_{t+\tau}^U}{M_t^U}\right)\left[\psi(C_{t+\tau}) + \left(\frac{\theta\delta}{2}\right)\xi_{t+\tau}(U_{t+\tau})\right] d\tau \mid \mathcal{F}_t\right) \tag{15}$$

where ψ is an instantaneous utility function. Equation (15) builds in a recursive structure that can be expressed as

$$\begin{aligned} V_t = \min_{\{U_\tau : t \leqslant \tau < t+\epsilon\}} \Bigg\{ & E\left[\int_0^\epsilon \exp(-\delta\tau)\left(\frac{M_{t+\tau}^U}{M_t^U}\right)\left[\psi(C_{t+\tau}) + \left(\frac{\theta\delta}{2}\right)\xi_{t+\tau}(U_{t+\tau})\right] d\tau \mid \mathcal{F}_t\right] \\ & + \exp(-\delta\epsilon) E\left[\left(\frac{M_{t+\epsilon}^U}{M_t^U}\right) V_{t+\epsilon} \mid \mathcal{F}_t\right]\Bigg\} \end{aligned} \tag{16}$$

for $\epsilon > 0$.

Heuristically, we can "differentiate" the right-hand side of (16) with respect to ϵ to obtain an instantaneous counterpart to a Bellman equation. Viewing the continuation value process $\{V_t\}$ as an Ito process, write:

$$dV_t = \nu_t dt + \varsigma_t \cdot dW_t. \tag{17}$$

A local counterpart to (16) is:

$$0 = \min_{U_t} \left[\psi(C_t) - \frac{\theta\delta}{2}\xi_t(U_t) - \delta V_t + U_t \cdot \varsigma_t + \nu_t \right], \tag{18}$$

where U_t is restricted to be $\mathcal{F}_t$ measurable. The term $U_t \cdot \varsigma_t$ comes from an Ito adjustment to the local covariance between $\frac{dM_t^U}{M_t^U}$ and dV_t. Alternatively, it is an adjustment to the drift ν_t of dV_t that is induced by using martingale M^U to change the probability measure. Preferences defined in this way are a continuous-time counterpart to the dynamic variational preferences of Maccheroni et al. (2006b). Their recursive structure demonstrates that these preferences are dynamically consistent.[12]

4.2 Markovian Consumption Processes

By ranking consumption processes with continuation value processes satisfying (18), a decision maker evaluates utility consequences of unstructured models that our relative entropy measure asserts are difficult to distinguish from members of the set of structured models $\mathcal{M}^o$ that also concern him. We now illustrate this by deliberately considering a setup that starts with a Markovian consumption process and eventually adds non-Markovian processes.

To construct a set of models, the decision maker:

1) Begins with a Markovian baseline model.

2) Creates a set $\mathcal{M}^o$ of *structured* models by naming a sequence of closed convex sets $\{\Xi_t\}$ and associated drift distortion processes $\{S_t\}$ that satisfy structured model constraint (10).

[12]The term $\frac{\theta\delta}{2}\xi_t(U_t)$ in our analysis is γ_t in Maccheroni et al. (2006b) and our equation (18) is a continuous time counterpart to equation (12) in their paper. In Hansen and Sargent (2001) and Hansen et al. (2006), their $\gamma_t = \frac{\theta}{2}|U_t|^2$ where Maccheroni et al.'s θ is a scaled version of ours. This construction contrasts with how equation (17) of Maccheroni et al. (2006b) describes Hansen and Sargent and Hansen et al.'s "multiplier preferences". We regard the disparity as a minor blemish in Maccheroni et al. (2006b). It is pertinent to point this out here only because the analysis in this paper generalizes our earlier work.

3) Augments $\mathcal{M}^o$ with additional *unstructured* models that violate (10) but are statistically close to models that do satisfy it according to discrepancy measure (9).

For step 1, we use the diffusion (1) as a Markovian baseline model. Step 3 includes statistically similar models that are not Markovian. We will describe two approaches for step 2.

4.2.1 Revisiting example 3.1

We begin with Markov alternatives to (1) of the form

$$dX_t = \mu^j(X_t) + \sigma(X_t)dW_t^{S^j}, \quad j = 1, \ldots, n$$

where W^{S^j} is a Brownian motion and (6) continues to describe the relationship between processes W and W^{S^j}. The vectors of drifts μ^j differ from $\widehat{\mu}$ in baseline model (1), but the volatility vector σ is common to all models. These structured models have drift distortions that are time-invariant functions of the Markov state, namely, linear combinations of $S_t^j = \eta^j(X_t)$, where

$$\eta^j(x) = \sigma(x)^{-1}\left[\mu^j(x) - \widehat{\mu}(x)\right].$$

As in example 3.1, we want to add structured models of the form (10) with Ξ_t satisfying (13) to an initial baseline model, so we represent an initial set of time invariant parameter models in terms of

$$s(x) = \sum_{j=1}^{n} \pi^j \eta^j(x), \;\; \pi \in \Pi, \tag{19}$$

where Π is a convex set of possible parameter values. We allow parameters and conditional expectations of them to vary over time. Our decision maker considers mixtures in which

$$\Pi = \left\{ \pi : \pi^j \geqslant 0, \quad \sum_{j=1}^{n} \pi^j = 1 \right\}$$

represents alternative posterior probabilities that at a given date can be assigned to parameter configurations present within the set of structured models.

Where C_t is a time invariant function of the Markov state, we depict preferences with an instantaneous objective function $\delta\phi(x)$ and a subjective discount rate δ where we write $\psi(C_t) = \delta\phi(X_t)$. We seek a continuation value process $\{V_t\}$ whose time t component can be represented as

$$V_t = \delta\Phi(X_t).$$

We exploit the Markovian assumption to represent continuation values in terms of a function Φ that solves a functional equation. Specifically, the local mean ν and volatility $|\varsigma|$ that govern the evolution of the continuation value process in equation (17) are described by

$$\begin{aligned}\nu_t &= \widehat{\mu}(X_t) \cdot \frac{\partial \Phi}{\partial x}(X_t) + \frac{1}{2}\text{trace}\left[\sigma(X_t)'\frac{\partial^2 \Phi}{\partial x \partial x'}(X_t)\sigma(X_t)\right] \\ \varsigma_t &= \sigma(X_t)'\left[\frac{\partial \Phi}{\partial x}(X_t)\right].\end{aligned}$$

Substituting these into equation (18) gives the HJB equation

$$\begin{aligned}0 = \min_{u,s} & -\delta\Phi(x) + \phi(x) + \widehat{\mu}(x) \cdot \frac{\partial \Phi}{\partial x}(x) + [\sigma(x)u] \cdot \frac{\partial \Phi}{\partial x}(x) \\ & + \frac{1}{2}\text{trace}\left[\sigma(x)'\frac{\partial^2 \Phi}{\partial x \partial x'}(x)\sigma(x)\right] + \frac{\theta}{2}|u-s|^2 \end{aligned} \tag{20}$$

where minimization over u, s is subject to (19).[13] Here s represents structured models in $\mathcal{M}^o$ and u represents unstructured models that are statistically similar to models in $\mathcal{M}^o$.

The problem on the right side of HJB equation (20) can be simplified by first minimizing with respect to u given s, or equivalently, by minimizing with respect to $u-s$ given s. First-order conditions for this simpler problem lead to

$$u - s = -\frac{1}{\theta}\sigma(x)'\frac{\partial \Phi}{\partial x}(x). \tag{21}$$

Substituting from (21) into HJB equation (20) gives the reduced HJB equation:

Problem 4.1.

$$\begin{aligned}0 = \min_{s} & -\delta\Phi(x) + \phi(x) + \widehat{\mu}(x) \cdot \frac{\partial \Phi}{\partial x}(x) + [\sigma(x)s] \cdot \frac{\partial \Phi}{\partial x}(x) \\ & + \frac{1}{2}\text{trace}\left[\sigma(x)'\frac{\partial^2 \Phi}{\partial x \partial x'}(x)\sigma(x)\right] - \frac{1}{2\theta}\left[\frac{\partial \Phi}{\partial x}(x)\right]'\sigma(x)\sigma(x)'\left[\frac{\partial \Phi}{\partial x}(x)\right] \end{aligned} \tag{22}$$

where minimization is subject to (19). *Given the minimizing* $s^*(x)$*, we can recover the minimizing* u *from* $u^*(x) = s^*(x) - \frac{1}{\theta}\sigma(x)'\frac{\partial \Phi^*}{\partial x}(x)$*, where* Φ^* *solves HJB equation* (22).

[13] We have divided by δ for notational convenience.

Minimizers of the right side of (20) are s^* and u^*. The minimizing s is a structured state-dependent drift distortion taking the form $s^*(x) = \sum_{j=1}^n \pi^{j*}(x)\eta^j(x)$. The minimizing u is a worst-case drift distortion $u^*(x)$ relative to $s^*(x)$ that adjusts for the decision maker's suspicion that the data are generated by a model not in $\mathcal{M}^o$.

The solution of the HJB equation in problem 4.1 should in general be interpreted as a viscosity solution that satisfies appropriate boundary conditions as well as conditions that justify a verification theorem. In an example in section 7, the first derivative of the value function has a kink, but the value function is still a viscosity solution.

More generally, the decision maker or, as in our example, a fictitious planner, could face a resource allocation problem that involves accumulating physical capital and other factors of production that make consumption endogenous. A counterpart to problem 4.1 in such settings would be a two-player, zero-sum stochastic differential game of a type studied by Fleming and Souganidis (1989).

4.2.2 Restricting relative entropy

In this subsection, we describe how, instead of forming a set of structured model according to equation (19), we form a set indirectly by restricting relative entropies. We will use this approach in our quantitative application. For special cases that include our application in section 7, the two ways of forming a set of structured models coincide, but in other applications they would not.[14]

Section 3 defined relative entropy for a stochastic process M^S to be

$$\varepsilon(M^S) = \lim_{t\to\infty} \frac{1}{2t}\int_0^t E\left(M_\tau^S |S_\tau|^2 \Big| \mathcal{F}_0\right) d\tau.$$

Notice that ε is the limit as $t \to +\infty$ of a process of mathematical expectations of time series averages

$$\frac{1}{2t}\int_0^t |S_\tau|^2$$

under the probability measure implied by M^S. Suppose that M^S is defined by the drift distortion $S_t = \eta(X_t)$, where X is an ergodic Markov process with transition probabilities that converge to a unique well-defined stationary distribution Q under the M^S probability.

[14] See appendix A.

In this case, we can use Q to evaluate relative entropy by computing:

$$\frac{1}{2}\int |\eta|^2 dQ.$$

In continuous time, we can represent the instantaneous counterpart to the one-period transition distribution for a Markov process in terms of an infinitesimal generator. A generator tells how conditional expectations of the Markov state evolve locally. It can be derived heuristically by differentiating the family of conditional expectation operators with respect to the gap of elapsed time. For a diffusion, the infinitesimal generator $\mathcal{A}$ of transitions under the M^S probability is the second-order differential operator:

$$\mathcal{A}\rho = \frac{\partial \rho}{\partial x} \cdot (\widehat{\mu} + \sigma\eta) + \frac{1}{2}\text{trace}\left(\sigma' \frac{\partial^2 \rho}{\partial x \partial x'}\sigma\right),$$

where ρ resides in an appropriately defined domain of the generator $\mathcal{A}$. A stationary distribution Q for a continuous-time Markov process with generator $\mathcal{A}$ satisfies:

$$\int \mathcal{A}\rho dQ = 0. \tag{23}$$

This equation can be derived heuristically by applying the Law of Iterated expectations.

We apply characterization (23) of a stationary distribution to obtain the following second-order differential equation whose solution gives the implied relative entropy:

$$\mathcal{A}\rho = \frac{\mathsf{q}^2}{2} - \frac{|\eta|^2}{2}, \tag{24}$$

where $\varepsilon(M^S) = \frac{\mathsf{q}^2}{2}$ is the relative entropy and q measures the magnitude of the corresponding drift distortion. To compute relative entropy associated with a process defined by generator $\mathcal{A}$, we solve equation (24) simultaneously for q and the function ρ. Since Q is a stationary distribution, the right side of (24) must have mean zero under the stationary distribution Q, which justifies our interpretation of $\frac{\mathsf{q}^2}{2}$. The function ρ is well defined only up to translation by a constant. This approach for computing relative entropy has direct extensions for Markov jump processes and mixed jump diffusion processes. For diffusion processes, equation (24) is a special case of a Feynman-Kac equation.[15]

[15] Had our interest been to compute discounted relative entropy, equation (24) would include a term $-\delta\rho$ on the left-side and the term $\frac{\mathsf{q}}{2}$ would be omitted. Discounted relative entropy is state dependent and given by $\delta\rho(x)$ with ρ satisfying a different Feynman-Kac equation.

We have just illustrated how to compute relative entropy for a Markov process. We want to use similar calculations to restrict the family of potential structured models in terms of their relative entropies $\varepsilon(M^S)$. Instead of specifying η, we use relative entropy as a constraint on η, , for instance by restricting η. One approach is to specify only q and then to find all η that satisfy (26), or more generally all S that satisfy $\varepsilon(M^S) \leqslant \frac{\mathsf{q}^2}{2}$. However, imposing a constraint on relative entropy by pre-specifying only q produces a family of probabilities that fails to satisfy an instant-by-instant constraint $S_t \in \Xi_t$ for all $t \geqslant 0$ in (10) for some collection $\{\Xi_t : t \geqslant 0\}$.[16] Therefore, we use the following approach instead. For ρ and q specified *a priori*, we impose the following inequality on the local evolution s of alternative structured models:

$$\frac{\partial \rho}{\partial x}(x) \cdot [\widehat{\mu}(x) + \sigma(x)s] + \frac{1}{2}\text{trace}\left[\sigma(x)' \frac{\partial^2 \rho}{\partial x \partial x'}(x)\sigma(x)\right] + \frac{|s|^2}{2} \leqslant \frac{\mathsf{q}^2}{2}. \tag{25}$$

Integrating both sides with respect to dQ implies that

$$\int \frac{|\eta|^2}{2} dQ \leqslant \frac{\mathsf{q}^2}{2} \tag{26}$$

where Q satisfies

$$\int \left[\frac{\partial \rho}{\partial x} \cdot (\widehat{\mu} + \sigma\eta) + \frac{1}{2}\text{trace}\left(\sigma' \frac{\partial^2 \rho}{\partial x \partial x'}\sigma\right)\right] dQ = 0.$$

We impose structure on models by specifying ρ. Because it is quadratic in s, inequality (25) imposes a state-dependent restriction on s that is easy to implement. The local restriction (25) goes beyond restricting entropy because it deliberately precludes the intertemporal tradeoffs that are allowed by a relative entropy constraint.

One way to proceed is to posit an alternative drift configuration $\eta(x)$ and then to solve (24) for ρ. But other models also satisfy inequality (25) for the same ρ. We provide an illustration in section 7. An extreme example imposes that $\frac{\partial \rho}{\partial x}(x) = 0$, which is equivalent to restricting $\frac{|S_t|^2}{2}$ to be less than or equal to a constant $\frac{\mathsf{q}^2}{2}$ every instance. Our section 7 application will lead us naturally to consider state-dependent (in particular, quadratic) specifications of ρ.

To illustrate how we use local restriction (25), we again depict preferences with an

[16] Furthermore, embedding this set in one that is rectangular would yield too large a set in a sense described in section 10.

instantaneous utility function $\delta\phi(x)$ and a subjective discount rate δ. The decision problem that replaces problem 4.1 has HJB equation

Problem 4.2.

$$
\begin{aligned}
0 = \min_s & -\delta\Phi(x) + \phi(x) + \widehat{\mu}(x) \cdot \frac{\partial \Phi}{\partial x}(x) + [\sigma(x)s] \cdot \frac{\partial \Phi}{\partial x}(x) \\
& + \frac{1}{2}\text{trace}\left[\sigma(x)' \frac{\partial^2 \Phi}{\partial x \partial x'}(x)\sigma(x)\right] - \frac{1}{2\theta}\left[\frac{\partial \Phi}{\partial x}(x)\right]' \sigma(x)\sigma(x)'\left[\frac{\partial \Phi}{\partial x}(x)\right]
\end{aligned} \tag{27}
$$

where minimization over s is subject to

$$
\frac{\partial \rho}{\partial x}(x) \cdot [\widehat{\mu}(x) + \sigma(x)s] + \frac{1}{2}\text{trace}\left[\sigma(x)' \frac{\partial^2 \rho}{\partial x \partial x'}(x)\sigma(x)\right] + \frac{|s|^2}{2} \leqslant \frac{\mathsf{q}^2}{2}.
$$

We can recover a minimizing u from a minimizing $s^*(x)$ via $u^*(x) = s^*(x) - \frac{1}{\theta}\sigma(x)'\frac{\partial \Phi^*}{\partial x}(x)$, where Φ^* solves HJB equation (27).

In appendix A we construct a different representation of the constraint set and verify that it is not empty.

5 Uncertainty about Macroeconomic Growth

To prepare the way for the quantitative illustration in section 7, this section describes a particular macro-finance setting. In the tradition of Lucas (1978), it features a representative agent who faces an exogenous aggregate consumption process.[17] From a robust planning problem, we deduce shadow prices that equal prices of risk and uncertainty in a competitive equilibrium.

We start with a baseline parametric model for a representative investor's consumption process, then form a family of parametric structured probability models. We deduce the pertinent version of HJB equation (27) that describes the value function attained by worst-case drift distortions S and U. The baseline model is

$$
\begin{aligned}
dY_t &= .01\left(\widehat{\alpha}_y + \widehat{\beta} Z_t\right) dt + .01\sigma_y \cdot dW_t \\
dZ_t &= (\widehat{\alpha}_z - \widehat{\kappa} Z_t)\, dt + \sigma_z \cdot dW_t.
\end{aligned} \tag{28}
$$

[17]More generally, a fictitious planner could solve a resource allocation problem that involves accumulating physical capital and other factors of production and that makes consumption endogenous.

We scale by .01 because we want to work with growth rates and Y is typically expressed in logarithms. Let

$$X = \begin{bmatrix} Y \\ Z \end{bmatrix}.$$

Notice that the drift $\widehat{\alpha}_z - \widehat{\kappa} z$ has a zero at

$$\bar{z} = \frac{\widehat{\alpha}_z}{\widehat{\kappa}},$$

and that $\widehat{\alpha}_z - \widehat{\kappa} z = -\widehat{\kappa}(z - \bar{z})$; $\bar{z}$ is the mean of Z_t in the stationary distribution under the baseline model.

We focus on the following collection of structured parametric models:

$$\begin{aligned} dY_t &= .01\left(\alpha_y + \beta Z_t\right) dt + .01 \sigma_y \cdot dW_t^S \\ dZ_t &= \left(\alpha_z - \kappa Z_t\right) dt + \sigma_z \cdot dW_t^S, \end{aligned} \tag{29}$$

where W^S is a Brownian motion and (6) continues to describe the relationship between the processes W and W^S. Collection (29) nests the baseline model (28). Here $(\alpha_y, \beta, \alpha_z, \kappa)$ are parameters that distinguish the structured models (29) from the baseline model, and (σ_y, σ_z) are parameters common to models (28) and (29).

We represent members of a parametric class defined by (29) in terms of our section 2.1 structure with drift distortions S of the form

$$S_t = \eta(Z_t) \equiv \eta_0 + \eta_1 (Z_t - \bar{z}),$$

then use (1), (6), and (29) to deduce the following restrictions on η_1:

$$\sigma \eta_1 = \begin{bmatrix} \beta - \widehat{\beta} \\ \widehat{\kappa} - \kappa \end{bmatrix}.$$

where

$$\sigma = \begin{bmatrix} (\sigma_y)' \\ (\sigma_z)' \end{bmatrix}.$$

Relative entropy $\frac{\mathsf{q}^2}{2}$ emerges from the solution to differential equation (24) appropriately specialized:

$$\frac{|\eta(z)|^2}{2} + \frac{d\rho}{dz}(z)[-\widehat{\kappa}(z-\bar{z}) + \sigma_z \cdot \eta(z)] + \frac{|\sigma_z|^2}{2}\frac{d^2\rho}{dz^2}(z) - \frac{\mathsf{q}^2}{2} = 0. \tag{30}$$

Under the parametric alternatives (29), the solution for ρ is quadratic in $z - \bar{z}$. Write:

$$\rho(z) = \rho_1(z-\bar{z}) + \frac{1}{2}\rho_2(z-\bar{z})^2.$$

As described in Appendix B, we compute ρ_1 and ρ_2 by matching coefficients on the terms $(z-\bar{z})$ and $(z-\bar{z})^2$, respectively. Matching constant terms then implies $\frac{\mathsf{q}^2}{2}$. In restricting structured models, we impose:

$$\frac{|S_t|^2}{2} + [\rho_1 + \rho_2(Z_t - \bar{z})]\left[-\widehat{\kappa}(Z_t - \bar{z}) + \sigma_z \cdot S_t\right] + \frac{|\sigma_z|^2}{2}\rho_2 - \frac{\mathsf{q}^2}{2} \leqslant 0. \tag{31}$$

Suppose that $Y = \log C$, where C is consumption, δ is a subjective rate of discount and instantaneous utility $\phi(x) = y$. Let $r = \sigma s$. Let $\Phi(x) = y + \widehat{\Phi}(z)$ be a value function for a robust planner; $\hat{\Phi}(x)$ solves the HJB equation

$$\begin{aligned} 0 = \min_r & -\delta\widehat{\Phi}(z) + .01[\widehat{\alpha}_y + \widehat{\beta}z + r_1] + [-\widehat{\kappa}(z-\bar{z}) + r_2]\frac{d\widehat{\Phi}}{dz}(z) \\ & + \frac{1}{2}|\sigma_z|^2\frac{d^2\widehat{\Phi}}{dz^2}(z) - \frac{1}{2\theta}\begin{bmatrix} .01 & \frac{d\hat{\Phi}}{dz}(z)\end{bmatrix}\sigma\sigma'\begin{bmatrix} .01 \\ \frac{d\hat{\Phi}}{dz}(z)\end{bmatrix} \end{aligned} \tag{32}$$

where the minimization is subject to

$$\frac{1}{2}r'\Lambda r + [\rho_1 + \rho_2(z-\bar{z})]\left[-\widehat{\kappa}(z-\bar{z}) + r_2\right] + \frac{|\sigma_z|^2}{2}\rho_2 - \frac{\mathsf{q}^2}{2} \leqslant 0 \tag{33}$$

and $\Lambda = (\sigma')^{-1}\sigma^{-1}$. A worst-case structured model induces a worst-case unstructured model via equation (21). (In the portfolio problem of section 8, we will also maximize over portfolio weights and the consumption process C.)

For a given $\widehat{\Phi}$ and state realization z, the component of the objective that depends on r is the inner product

$$\begin{bmatrix} .01 & \frac{d\hat{\Phi}}{dz}(z)\end{bmatrix} r.$$

That this component is linear in r pushes the solution to an ellipsoid that is the boundary of the convex constraint set for each z. Figure **??** shows ellipoids associated with two alternative values of z and baseline parameters that we present in section 7. For every feasible choice of r_2, two choices of r_1 satisfy the implied quadratic equation. Provided that $\frac{d\hat{\Phi}}{dz}(z) > 0$, which is true in our calculations, we take the lower of the two solutions for r_1. The solution occurs at a point on the lower left of the ellipsoid where $\frac{dr_1}{dr_2} = -100\frac{d\hat{\Phi}}{dz}(z)$ and depends on z, as figure **??** indicates.

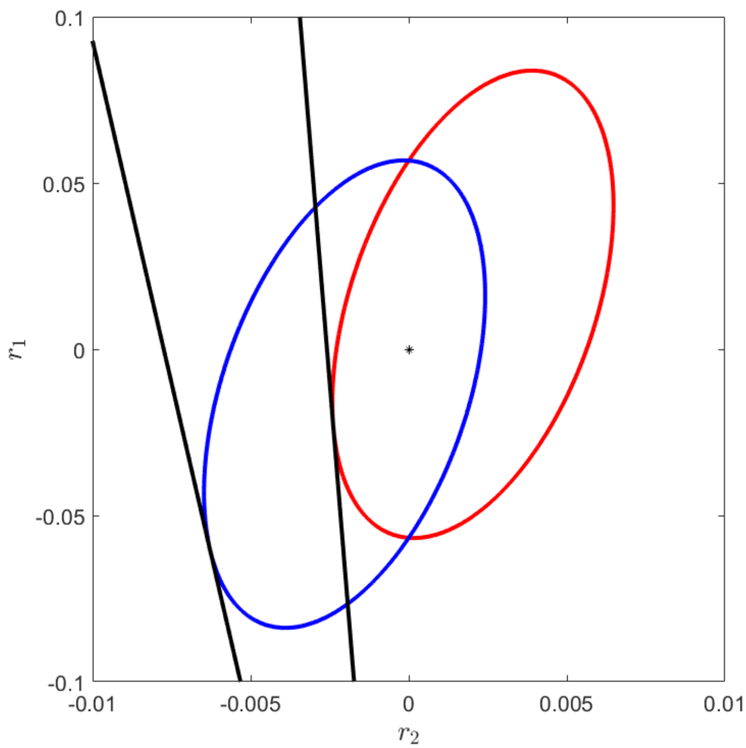


Figure 1: An illustration for section 7 parameter configuration, the figure 3 configuration for $\mathsf{q}_{s,0} = .1$ and $\mathsf{q}_{u,s} = .2$. The figure displays parameter contours for (r_1, r_2), holding relative entropy fixed. The upper right contour depicted in red is for z equal to the .1 quantile of the stationary distribution under the baseline model and the lower left contour is for z at the .9 quantile. The dot depicts the $(r_1, r_2) = (0, 0)$ point corresponding to the baseline model. Tangency points denote worst-case structured models.

By prespecifying $(\rho_1, \rho_2, \mathsf{q})$, we trace out a one-dimensional family of parametric models with the same relative entropy. For instance, we can solve equation (30) for η_0 and η_1. By matching a constant, a linear term, and a quadratic term in $z - \bar{z}$, we obtain three equations in four unknowns that imply a one dimensional curve for η_0 and η_1 that imply nonlinear

S_t's. In this way, nonlinear structured models emerge endogenously. These nonlinear models will also have relative entropy $\frac{\mathsf{q}^2}{2}$. We can represent the resulting nonlinear model as a time-varying coefficient model by solving

$$r^*(z) = \sigma \left[\eta_0 + \eta_1(z - \bar{z})\right]$$

for η_0 and η_1 along the one-dimensional curve z by z as illustrated in example 5.1. By using this construction, our relative entropy restriction (31) can also be depicted as in example 3.1. See appendix A for a more complete derivation.

We will feature the following special case in some of our calculations.

Example 5.1. *Suppose that*

$$\eta(z) = \eta_1(z - \bar{z}),$$

which focuses structured uncertainty on how drifts for (Y, Z) respond to the state variable Z. In this case, $\rho_1 = 0$ and

$$-\frac{\mathsf{q}^2}{2} + \frac{1}{2}\rho_2|\sigma_z|^2 = 0,$$

or equivalently,

$$\rho_2 = \frac{\mathsf{q}^2}{|\sigma_z|^2}.$$

Notice that restriction (31) implies that

$$S_t = 0$$

when $Z_t = \bar{z}$. To connect this to a time-varying parameter specification, first construct the convex set of η_1's that satisfy

$$\frac{1}{2}\eta_1 \cdot \eta_1 + \left(\frac{\mathsf{q}^2}{|\sigma_z|^2}\right)\left[-\widehat{\kappa} + \sigma_z \cdot \eta_1\right] \leqslant 0.$$

Next form the boundary of the convex set Π by solving

$$\sigma\eta_1 = \begin{bmatrix} (\beta - \widehat{\beta}) \\ (\widehat{\kappa} - \kappa) \end{bmatrix}$$

for (β, κ) associated with alternative choices of η_1. This illustrates how using a convex set Π constrained in this way in the HJB equation for problem 4.1 is equivalent to imposing *the restricted version ρ in the HJB equation of problem 4.2.*

6 Inspecting variational preferences

Preference orderings described in section 4 use the penalization parameter θ to limit the model misspecifications that concern a decision maker. We want the decision maker to follow Good's (1952) advice to evaluate a max-min expected utility approach by verifying that a worst-case model is plausible.[18] We implement Good's recommendation by first using entropy to measure how far a worst-case model is from a set of structured models, and then using the outcome to restrict the penalty parameter θ in HJB equation (20).

We consider two entropy concepts that quantify a statistical discrepancy of a probability model generated by a martingale M^S from a model generated by a martingale M^U, where logarithms of M^S and M^U both evolve according to appropriate versions of (7), namely,

$$\begin{aligned} d \log M_t^S &= -\frac{1}{2}|S_t|^2 dt + S_t \cdot dW_t \\ d \log M_t^U &= -\frac{1}{2}|U_t|^2 dt + U_t \cdot dW_t. \end{aligned}$$

Think of a pairwise model selection problem that statistically compares a structured model generated by a martingale M^S with an unstructured model generated by a martingale M^U. For a given value of θ, we compute worst-case structured and unstructured models in terms of the drift distortions

$$\begin{aligned} S_t &= \eta_s(Z_t) \\ U_t &= \eta_u(Z_t) \end{aligned}$$

implied for example by the minimization that appears in decision problem 4.2.

6.1 Relative entropy

Relative entropy is one measuure of divergence between probabilities. Relative entropy is an expected log likelihood ratio:

$$\Lambda(M^U, M^S) = \lim_{t \to \infty} \frac{1}{t} E\left[M_t^U \left(\log M_t^U - \log M_t^S\right) | \mathcal{F}_0\right].$$

[18]See Berger (1994) and Chamberlain (2000) for related discussions.

Since worst-case structured and unstructured probability models are both Markovian, we can compute $\Lambda(M^U, M^S)$ using the same procedures that we applied to compute entropy relative to the baseline model. In particular, we solve

$$\frac{\partial \rho}{\partial x}(x) \cdot (\widehat{\mu} + \sigma \eta_u) + \frac{1}{2}\text{trace}\left(\sigma' \frac{\partial^2 \rho}{\partial x \partial x'} \sigma\right) + \frac{|\eta_u - \eta_s|^2}{2} \leqslant \frac{\mathsf{q}^2}{2}$$

for ρ (up to a constant translation) and $\frac{\mathsf{q}^2}{2}$. Appendix C describes our computational approach. Entropy concept $\Lambda(M^U, M^S)$ is typically independent of the date zero conditioning information when the Markov process is asymptotically stationary. In our application, we find it enlightening to report the following transformed object that measures the magnitude of the drift distortion:

$$\mathsf{q}_{u,s} = \sqrt{2\Lambda(M^U, M^S)}.$$

6.2 Chernoff entropy

As an alternative measure of probability measure divergence, we also consider a version of Chernoff entropy that we construct as a dynamic counterpart to Chernoff's (1952) divergence concept. Chernoff entropy emerges from studying how, by disguising distortions of a baseline probability model, Brownian motions make it challenging to distinguish models statistically. Chernoff entropy's connection to a statistical decision problem makes it interesting, but it is less tractable than relative entropy. Anderson et al. (2003) used Chernoff entropy measured as a local rate to draw direct connections between magnitudes of market prices of uncertainty and statistical discrimination. That local rate is state dependent and for diffusion models proportional to the local drift in relative entropy. Quantitative differences arise when we measure statistical discrepancy globally as did Newman and Stuck (1979). We shall characterize a long-run version of Chernoff entropy and show how to compute it.

Think of a pairwise model selection problem that statistically compares a structured model generated by a martingale M^S with an unstructured model generated by a martingale M^U. Consider a statistical model selection rule based on a data history of length t that takes the form of a log likelihood comparison $\log M_t^U - \log M_t^S \geqslant \mathsf{h}$. This selection rule sometimes incorrectly chooses the unstructured model when the structured model governs the data. We can bound the probability of this outcome by using an argument from large deviations theory that starts from

$$
\begin{aligned}
\mathbf{1}_{\left\{\log M_t^U - \log M_t^S \geqslant \mathsf{h}\right\}} &= \mathbf{1}_{\left\{-\gamma\left(\mathsf{h} + \log M_t^U - \log M_t^S\right) \geqslant 0\right\}} \\
&= \mathbf{1}_{\left\{\exp(-\gamma \mathsf{h})(M_t^U)^{\gamma}(M_t^S)^{-\gamma} \geqslant 1\right\}} \\
&\leqslant \exp(-\gamma \mathsf{h})(M_t^U)^{\gamma}(M_t^S)^{-\gamma},
\end{aligned}
$$

where the inequality holds for $0 \leqslant \gamma \leqslant 1$. Under a structured model, the mathematical expectation of the term on the left side multiplied by M_t^S equals the probability of mistakenly selecting the alternative model when data are a sample of size t generated under the structured model. We bound this mistake probability for large t by following Donsker and Varadhan (1976) and Newman and Stuck (1979) and studying

$$
\lim_{t \to \infty} \frac{1}{t} \log E\left[\exp(-\gamma \mathsf{h}) \left(M_t^U\right)^{\gamma} \left(M_t^S\right)^{1-\gamma} | \mathcal{F}_0\right] = \lim_{t \to \infty} \frac{1}{t} \log E\left[\left(M_t^U\right)^{\gamma} \left(M_t^S\right)^{1-\gamma} | \mathcal{F}_0\right]
$$

for alternative choices of γ. We apply these calculations for specifications of U and S, checking that the limits are well defined. The threshold h does not affect the limit. Furthermore, the limit is often independent of the initial conditioning information. To get the best bound, we compute

$$
\inf_{0 \leqslant \gamma \leqslant 1} \lim_{t \to \infty} \frac{1}{t} \log E\left[\left(M_t^U\right)^{\gamma} \left(M_t^S\right)^{1-\gamma} | \mathcal{F}_0\right],
$$

which is typically negative because mistake probabilities decay with sample size. Chernoff entropy is then

$$
\Gamma(M^U, M^S) = -\inf_{0 \leqslant \gamma \leqslant 1} \liminf_{t \to \infty} \frac{1}{t} \log E\left[\left(M_t^U\right)^{\gamma} \left(M_t^S\right)^{1-\gamma} | \mathcal{F}_0\right]. \tag{34}
$$

Setting $\Gamma(M^U, M^S) = 0$ would include only alternative models M^U that cannot be distinguished from M^S on the basis of histories of infinite length.[19] Because we want to include more possible alternative models than that, we entertain positive values of $\Gamma(M^U, M^S)$.

To interpret $\Gamma(M^U, M^S)$, consider the following. If the decay rate of mistake probabili-

[19] That is what is done in models that extend the rational expectations equilibrium concept to self-confirming equilibria that allow probability models that are wrong only off equilibrium paths, i.e., for events that in equilibrium do not occur infinitely often. See Fudenberg and Levine (1993, 2009) and Sargent (1999). Our decision theory differs from that used in most of the literature on self confirming equilibria because our decision maker acknowledges model uncertainty and wants to adjust decisions accordingly. But see Battigalli et al. (2015).

ties were constant, say d, then mistake probabilities for two sample sizes $T_i, i = 1, 2$, would be

$$\text{mistake probability}_i = \frac{1}{2} \exp\left(-T_i \mathsf{d}_{u,s}\right)$$

for $\mathsf{d}_{u,s} = \Gamma(M^U, M^S)$. We define a half-life as an increase in sample size $T_2 - T_1 > 0$ that multiplies a mistake probability by a factor of one half:

$$\frac{1}{2} = \frac{\text{mistake probability}_2}{\text{mistake probability}_1} = \frac{\exp\left(-T_2 \overline{\chi}\right)}{\exp\left(-T_1 \mathsf{d}\right)},$$

so the half-life is approximately

$$T_2 - T_1 = \frac{\log 2}{\mathsf{d}}. \tag{35}$$

The bound on the decay rate should be interpreted cautiously because the actual decay rate is not constant. Furthermore, the pairwise comparison understates the true challenge, which is statistically to discriminate among *multiple* models.

A symmetrical calculation reverses the roles of the two models and instead conditions on the perturbed model implied by martingale M^U. The limiting rate remains the same. Thus, when we select a model by comparing a log likelihood ratio to a constant threshold, the two types of mistakes share the same asymptotic decay rate.

To implement Chernoff entropy, we follow an approach suggested by Newman and Stuck (1979). Because our worst case models are Markovian, we use Perron-Frobenius theory to characterize

$$\lim_{t \to \infty} \frac{1}{t} \log E\left[\left(M_t^U\right)^{\gamma} \left(M_t^S\right)^{1-\gamma} | \mathcal{F}_0\right]$$

for a given $\gamma \in (0, 1)$ as a dominant eigenvalue of a semigroup of linear operators. The limit does not depend on the initial state x and is characterized as a dominant eigenvalue associated with an eigenfunction that is strictly positive.

Appendix C describes how we evaluate Chernoff entropy numerically for the nonlinear Markov specifications that we use in subsequent sections.

7 Quantitative example

Our quantitative example builds on section 5 and features a representative investor who wants to explore utility consequences of alternative models portrayed by $\{M_t^U\}$ and $\{M_t^S\}$ processes, some of which contribute a troublesome and difficult to detect predictable components of consumption growth.[20] Relative entropy and Chernoff entropy shape and quantify the doubts that we impute to investors.

7.1 Baseline model

Our example blends elements of Bansal and Yaron (2004) and Hansen et al. (2008). We use a vector autoregression (VAR) to construct a quantitative version of a baseline model like (28) that approximates responses of consumption to permanent shocks. In contrast to Bansal and Yaron (2004), we assume no stochastic volatility because we want to focus exclusively on fluctuations in uncertainty prices that are induced by the representative investor's specification concerns.

Our VAR follows Hansen et al. (2008) in using additional macroeconomic time series to infer information about long-term consumption growth. We deduce a calibration of our baseline model (28) from a trivariate VAR for the first difference of log consumption, the difference between logs of business income and consumption, and the difference between logs of personal dividend income and consumption. This specification makes consumption, business income, and personal dividend income be cointegrated.[21] Since we presume that all three time series grow, we know the coefficients in the cointegrating relation. In Appendix D we tell how we used the discrete time VAR estimates to deduce the following parameters for the baseline model (28):

$$\begin{array}{rclcrcl} \widehat{\alpha}_y & = & .386 & \quad & \widehat{\beta} & = & 1 \\ \widehat{\alpha}_z & = & 0 & \quad & \widehat{\kappa} & = & .019 \end{array}$$

$$\sigma \doteq \begin{bmatrix} (\sigma_y)' \\ (\sigma_z)' \end{bmatrix} = \begin{bmatrix} .488 & 0 \\ .013 & .028 \end{bmatrix} \tag{36}$$

[20] While we appreciate the value of a more comprehensive empirical investigation with multiple macroeconomic time series, here our aim is to illustrate a mechanism within the context of relatively simple time series models of predictable consumption growth.

[21] Business income is measured as proprietor's income plus corporate profits per capita. Dividends are personal dividend income per capita. The time series are quarterly data from 1948 Q1 to 2015 Q1. Our consumption measure is nondurables plus services consumption per capita. The business income data are from NIPA Table 1.12 and the dividend income from NIPA Table 7.10. By including proprietors' income in addition to corporate profits, we use a broader measure of business income than Hansen et al. (2008) who used only corporate profits. Hansen et al. (2008) did not include personal dividends in their VAR analysis.

We suppose that $\delta = .002$ and $\phi = y$, where y is the logarithm of consumption. Under this model, the standard deviation of the Z process in the implied stationary distribution is .158.

7.2 Structured models and a robust plan

We solve HJB equation (32) for three different configurations of structured models.

7.2.1 Uncertain growth rate responses

We compute a solution by first focusing on an Example 5.1 specification in which $\rho_1 = 0$ and ρ_2 satisfies:

$$\rho_2 = \frac{\mathsf{q}^2}{|\sigma_z|^2}.$$

When η is restricted to be $\eta_1(z - \bar{z})$, a given value of q imposes a restriction on η_1 and implicitly on (β, κ). Figure 2 plots iso-entropy contours for (β, κ) for $\mathsf{q} = .1$ and $\mathsf{q} = .05$.

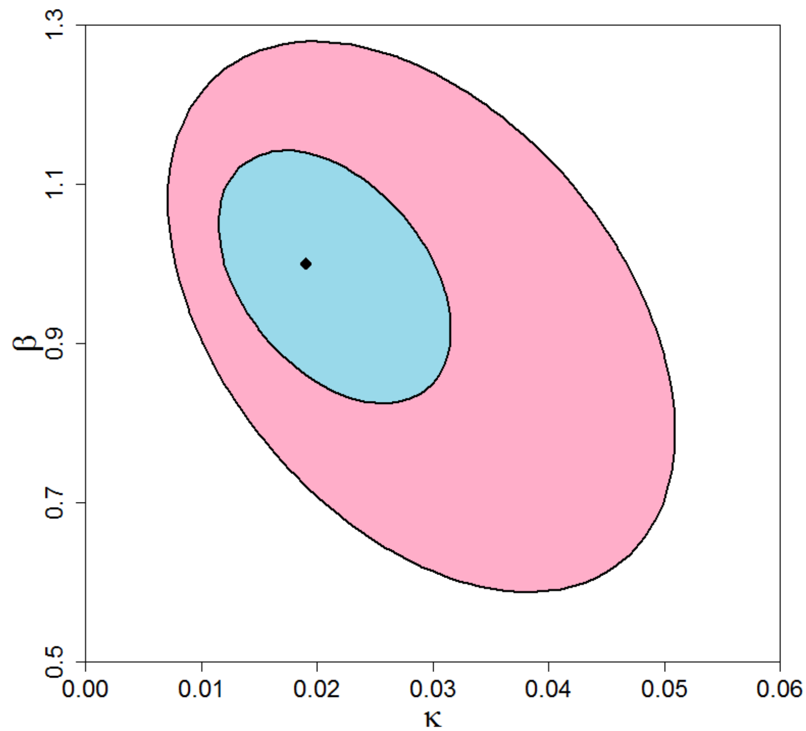


Figure 2: Parameter contours for (β, κ) holding relative entropy fixed. The outer curve depicts $\mathsf{q}_{s,0} = .1$ and the inner curve $\mathsf{q}_{s,0} = .05$. The small diamond depicts the baseline model.

While Figure 2 displays contours of time invariant parameters with the same relative entropy, the robust planner chooses a two-dimensional vector of drift distortions r for a structured model in a more flexible way. As happens when there is uncertainty about (β, κ), the set of possible r's differs depending on the state z. As we remarked earlier, when $z = \bar{z}$ the only feasible r is $r = 0$. Figure 1 also reported iso-entropy contours when z is at the .01 and .9 quantile of the stationary distribution under the baseline model. The larger value of z results in a lower downward shift of the contour relative to the smaller value of z. The tangent lines in figure 1 have slopes equal to $-100\frac{d\hat{V}}{dz}$ where the point of tangency is the worst-case structured model. This point occurs at a lower drift distortion for the .9 quantile than for the .1 quantile.

Consider next the adjustment for model misspecification. Since

$$\sigma(u^* - s^*) = -\frac{1}{\theta}\sigma\sigma' \begin{bmatrix} .01 \\ \frac{d\hat{V}}{dz} \end{bmatrix}$$

and entries of $\sigma\sigma'$ are positive, the adjustment for model misspecification is smaller in magnitude for larger values of the state z. Taken together, the vector of drift distortions is:

$$\sigma u^* = \sigma(u^* - s^*) + r^*.$$

The first term on the right is smaller in magnitude for a larger z and conversely, the second term is larger in magnitude for smaller z.

Under the restrictions on structured models now under study that $\rho_1 = 0$ and $\rho_2 = \frac{\mathsf{q}^2}{|\sigma_z|^2}$ and that η is restricted to be $\eta_1(z - \bar{z})$, the first derivative of the value function is not differentiable at $z = \bar{z}$. We can compute the value function and the worst-case models by solving two coupled HJB equations, one for $z < \bar{z}$ and another for $z > \bar{z}$. We obtain two second-order differential equations in value functions and their derivatives; these value functions coincide at $z = \bar{z}$, as do their first derivatives.

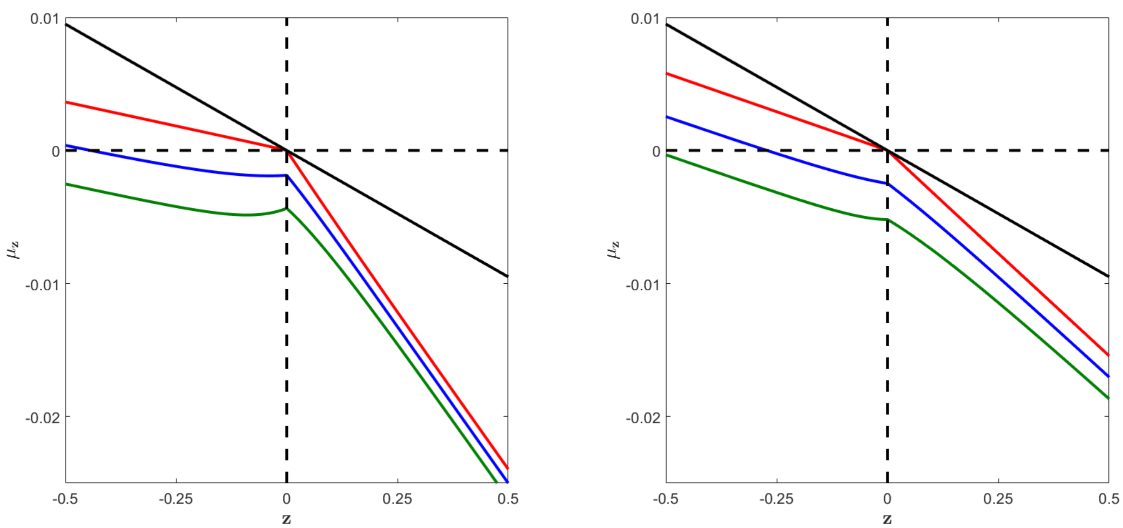


Figure 3: Worst-case structured model growth rate drifts. Left panel: larger structured entropy ($\mathsf{q}_{s,0} = .1$). Right panel: smaller structured entropy ($\mathsf{q}_{s,0} = .05$). The penalization parameter θ set to hit targeted values of $\mathsf{q}_{u,s}$. **Red**: worst-case structured model; **blue**: $\mathsf{q}_{u,s} = .1$; and **green**: $\mathsf{q}_{u,s} = .2$.

Figure 3 shows adjustments of the drifts due to ambiguity aversion and concerns about misspecification of the structured models. Setting $\theta = \infty$ silences concerns about misspecification of the structured models, all of which are expressed through minimization over s. When we set $\theta = \infty$, the implied worst-case structured model has state dynamics that take the form of a threshold autoregression with a kink at zero. The distorted drifts again show less persistence than does the baseline model for negative values of z and more persistence for larger values of z. We activate a concern for misspecification of the structured models by setting θ to attain targeted values of $\mathsf{q}_{u,s}$ computed using the structured and unstructured worst-case models. This adjustments shifts the implied worst-case drift as a function of the state downwards, even more so for negative values of z than for positive ones. The impact of the drift for y is much more modest.

$\mathsf{q}_{s,0}$	$\mathsf{q}_{u,s}$	$\mathsf{d}_{u,s}$	half life u, s	$\mathsf{q}_{u,0}$	$\mathsf{d}_{u,0}$	half life $u, 0$
.10	.10	.0010	671	.29	.0036	192
.10	.20	.0047	148	.54	.0107	65
.05	.10	.0011	623	.18	.0024	288
.05	.20	.0047	148	.33	.0077	89

Table 1: Entropies and half lives. $\frac{1}{2}\mathsf{q}^2$ measures relative entropy and d measures Chernoff entropy. The subscripts denote the probability models used in performing the computations.

Table 1 reports Chernoff and relative entropies implied by structured and unstructured worst-case models. The first two columns tell the relative entropy magnitudes that we imposed by adjusting the value of θ. The remaining columns report other measures of entropy as implied by these settings. Recall that the q's measure magnitudes of the drift distortions under associated distorted measures. Thus, $\mathsf{q}_{u,0}$ measures how large the drift distortion is relative to the baseline model. As expected, increasing the targeted values of $\mathsf{q}_{s,0}$ and $\mathsf{q}_{u,s}$ increases the implied values $\mathsf{q}_{u,0}$. There is one perhaps puzzling finding. From table 1, we see that

$$\mathsf{q}_{u,s} + \mathsf{q}_{s,0} < \mathsf{q}_{u,0},$$

which does not satisfy a Triangle Inequality because while $\mathsf{q}_{u,s}$ and $\mathsf{q}_{u,0}$ are computed under the stationary probability measure implied by the worst-case unstructured model induced by U, $\mathsf{q}_{s,0}$ is computed under the measure implied by worst-case structured model.

Table 1 also reports Chernoff entropies and their implied half lives. . These numbers indicate that statistical discrimination is challenging for all four configurations, since even the smallest half-life exceeds 65 quarters. Discrimination is especially challenging when we limit the extent of model misspecification by setting $q_{u,s} = .1$. In terms of how the entropy measures are related, we know of no formula that transforms relative entropy into long-run Chernoff entropy, but a formula from by Anderson et al. (2003) is valid locally and leads us to expect that

$$\frac{\mathsf{q}^2}{2} \approx 4\mathsf{d},$$

an approximation that becomes exact when relative drift distortions are constant. This is evidently a good approximation for the computed $\mathsf{q}_{u,s}$ and $\mathsf{d}_{u,s}$, but not for $\mathsf{q}_{u,0}$ and $\mathsf{d}_{u,0}$. As we have seen, the composite drift distortions show substantial state dependence because of the impact of the worst-case structured model.

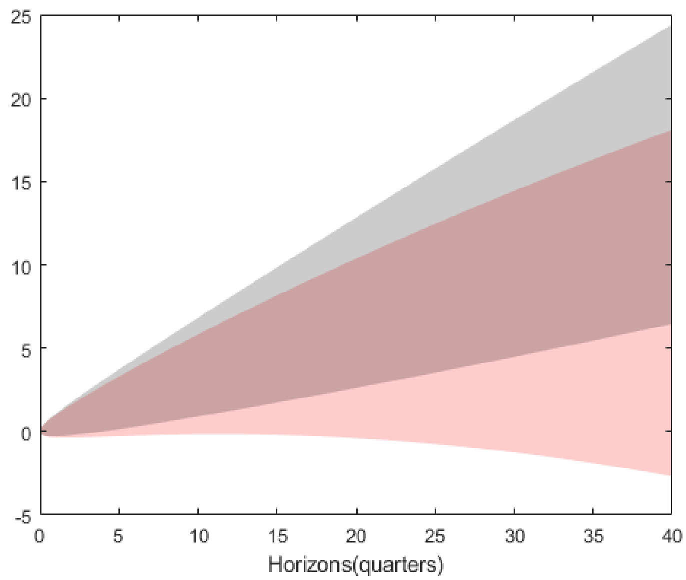


Figure 4: Distribution of $Y_t - Y_0$ under the baseline model and worst-case model for $\mathsf{q}_{s,0} = .1$ and $\mathsf{q}_{u,s} = .2$. The gray shaded area depicts the interval between the .1 and .9 deciles for every choice of the horizon under the baseline model. The red shaded area gives the region within the .1 and .9 deciles under the worst-case model.

Figure 4 extrapolates impacts of the drift distortion on distributions of future consumption growth over alternative horizons. It shows how the consumption growth distribution adjusted for ambiguity aversion and misspecification tilts down relative to the baseline distribution.

7.2.2 Altering the scope of uncertainty

Until now, we have restricted

$$\rho_2 = \frac{\mathsf{q}}{|\sigma_z|^2}$$

with the implication that the alternative structured models have no drift distortions for Z at $Z_t = \bar{z}$. We now alter this restriction by cutting the value of ρ_2 in half. Consequences of this change are depicted in the right panel of Figure 5. For sake of comparison, this figure includes the previous specification in the left panel. The worst-case structured drifts no longer coincide with the baseline drift at $z = \bar{z}$ and vary smoothly in the vicinity of $z = \bar{z}$.

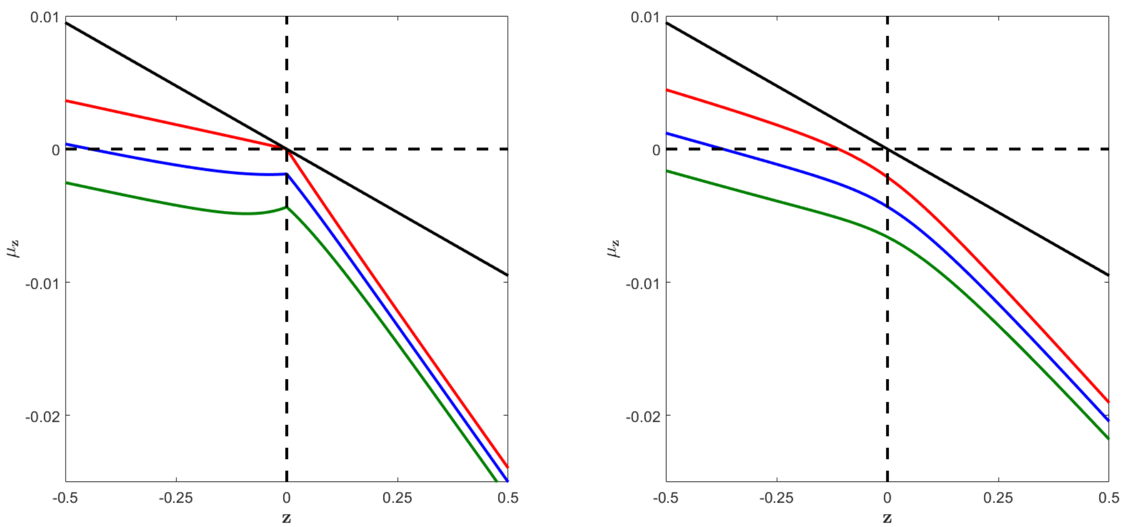


Figure 5: Distorted growth rate drift for Z. Relative entropy $\mathsf{q}_{s,0} = .1$. Left panel: $\rho_2 = \frac{(.01)}{|\sigma_z|^2}$. Right panel: $\rho_2 = \frac{(.01)}{2|\sigma_z|^2}$. **red**: worst-case structured model; **blue**: $\mathsf{q}_{u,s} = .1$; and **green**: $\mathsf{q}_{u,s} = .2$.

Adding the restriction that $\rho_2 = 0$ makes the robust planner's value function become linear and the minimizing s and u become constant and therefore independent of z. Specifically,

$$\frac{d\widehat{\Phi}}{dz} = .01 \frac{\widehat{\beta}}{\delta + \widehat{\kappa}},$$

and

$$s^* \propto -\sigma' \begin{bmatrix} .01 \\ \frac{.01}{\delta + \widehat{\kappa}} \end{bmatrix}$$

$$u^* - s^* = -\frac{1}{\theta}\sigma' \begin{bmatrix} .01 \\ \frac{.01}{\delta+\widehat{\kappa}} \end{bmatrix}$$

The constant of proportionality for s^* is determined by the constraint $|s^*| = \mathsf{q}$. So setting ρ_1 and ρ_2 to zero results in parallel downward shifts of the baseline drifts of worst-case drifts for both Y and Z. This amounts to changing the coefficients α_y and α_z in ways that are time invariant and leave $\kappa = \widehat{\kappa}$ and $\beta = \widehat{\beta}$.

8 Robust portfolio choice and pricing

In this section, we describe equilibrium prices that make a representative investor willing to bear risks accurately approximated by baseline model (1) in spite of his concerns about model misspecification. We construct equilibrium prices by computing shadow prices from the robust planning problem of section 4. We decompose equilibrium risk prices into distinct compensations for bearing risk and for bearing model uncertainty. We begin by posing the representative investor's portfolio choice problem.

8.1 Robust investor portfolio problem

A representative investor solves a continuous-time Merton portfolio problem in which individual wealth K evolves as

$$dK_t = -C_t dt + K_t \iota(Z_t) dt + K_t A_t \cdot dW_t + K_t \omega(Z_t) \cdot A_t dt, \tag{37}$$

where $A_t = a$ is a vector of chosen risk exposures, $\iota(z)$ is an instantaneous risk free rate, and $\omega(z)$ is a vector of risk prices evaluated at state $Z_t = z$. Initial wealth is K_0. The investor discounts the logarithm of consumption and distrusts his probability model.

Key inputs to a representative investor's robust portfolio problem are the baseline model (1), the wealth evolution equation (37), the vector of risk prices $\omega(z)$, and the quadratic function ρ and relative entropy $\frac{\mathsf{q}^2}{2}$ that define alternative structured models.

Under a guess that the value function takes the form $\widetilde{\Phi}(z) + \log k + \log \delta$, the HJB equation for the robust portfolio allocation problem is

$$0 = \max_{a,c} \min_{u,s} -\delta\widetilde{\Phi}(z) - \delta \log k - \delta \log \delta + \delta \log c - \frac{c}{k} + \iota(z)$$

$$+ \omega(z) \cdot a + a \cdot u - \frac{|a|^2}{2} + \frac{d\widetilde{\Phi}}{dz}(z)\left[-\widehat{\kappa}(z-\bar{z}) + \sigma_z \cdot u\right]$$
$$+ \frac{1}{2}|\sigma_z|^2 \frac{d^2\widetilde{\Phi}}{dz^2}(z) + \frac{\theta}{2}|u-s|^2 \tag{38}$$

subject to
$$\frac{|s|^2}{2} + \frac{d\rho}{dz}(z)[-\widehat{\kappa}(z-\bar{z}) + \sigma_z \cdot s] + \frac{|\sigma_z|^2}{2}\frac{d^2\rho}{dz^2}(z) - \frac{\mathsf{q}^2}{2} = 0. \tag{39}$$

First-order conditions for consumption are

$$\frac{\delta}{c^*} = \frac{1}{k},$$

which imply that $c^* = \delta k$, an implication that follows from the unitary elasticity of intertemporal substitution associated with the logarithmic instantaneous utility function. First-order conditions for a and u are

$$\omega(z) + u^* - a^* = 0 \tag{40a}$$
$$a^* + \theta(u^* - s^*) + \frac{d\widetilde{\Phi}}{dz}(z)\sigma_z = 0. \tag{40b}$$

These two equations determine a^* and $u^* - s^*$ as functions of $\omega(z)$ and the value function $\widetilde{\Phi}$. We determine s^* as a function of u^* by solving

$$\min_s \frac{\theta}{2}|u-s|^2$$

subject to (39). Taken together, these determine (a^*, u^*, s^*). We can appeal to arguments like those of Hansen and Sargent (2008, ch. 7) to justify stacking first-order conditions as a way to collect equilibrium conditions for the two-person zero-sum game that the robust portfolio problem solves.[22]

8.2 Competitive equilibrium prices

We now impose $\log C = Y$ as an equilibrium condition. We show here that the drift distortion η^* that emerges from the robust planner's problem of section 5 determines prices

[22] An alternative timing protocol that allows the maximizing player to take account of the impact of its decisions on the minimizing agent implies the same equilibrium decision rules described in the text. See Hansen and Sargent (2008, ch. 5).

that a competitive equilibrium awards for bearing model uncertainty. To compute a vector $\omega(x)$ of competitive equilibrium risk prices, we find a robust planner's marginal valuations of exposures to the W shocks. We decompose that price vector into separate compensations for bearing *risk* and for accepting *model uncertainty*.

Noting from the robust planning problem that the shock exposure vectors for $\log K$ and Y must coincide implies

$$a^* = (.01)\sigma_y.$$

From (40b) and the solution for s^*

$$u^* = \eta^*(z),$$

where η^* can be shown to be the worst-case drift from the robust planning problem provided that we can show that $\widetilde{\Phi} = \widehat{\Phi}$, where $\widehat{\Phi}$ is the value function for the robust planning problem. Thus, from (40a), $\omega = \omega^*$, where

$$\omega^*(z) = (.01)\sigma_y - \eta^*(z). \tag{41}$$

Similarly, in the problem for a representative investor within a competitive equilibrium, the drifts for $\log K$ and Y coincide:

$$-\delta + \iota(z) + [(.01)\sigma_y - \eta^*(z)] \cdot a^* - \frac{.0001}{2}\sigma_y \cdot \sigma_y = (.01)(\hat{\alpha}_y + \hat{\beta} z),$$

so that $\iota = \iota^*$, where

$$\iota^*(z) = \delta + .01(\widehat{\alpha}_y + \widehat{\beta} z) + .01\sigma_y \cdot \eta^*(z) - \frac{.0001}{2}\sigma_y \cdot \sigma_y. \tag{42}$$

We use these formulas for equilibrium prices to construct a solution to the HJB equation of a representative investor in a competitive equilibrium by letting $\widetilde{\Phi} = \widehat{\Phi}$.

8.3 Local uncertainty prices

The equilibrium stochastic discount factor process Sdf for our robust representative investor economy is

$$d\log Sdf_t = -\delta dt - .01\left(\widehat{\alpha}_y + \widehat{\beta} Z_t\right) dt - .01\sigma_y \cdot dW_t + U_t^* \cdot dW_t - \frac{1}{2}|U_t^*|^2 dt. \tag{43}$$

The components of the vector $\omega^*(Z_t)$ given by (41) equal minus the local exposures to the Brownian shocks. While these are usually interpreted as local "risk prices," we shall reinterpret them. The decomposition

$$\textbf{minus stochastic discount factor exposure} = \underbrace{.01\sigma_y}_{\textbf{risk price}} \quad \underbrace{-U_t^*}_{\textbf{uncertainty price}},$$

motivates us to think of $.01\sigma_y$ as risk prices induced by the curvature of log utility and $-U_t^*$ as "uncertainty" prices induced by a representative investor's doubts about the baseline model. Here U_t^* is state dependent. Local prices are large in both good and bad macroeconomic growth states. Prices of longer horizons will behave differently.

8.4 Uncertainty prices over alternative investment horizons

We now report shock-price elasticities for exposures to future uncertainty. These are related to but distinct from objects computed by Borovička et al. (2014). We use a different timing convention than they do because we want to answer a different set of questions. In particular, Borovička et al. report elasticities that tell the impact of changing the next period exposure to a shock on the expected return on a hypothetical asset with payoff τ periods into the future. In contrast, here we shift the change in the exposure to τ time periods in the future, the same time as the asset payoff. We then study the current period impact on the expected return as the investment horizon $\tau > 0$ varies. We express responses as elasticities by normalizing the exposure change to be a unit standard deviation and focusing on logs of expected returns. The shock-price elasticities we report here are designed to enlighten us about how state dependence in exposures to future shocks impacts the current period expected return over alternative investment horizons. As we will see, elasticities defined in this way link pricing to relative entropy.

We let consumption be the hypothetical payoff of interest. The logarithm of the expected return from a consumption payoff at date t consists of two terms:

$$\log E\left(\frac{C_t}{C_0}\middle| X_0 = x\right) - \log E\left[Sdf_t\left(\frac{C_t}{C_0}\right)\middle| X_0 = x\right]. \tag{44}$$

where $\log C_t = Y_t$. The first term is an expected payoff and the second is the cost of purchasing that payoff. Our example imposes a unitary elasticity of substitution

$$Sdf_t \left(\frac{C_t}{C_0} \right) = M_t^{U*},$$

so the second term features a martingale contributed by the representative investor's concerns about misspecificaion.

An elasticity tells changes in an expected return that result from a local change in the exposure of consumption to the underlying Brownian motion. Malliavin derivatives are important inputs into calculating a shock-price elasticity. These derivatives measure how a shock at a given date affects consumption and stochastic discount factor processes. Both Sdf_t and C_t depend on the Brownian motion between dates zero and t. We are particularly interested in the impact of a date t shock. Computing the derivative of the logarithm of the expected return given in (44) results in

$$\frac{E\left[\mathcal{D}_t C_t | \mathcal{F}_0\right]}{E\left[C_t | \mathcal{F}_0\right]} - E\left[\mathcal{D}_t M_t^{U*} | \mathcal{F}_0\right]$$

where $\mathcal{D}_t C_t$ and $\mathcal{D}_t M_t^{U*}$ denote two-dimensional vectors of Malliavin derivatives with respect to the two dimensional Brownian increment at date t for consumption and the worst-case martingale, respectively.

A formula familiar from other forms of differentiation implies

$$\mathcal{D}_t C_t = C_t \left(\mathcal{D}_t \log C_t \right).$$

The Malliavin derivative of $\log C_t = Y_t$ is the vector $.01\sigma_y$, which is the exposure vector of $\log C_t$ to the Brownian increment dW_t:

$$\mathcal{D}_t C_t = .01 C_t \sigma_y,$$

so

$$\frac{E\left[\mathcal{D}_t C_t | \mathcal{F}_0\right]}{E\left[C_t | \mathcal{F}_0\right]} = .01\sigma_y.$$

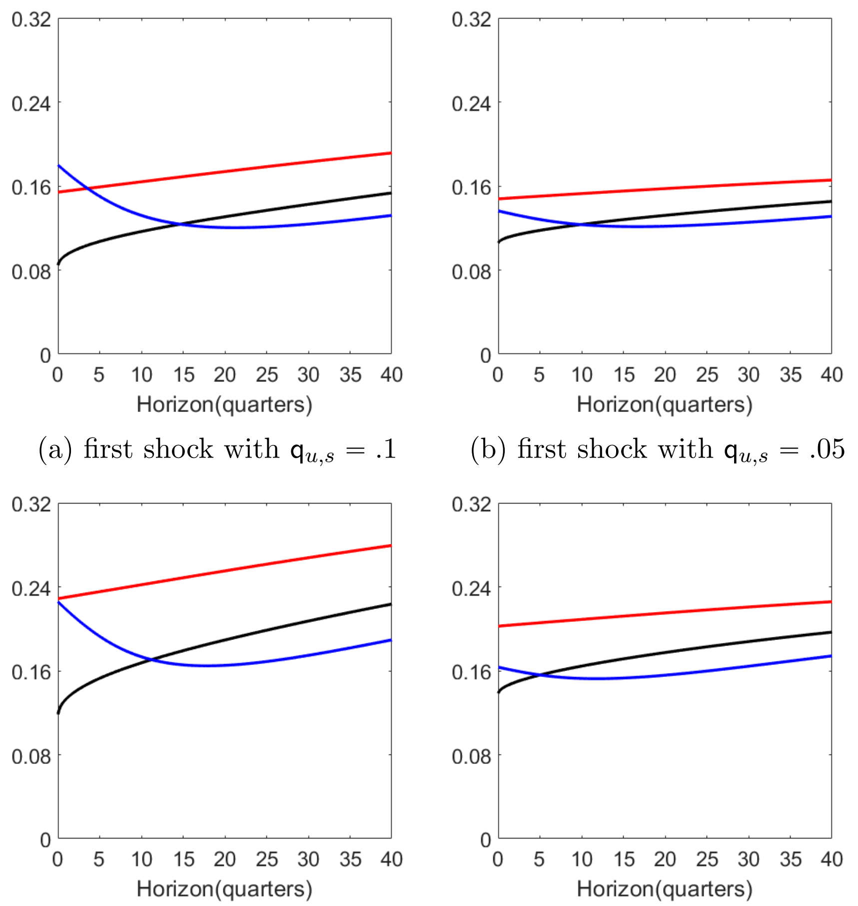
Similarly,

$$\mathcal{D}_t M_t^{U*} = U_t^*.$$

Therefore, the term structure of prices that interests us is

$$.01\sigma_y - E\left[M_t^{U^*}U_t^*|\mathcal{F}_0\right]. \tag{45}$$

The first term is the risk price familiar from consumption-based asset pricing. It is state independent and contributes a (small) term that is independent of the horizon. In contrast, the equilibrium drift distortion in the second term provides a state dependent component; its expectation under the distorted probability measure



(a) first shock with $\mathsf{q}_{u,s} = .1$ (b) first shock with $\mathsf{q}_{u,s} = .05$

(c) second shock with $\mathsf{q}_{u,s} = .1$ (d) second shock with $\mathsf{q}_{u,s} = .05$

Figure 6: Shock price elasticities for alternative horizons. The change in the exposure occurs at the same future date as the consumption payoff. The figure reports the median and deciles for the section 5 specification with (β, κ) structured uncertainty. **Black**: median of the Z stationary distribution **red**: .1 decile; and **blue**: .9 decile.

Figure 6 shows shock price elasticities for our economy. Notice that although the price elasticity is initially smaller for the median specification of z than for the .9 quantile, this inequality is eventually reversed as the horizon increases. (The blue and black curves cross.) The uncertainty price for positive z initially diminishes because the probability measure implied by the martingale M_t^U has reduced persistence for positive states. Under the M_t^U probability, the growth rate state variable is expected to spend less time in the positive region. This is reflected in smaller prices at the .9 quantile than at the median over longer investment horizons. For longer investment horizons, but not necessarily for very short ones, an endogenous nonlinearity makes uncertainty prices larger for negative values of z than for positive values of z. Horizon dependence is an important avenue through which concerns about misspecification and ambiguity aversion influence valuations of assets.

There is an intriguing connection between long-horizon prices and relative entropy. While the uncertainty price trajectories do not converge over the time span reported in figure 6, well defined limiting uncertainty prices do emerge over longer time horizons.[23] These limits equal $E^{M^{U^*}}[-u^*]$, i.e., expectations of the corresponding drift distortions computed under the worst-case probability measures. In table 2, we compare these limit prices to the relative entropy divergence $\mathsf{q}_{u,0}$, which measures the overall magnitude of these distortions by $\sqrt{2E^{M^{U^*}}[|u^*|^2]}$, i.e., the square root of twice the expected square of the absolute value of the vector of drift distortions, also under the worst-case probability measures. Indeed these mean contributions account for most of the relative entropy measures.

$\mathsf{q}_{s,0}$	$\mathsf{q}_{u,s}$	$\mathsf{q}_{u,0}$	shock one price	shock two price
.10	.20	.54	.30	.44
.05	.20	.33	.19	.27

Table 2: Entropies and limit prices. $\frac{1}{2}\mathsf{q}^2$ denotes relative entropy. The limiting long-horizon prices are the expectations of $-U^*$ under the probability model implied by U^*.

We have designed our quantitative examples to investigate a particular mechanism that causes statistically plausible amounts of uncertainty to generate fluctuations in uncertainty prices. We infer parameters of the baseline model for these examples solely from time series of macroeconomic quantities, thus completely ignoring asset prices during calibration. As

[23] Hansen and Scheinkman (2012) study a limiting growth rate risk price that is based on a different conceptual experiment but leads to a similar characterization. Whereas formula (45) has an adjustment for the current consumption exposure to shocks, the limiting Hansen and Scheinkman measure replaces this term by the proportionate exposure of the martingale component of consumption. Both adjustments are small in our quantitative example.

a consequence, we do not expect to track high frequency movements in financial markets well. By limiting our empirical inputs, we respect concerns that Hansen (2007) and Chen et al. (2015) expressed about using asset market data to calibrate macro-finance models that assign a special role to investors' beliefs about the future asset prices.[24]

9 Learning and dynamic consistency

We have made the set of unstructured models that concerns our decision maker so vast and some of the structured models themselves so complicated that our decision maker thinks that it is pointless to learn his way out of model ambiguity as he observes more data. Had we featured only time invariant models, there would be ways a decision maker could learn, but ambiguity would still add a source of variation to valuations. Even if we were to begin with a family of time invariant models, confining $\mathcal{M}^o$ to time-invariant parameter models would be too restrictive for at least two reasons One is that time invariance precludes learning from new information. Another is that the passage of time alters what a decision maker cares about.

Consider first learning. Until now, we have supposed that the set of models of interest to our representative investor makes learning particularly difficult. But had we restricted that set of models enough, learning would be possible. For time invariant parameter models with unknown parameters, endowing a decision maker with a family of conjugate priors could make it tractable to construct a coresponding family of posteriors by repeatedly applying Bayes law. But as the following example illustrates, learning breaks time invariance:

Example 9.1. *Apply Bayes' rule to a finite collection of models characterized by S^j where M^{S^j} is in $\mathcal{M}^o$ for $j = 1, \ldots, n$. Let $\pi_0^j \geqslant 0$ be a prior probability assigned to model S^j, where $\sum_{i=1}^n \pi_0^j = 1$. A martingale*

$$M = \sum_{j=1}^{n} \pi_0^j M^{S^j}$$

corresponds to a mixture of S^j models. The mathematical expectation of M conditioned on

[24]Hansen (2007) and Chen et al. (2015) describe situations in which it is the behavior of rates of return on assets that, through the cross-equation restrictions, lead an econometrician to make inferences about the behavior of macroeconomic quantities like consumption that are much more confident than can be made from the quantity data alone. That opens questions about how the investors who are supposedly putting those cross-equation restrictions into returns came to know those quantity processes before *they* observed returns.

date zero information equals unity. The law of motion of M is

$$\begin{aligned} dM_t &= \sum_{j=1}^{n} \pi_0^j dM_t^{S^j} \\ &= \sum_{j=1}^{n} \pi_0^j M_t^{S^j} S_t^j \cdot dW_t \\ &= M_t \left(\pi_t^j S_t^j \right) \cdot dW_t \end{aligned}$$

where π_t^j is the date t posterior

$$\pi_t^j = \frac{\pi_0^j M_t^{S^j}}{M_t}.$$

The drift distortion is

$$S_t = \sum_{j=1}^{n} \pi_t^j S_t^j.$$

The example illustrates how Bayes' rule leads naturally to a particular form of history-dependent weights on the S_t^j's that characterize alternative models.

Another reason for history dependence is that a decision maker with a nontrivial set of priors (i.e., a robust Bayesian) would want to evaluate the utility consequences of a set of posteriors implied by Bayes' law from different perspectives as time passes. With an aversion to ambiguity, a robust Bayesian would rank alternative plans by minimizing expected continuation utilities over the set of posteriors. Epstein and Schneider (2003) note that for many possible sets of models and priors, this approach induces a form of dynamic inconsistency.

Thus, consider a given plan. A decision maker has more information at $t > 0$ than at $t = 0$ and he cares only about the continuation of the plan for dates $s \geqslant t$. To evaluate a plan under ambiguity aversion at $t > 0$, the decision maker would minimize continuation utility over the set of date zero priors. Changes in perspective would in general lead the decision maker to choose different worst-case date zero priors as time passes. A date t conditional preference order could conflict with a date 0 preference order. This possibility led Epstein and Schneider to study implications of a dynamic consistency axiom.

To make preferences satisfy that axiom, they argue that the decision maker's set of probabilities should satisfy a property that they call rectangularity. A rectangular family of probabilities is formed by i) specifying a set of possible local (i.e., instantaneous) transitions for each t, and ii) constructing *all* possible joint probabilities having such local transitions. The set of probabilities implied by martingales in $\mathcal{M}^o$ satisfying the time-separability restriction (10) satisfies this property.[25]

Epstein and Schneider make

> ... an important conceptual distinction between the set of probability laws that the decision maker views as possible, such as *Prob*, and the set of priors P that is part of the representation of preference.

Regardless of whether they are subjectively or statistically plausible, Epstein and Schneider recommend augmenting a decision maker's original set of "possible" probabilities (i.e., their *Prob*) with enough additional probabilities to make an enlarged set (i.e., their P) satisfy a condition that suffices to render conditional preferences orderings dynamically consistent as required by their axioms.

We can illustrate what Epstein and Schneider's procedure does and does not accomplish within the setting of Example 9.1 with $n = 2$. Suppose that we have a set of priors $\underline{\pi}_0^1 \leqslant \pi_0^1 \leqslant \overline{\pi}_0^1$. For each π_0^1, we can use Bayes' rule to construct a posterior residing in an interval $[\underline{\pi}_t^1, \overline{\pi}_t^1]$, an associated set of drift processes $\{S_t : t \geqslant 0\}$, and implied probability measures over the filtration $\{\mathcal{F}_t : t \geqslant 0\}$. This family of probabilities is typically not rectangular in the sense of Epstein and Schneider. To obtain a smallest rectangular family that contains these probabilities, we construct the larger space $\{S_t : t \geqslant 0\}$ with $S_t \in \Xi_t$, where

$$\Xi_t = \left\{\pi_t^1 S_t^1 + (1 - \pi_t^1) S_t^2, \ \underline{\pi}_t^1 \leqslant \pi_t \leqslant \overline{\pi}_t^1, \ \pi_t \text{ is } \mathcal{F}_t \text{ measurable}\right\} \tag{46}$$

Augmenting the set $\{S_t : t \geqslant 0\}$ in this way makes conditional preference orderings over plans remain the same as time passes. But this expanded set of probabilities includes elements that can emerge from *no* single date zero prior. Thus, in constructing the set $\{S_t : t \geqslant 0\}$, Epstein and Schneider allow different date zero priors at each calendar date t. Doing that intertemporally disconnects restrictions on local transition probabilities.[26]

The failure of Epstein and Schneider's procedure to yield a unique prior capable of justifying their dynamically consistent preference ordering undermines the useful concept called admissibility that is widely applied in statistical decision theory. An admissible decision rule is one that cannot be dominated under all possible probability specifications entertained by a decision maker. Verifying optimality against a unique worst-case model

[25]Rectangularity, per se, does not require Ξ_t to be convex, a property that we impose for other reasons.

[26]This approach could be made tractable by using a family of conjugate priors that enable updating via Bayes law by applying recursive methods.

is a common way to establish that a statistical decision rule is admissible. Epstein and Schneider's proposal to achieve dynamic consistency by adding probabilities to those that the decision maker thinks are possible renders the resulting decision rule inadmissible and disables Good (1952)'s sensible recommendation for assessing the suitability of max-min decision making.[27]

10 Relative entropy and rectangularity

Our decision maker starts with a set of structured probability models that happen to be rectangular in the sense of Epstein and Schneider. But our decision maker's concern that all structured models are misspecified leads him to explore the utility consequences of unstructured probability models that are not rectangular, even though as measured by relative entropy they are statistically close to models in the rectangular set.

An alternative approach would be first to construct a set that includes relative entropy neighborhoods of all martingales in $\mathcal{M}^o$. For instance, we could start with a set

$$\overline{\mathcal{M}} = \left\{ M^U \in \mathcal{M} : \Theta(M^U | \mathcal{F}_0) < \epsilon \right\} \tag{47}$$

that yields a set of implied probabilities that are not rectangular. At this point, why not follow Epstein and Schneider's (2003) recommendation to add enough martingales to attain a rectangular set of probability measures? Our answer is that doing so would include all martingales in $\mathcal{M}$ – a set much too large for a max-min decision analysis.

To show this, it suffices to look at relative entropy neighborhoods of the baseline model.[28] To construct a rectangular set of models that includes the baseline model, for a fixed date τ, consider a random vector $\overline{U}_\tau$ that is observable at τ and that satisfies

$$E\left(|\overline{U}_\tau|^2 \mid \mathcal{F}_0 \right) < \infty. \tag{48}$$

[27]Presumably, an advocate of Epstein and Schneider's dynamic consistency axiom could respond that admissibility is too limiting in a dynamic context because it commits to a time 0 perspective and does not allow a decision maker to reevaluate later. Nevertheless, it is common in the control theory literature to maintain just such a date zero perspective and in effect solve a commitment problem under ambiguity aversion.

[28]Including additional structured models would only make the set of martingales larger.

Form a stochastic process

$$U_t^h = \begin{cases} 0 & 0 \leqslant t < \tau \\ \overline{U}_\tau & \tau \leqslant t < \tau + h \\ 0 & t \geqslant \tau + h. \end{cases} \tag{49}$$

The martingale M^{U^h} associated with U^h equals one both before time τ and after time $\tau + h$. Compute relative entropy:

$$\begin{aligned} \Delta(M^{U^h}|\mathcal{F}_0) &= \left(\frac{1}{2}\right) \int_\tau^{\tau+h} \exp(-\delta t) E\left[M_t^{U^h} |\overline{U}_\tau|^2 dt \Big| \mathcal{F}_0 \right] dt \\ &= \left[\frac{1 - \exp(-\delta h)}{2\delta} \right] \exp(-\delta \tau) E\left(|\overline{U}_\tau|^2 \mid \mathcal{F}_0 \right). \end{aligned}$$

Evidently, relative entropy $\Delta(M^{U^h}|\mathcal{F}_0)$ can be made arbitrarily small by shrinking h to zero. This means that any rectangular set that contains $\overline{\mathcal{M}}$ must allow for a drift distortion $\overline{U}_\tau$ at date τ. We summarize this argument in the following proposition:

Proposition 10.1. *Any rectangular set of probabilities that contains the probabilities induced by martingales in* (47) *must also contain the probabilities induced by any martingale in* $\mathcal{M}$.

This rectangular set of martingales allows us too much freedom in setting date τ and random vector $\overline{U}_\tau$: all martingales in the set $\mathcal{M}$ identified in definition 2.1 are included in the smallest rectangular set that embeds the set described by (47). That set is too big to pose a meaningful decision problem.

11 Concluding remarks

This paper formulates and applies a tractable model of the effects on equilibrium prices of exposures to macroeconomic uncertainties. We use models' consequences for discounted expected utilities to quantify investors' concerns about model misspecification. We characterize the effects of concerns about misspecification of a baseline stochastic process for individual consumption as shadow prices for a planner's problem that supports competitive equilibrium prices.

To illustrate our approach, we have focused on the growth rate uncertainty featured in the "long-run risk" literature initiated by Bansal and Yaron (2004). Other applications seem natural. For example, the tools developed here could shed light on a recent public debate between two groups of macroeconomists, one prophesying secular stagnation because of technology growth slowdowns, the other discounting those pessimistic forecasts. The tools that we describe can be used, first, to quantify how challenging it is to infer persistent changes in growth rates, and, second, to guide macroeconomic policy in light of available empirical evidence.

Specifically, we have produced a model of a log stochastic discount factor whose uncertainty prices reflect a robust planner's worst-case drift distortions U^* and we have argued that these drift distortions should be interpreted as prices of model uncertainty. The dependence of these uncertainty prices U^* on the growth state z is shaped partly by the alternative parametric models that the decision maker entertains. In this way, our theory of state dependence in uncertainty prices is all about how our robust investor responds to the presence of the alternative parametric models among a huge set of unspecified alternative models that also concern him.

It is worthwhile comparing this paper's way of inducing time varying prices of risk with three other macro/finance models that also get them. Campbell and Cochrane (1999) proceed in the standard rational expectations single-known-probability-model tradition and so exclude any fears of model misspecification from the mind of their representative investor. They construct a utility function in which the history of consumption expresses an externality. This history dependence makes the investor's local risk aversion depend in a countercyclical way on the economy's growth state. Ang and Piazzesi (2003) use an exponential quadratic stochastic discount factor in a no-arbitrage statistical model and explore links between the term structure of interest rates and other macroeconomic variables. Their approach allows movements in risk prices to be consistent with historical evidence without specifying an explicit general equilibrium model. A third approach introduces stochastic volatility into the macroeconomy by positing that the volatilities of shocks driving consumption growth are themselves stochastic processes. A stochastic volatility model induces time variation in risk prices via exogenous movements in the conditional volatilities impinging on macroeconomic variables. A related approach is implemented by Ulrich (2013) and Ilut and Schneider (2014), who posit exogenous stochastic fluctuations in ambiguity concerns to induce additional macroeconomic fluctuations.

In Hansen and Sargent (2010), countercyclical uncertainty prices are driven by a representative investor's robust model averaging. The investor carries along two difficult-to-distinguish models of consumption growth, one with substantial growth rate dependence and the other with little such dependence. The investor uses observations on consumption growth to update a Bayesian posterior over these models and expresses his specification distrust by pessimistically exponentially twisting a posterior over alternative models. That leads the investor to act as if good news is temporary and bad news is persistent, an outcome that is qualitatively similar to what we have found here. Learning occurs in Hansen and Sargent's analysis because the parameterized structured models are time invariant and hence learnable.

In this paper, we propose a different way to make uncertainty prices vary in a qualitatively similar way. We exclude learning and instead consider alternative models with parameters whose future variations cannot be inferred from historical data. These time-varying parameter models differ from the decision maker's baseline model, a fixed parameter model whose parameters can be well estimated from historical data. The alternative models include ones that allow parameters persistently to deviate from those of the baseline model in statistically subtle and time-varying ways. In addition to this particular parametric class of alternative models, the decision maker also worries about other specifications. The robust planner's worst-case model responds to these forms of model ambiguity partly by having more persistence in bad states and less persistence in good states. Adverse shifts in a worst-case shock distribution that drive up the absolute magnitudes of uncertainty prices were also present in some of our earlier work (for example, see Hansen et al. (1999) and Anderson et al. (2003)). In this paper, we induce state dependence in uncertainty prices in a new way, namely, by specifying a set of alternative models to capture concerns about the baseline model's specification of persistence in consumption growth.

Our continuous-time formulation (18) exploits mathematically convenient properties of a Brownian information structure. There is a discrete-time counterpart to our formulation that starts from a baseline model cast in terms of a nonlinear stochastic difference equation. In that formulation, there are counterparts to structured and unstructured models that play the same roles that they do in the present continuous time formulation. Furthermore, preference orderings defined in terms of continuation values are dynamically consistent.

While our example used entropy measures to restrict the decision maker's set of structured models, two other approaches could be employed instead. One would use a more direct implementation of a robust Bayesian approach; the other would refrain from imposing absolute continuity when constructing a family of structured models.

We illustrated how one might start with structured models that are time invariant and a convex set of priors over the invariant parameters. Provided that the resulting set of posteriors could be characterized date-by-date and computed easily, say through the use of conjugate priors, this approach could be tractable. But after a rectangular augmentation of a set of probabilities, the implied worst case structured model would typically not emerge from applying Bayes' rule to a single prior. That prevents applying Good's advice about assessing the plausibility of max-min choice theory. On the other hand, a rectangular structure may place models on the table that are substantively interesting in their own right, including possibly the worst case structured model. By incorporating a concern for misspecification, this would provide an alternative to the approach to robust learning in Hansen and Sargent (2007).

In this paper we assumed that the structured model probabilities can be represented as martingales with respect to a baseline model. An alternative approach invented by Peng (2004) uses a theory of stochastic differential equations under a broad notion of ambiguity that is rich enough to allow for uncertainty in the conditional volatility of the Brownian increments. Alternative probability specifications there fail to be absolutely continuous and standard likelihood ratio analysis ceases to apply. If we knew how to construct bounds on uncertainty under a nondegenerate rectangular embedding, we could extend the construction of worst-case structured models and still restrain relative entropy as a way to limit the unstructured models to be explored.[29]

[29]See Epstein and Ji (2014) for an application of the Peng analysis to asset pricing that does not use relative entropy.

Appendices

A Structured model restrictions

Consider the constraint in problem 4.2. To verify that the constraint set is not empty, suppose that there exists an η such that

$$\frac{|\eta(x)|^2}{2} + \frac{\partial \rho}{\partial x}(x) \cdot [\widehat{\mu}(x) + \sigma(x)\eta(x)] + \frac{1}{2}\text{trace}\left[\sigma(x)' \frac{\partial^2 \rho}{\partial x \partial x'}(x)\sigma(x)\right] - \frac{\mathsf{q}^2}{2} = 0.$$

Next pose the problem

$$\min_s \frac{|s|^2}{2} + \frac{\partial \rho}{\partial x}(x) \cdot [\widehat{\mu}(x) + \sigma(x)s] + \frac{1}{2}\text{trace}\left[\sigma(x)' \frac{\partial^2 \rho}{\partial x \partial x'}(x)\sigma(x)\right] - \frac{\mathsf{q}^2}{2},$$

whose solution

$$\tilde{s}(x) = -\sigma(x)' \frac{\partial \rho}{\partial x}(x)$$

attains a minimized objective function

$$\begin{aligned} -\frac{\Upsilon(x)}{2} &\equiv -\frac{1}{2}\left[\frac{\partial \rho}{\partial x}(x)\right]' \sigma(x)\sigma(x)' \left[\frac{\partial \rho}{\partial x}(x)\right] + \frac{\partial \rho}{\partial x}(x) \cdot \widehat{\mu}(x) + \frac{1}{2}\text{trace}\left[\sigma(x)' \frac{\partial^2 \rho}{\partial x \partial x'}(x)\sigma(x)\right] - \frac{\mathsf{q}^2}{2} \\ &\leqslant 0. \end{aligned}$$

For convenience, write the constraint as:

$$\frac{|s - \tilde{s}(x)|^2}{2} \leqslant \frac{\Upsilon(x)}{2}. \tag{50}$$

Since $\Upsilon(x)$ is nonnegative for each x, minimizing solutions exist and reside on an ellipsoid centered at $\tilde{s}(x)$.

B Computing relative entropy

In this appendix we show how to compute relative entropies for parametric models of the form (29). Recall that relative entropy $\frac{\mathsf{q}^2}{2}$ emerges as part of the solution to HJB equation

(27) appropriately specialized:

$$\frac{|\eta(z)|^2}{2} + \frac{d\rho}{dz}(z)[-\widehat{\kappa}(z-\bar{z}) + \sigma_z \cdot \eta(z)] + \frac{|\sigma_z|^2}{2}\frac{d^2\rho}{dz^2}(z) - \frac{\mathsf{q}^2}{2} = 0.$$

where $\bar{z} = \frac{\widehat{\alpha}_z}{\widehat{\kappa}}$ and

$$\eta(z) = \eta_0 + \eta_1(z-\bar{z}).$$

Under our parametric alternatives, the solution for ρ is quadratic in $z-\bar{z}$. Write:

$$\rho(z) = \rho_1(z-\bar{z}) + \frac{1}{2}\rho_2(z-\bar{z})^2.$$

Compute ρ_2 by targeting only the terms of the HJB equation that involve $(z-\bar{z})^2$:

$$\frac{\eta_1 \cdot \eta_1}{2} + \rho_2\left[-\widehat{\kappa} + \sigma_z \cdot \eta_1\right] = 0.$$

Thus

$$\rho_2 = \frac{\eta_1 \cdot \eta_1}{2\left(\widehat{\kappa} - \sigma_z \cdot \eta_1\right)}$$

Given ρ_2, compute ρ_1 by targeting only the terms in $(z-\bar{z})$:

$$\eta_0 \cdot \eta_1 + \rho_2\left(\sigma_z \cdot \eta_0\right) + \rho_1\left(-\widehat{\kappa} + \sigma_z \cdot \eta_1\right) = 0.$$

Thus

$$\rho_1 = \frac{\eta_0 \cdot \eta_1}{\widehat{\kappa} - \sigma_z \cdot \eta_1} + \frac{\left(\eta_1 \cdot \eta_1\right)\left(\sigma_z \cdot \eta_0\right)}{2\left(\widehat{\kappa} - \sigma_z \cdot \eta_1\right)^2}.$$

Finally, calculate q by targeting the remaining constant terms:

$$\frac{\eta_0 \cdot \eta_0}{2} + \rho_1\left(\sigma_z \cdot \eta_0\right) + \rho_2\frac{|\sigma_z|^2}{2} - \frac{\mathsf{q}^2}{2} = 0.$$

Thus

$$\frac{\mathsf{q}^2}{2} = \frac{\eta_0 \cdot \eta_0}{2} + \frac{\eta_0 \cdot \eta_1\left(\sigma_z \cdot \eta_0\right)}{\widehat{\kappa} - \sigma_z \cdot \eta_1} + \frac{\eta_1 \cdot \eta_1\left(+\sigma_z \cdot \eta_0\right)^2}{2\left(\widehat{\kappa} - \sigma_z \cdot \eta_1\right)^2} + \frac{\eta_1 \cdot \eta_1|\sigma_z|^2}{4\left(\widehat{\kappa} - \sigma_z \cdot \eta_1\right)}.$$

The formula could alternatively be derived by computing the expectation of $\frac{|\eta(Z_t)|^2}{2}$ under the altered distribution.

C Computing Chernoff and relative entropy

In this appendix we show how to compute Chernoff and relative entropies for Markov specifications where the associated S's and U's take the forms

$$
\begin{aligned}
U_t &= \eta_u(Z_t) \\
S_t &= \eta_s(Z_t).
\end{aligned}
$$

C.1 Chernoff entropy

Given the implied Markov structure of both models, we compute Chernoff entropy by using an eigenvalue approach of Donsker and Varadhan (1976) and Newman and Stuck (1979). We start by computing the drift of $\left(M_t^U\right)^\gamma \left(M_t^S\right)^{1-\gamma} g(Z_t)$ for $0 \leqslant \gamma \leqslant 1$ at $t = 0$:

$$
\begin{aligned}
[\mathbb{G}(\gamma)g](z) \doteq& -\frac{\gamma(1-\gamma)}{2}|\eta_u(z) - \eta_s(z)|^2 g(z) + g(z)'\sigma \cdot [\gamma\eta_u(z) + (1-\gamma)\eta_s(z)] \\
&+ g'(z)\left(\widehat{\alpha}_z - \widehat{\kappa} z\right) + \frac{g''(z)}{2}|\sigma_z|^2,
\end{aligned}
$$

where $[\mathbb{G}(\gamma)g](x)$ is the drift given that $Z_0 = z$. Next we solve the eigenvalue problem

$$
[\mathbb{G}(\gamma)]e(z,\gamma) = -\lambda(\gamma)e(z,\gamma).
$$

We seek the eigenvalue for which $\exp[-\lambda(\gamma)]$ is largest in magnitude; the associated eigenfunction is positive.

We compute Chernoff entropy by solving

$$
\Gamma(M^H, M^S) = \max_{\gamma\in[0,1]} \lambda(\gamma),
$$

where we compute $\lambda(\gamma)$ numerically using a finite-difference approach. For a prespecified γ, We evaluate $[\mathbb{G}(\gamma)]g$ at each of n grid points and replacing derivatives by two-sided symmetric differences except at the edges where we use corresponding one-sided differences. For each such grid point, this gives a linear transformation of g evaluated at the n grid points. The outcome of this calculation is an n by n matrix applied to a vector containing the entries of g evaluated at the n grid points. The eigenvalue of the resulting matrix that has the largest exponential equals $-\eta(\gamma)$. We we use a grid for z over the interval $[-2.5, 2.5]$ with grid increments equal to .01, choices that imply that $n = 501$.

C.2 Relative entropy

Using an approach similar to that applied in section C.1, we solve

$$\frac{\mathsf{q}^2}{2} - \frac{d\rho}{dz}(z)[\widehat{\alpha}_z - \widehat{\kappa} z + \sigma_z \cdot \eta_u(z)] - \frac{|\sigma_z|^2}{2}\frac{d^2\rho}{dz^2}(z) = \frac{|\eta_u(z) - \eta_s(z)|^2}{2} \tag{51}$$

for q numerically using a finite difference approach. Notice that left-hand side of (51) is linear in $\left(\rho, \frac{\mathsf{q}^2}{2}\right)$. We evaluate equation (51) at the n grid points for z and use a finite difference approximation for the derivatives. We write the resulting left-hand equations as a matrix times a vector containing $\frac{\mathsf{q}^2}{2}$ and ρ evaluated at $n-1$ grid points omitting $z = 0$ because we conveniently set $\rho(0) = 0$. We write the right-hand side as a vector evaluated at the n grid points and solve the resulting equation system via matrix inversion.

D Statistical calibration

We fit a trivariate VAR with the following variables:

$$\begin{aligned} &\log Y_{t+1} - \log Y_t \\ &\log G_{t+1} - \log Y_{t+1} \\ &\log D_{t+1} - \log Y_{t+1} \end{aligned}$$

where G_t is the sum of corporate profits and proprietors' income and D_t is personal income.

Provided that the VAR has stable coefficients, this is a co-integrated system. All three time series have stationary increments, but there one common martingale process. The shock to this process is identified as the only one with long-term consequences. We set $\widehat{\alpha}_z = 0$ and $\widehat{\beta}_y = 1$. For the remaining parameters we:

i) fit a VAR with a constant and four lags of the first variable and five of the other two;

ii) compute the implied mean for $\log Y_{t+1} - \log Y_t$ and set this to $\widehat{\alpha}_y$;

iii) compute the state dependent component of the expected long-term growth rate by calculating:

$$\log Y_t^p = \lim_{j\to\infty} E\left(\log Y_{t+j} - \log Y_t - j\widehat{\alpha}_y | \mathcal{F}_t\right)$$

implied by the VAR estimates, to compare to the counterpart calculation in the continuous-time model:

$$Z_t^p = \lim_{j \to \infty} E\left(\log Y_{t+j} - \log Y_t - j\widehat{\alpha} | Z_t\right) = \frac{1}{\widehat{\kappa}} Z_t.$$

iv) compute the implied autoregressive coefficient for $\{\log Y_t^p\}$ in the discrete-time specification using the VAR parameter estimates and equate this coefficient to $1 - \widehat{\kappa}$.

v) compute the VAR implied covariance matrix for the one-step-ahead forecast error for $\{\log Y^p\}$, the direct shock to consumption and equate this to

$$\begin{bmatrix} (\sigma_y)' \\ \frac{1}{\widehat{\kappa}}(\sigma_z)' \end{bmatrix} \begin{bmatrix} (\sigma_y) & \frac{1}{\widehat{\kappa}}(\sigma_z) \end{bmatrix}$$

where we achieve identification of σ_z and σ_y by imposing a zero restriction on the second entry of σ_y and positive signs on the first coefficient of σ_y and on the second coefficient of σ_z.

E Solving the ODE's

The value function is approximately linear in the state variable for large $|z|$. This gives a good Neumann boundary condition to use in an approximation in which z is restricted to a compact interval that includes $z = \bar{z}$. Recall the constraint:

$$\frac{1}{2} r' \Lambda r + \left[\rho_1 + \rho_2 (z - \bar{z})\right]\left[-\widehat{\kappa}(z - \bar{z}) + r_2\right] + \frac{|\sigma_z|^2}{2}\rho_2 - \frac{\mathsf{q}^2}{2} \leqslant 0.$$

Consider an affine solution $r = r_0 + r_1(z - \bar{z})$. The vector r_1 satisfies

$$\frac{1}{2}(r_1)' \Lambda r_1 - \rho_2 \widehat{\kappa} + \rho_2 r_{1,2} = 0$$

where $r_1 = (r_{1,1}, r_{1,2})'$ When we view this relation as a quadratic equation in $r_{1,1}$ given $r_{1,2}$, there will be two solutions. We pick the solution that makes $r_{1,1}(z - \bar{z})$ the smallest this will differ depending on whether we use a left boundary point $z^- << \bar{z}$ or a right boundary point $z^+ >> \bar{z}$.

It remains to pick the two boundary conditions for the derivative of the value function

ϕ^- and ϕ^+. From the HJB equation:

$$(-\delta - \widehat{\kappa} + r_{1,2})\phi + .01(\widehat{\beta} + r_{1,1}) = 0$$
$$\Lambda r_1 + \begin{bmatrix} 0 \\ \rho_2 \end{bmatrix} \propto \begin{bmatrix} .01 \\ \phi \end{bmatrix}.$$

The first equation is the derivative of the value function for constant coefficients, putting aside the minimization. The next is the large z approximation to the first-order conditions implied by (32). By taking ratios of the latter condition, we obtain an equation in r_1 and ϕ. Solving the resulting three equations determines $(r_{1.1}^-, r_{1,2}^-, \phi^-)$ and $(r_{1,1}^+, r_{1,2}^+, \phi^+)$, where ϕ^- and ϕ^+ are the two approximate boundary conditions for the derivative of the value function.

We used `bvp4c` in Matlab to solve the ode's over the two intervals $[-2.5, 0]$ and $[0, 2.5]$ where $\overline{z} = 0$.

References

Anderson, Evan W., Lars P. Hansen, and Thomas J. Sargent. 1998. Risk and Robustness in Equilibrium. Available on webpages.

———. 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.

Ang, Andrew and Monika Piazzesi. 2003. A No-Arbitrage Vector Autoregression of the Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics* 50:745–787.

Bansal, Ravi and Amir Yaron. 2004. Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles. *Journal of Finance* 59 (4):1481–1509.

Barillas, Francisco, Lars P. Hansen, and Thomas J. Sargent. 2009. Doubts or Variability? *Journal of Economic Theory* 144 (6):2388–2418.

Battigalli, Pierpaolo, Simone Cerreia-Vioglio, Fabio Maccheroni, and Massimo Marinacci. 2015. Self-Confirming Equilibrium and Model Uncertainty. *American Economic Review* 105 (2):646–677.

Berger, James O. 1994. An Overview of Robust Bayesian Analysis (with discussion). *Test* 3 (1):5–124.

Bhandari, Anmol. 2014. Doubts, Asymmetries, and Insurance. University of Minnesota.

Borovička, Jaroslav, Lars Peter Hansen, and José A. Scheinkman. 2014. Shock Elasticities and Impulse Response Functions. *Mathematics and Financial Economics* 8 (4):333–354.

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

Chamberlain, Gary. 2000. Econometric Applications of Maxmin Expected Utility. *Journal of Applied Econometrics* 15 (6):625–644.

Chen, Hui, Winston Wei Dou, and Leonid Kogan. 2015. Measuring the 'Dark Matter' in Asset Pricing Models. MIT Sloan School of Management.

Chen, Zengjing and Larry Epstein. 2002. Ambiguity, Risk, and Asset Returns in Continuous Time. *Econometrica* 70:1403–1443.

Chernoff, Herman. 1952. A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the Sum of Observations. *Annals of Mathematical Statistics* 23 (4):pp. 493–507.

Donsker, Monroe E. and S. R. Srinivasa Varadhan. 1976. On the Principal Eigenvalue of Second-Order Elliptic Differential Equations. *Communications in Pure and Applied Mathematics* 29:595–621.

Epstein, Larry G. and Shaolin Ji. 2014. Ambiguous Volatility, Possibility and Utility in Continuous Time. *Journal of Mathematical Economics* 50:269 – 282.

Epstein, Larry G. and Martin Schneider. 2003. Recursive Multiple-Priors. *Journal of Economic Theory* 113 (1):1–31.

Fleming, Wendell H. and Panagiotis E. Souganidis. 1989. On the Existence of Value Functions of Two Player, Zero-sum Stochastic Differential Games. *Indiana Univesity Mathematics Journal* 38:293–314.

Fudenberg, Drew and David K. Levine. 1993. Self-Confirming Equilibrium. *Econometrica* 61:523–46.

———. 2009. Self-confirming Equilibrium and the Lucas Critique. *Journal of Economic Theory* 144 (6):2354–2371.

Good, Irving J. 1952. Rational Decisions. *Journal of the Royal Statistical Society. Series B (Methodological)* 14 (1):pp. 107–114.

Hansen, Lars P. 2007. Beliefs, Doubts and Learning: Valuing Macroeconomic Risk. *American Economic Review* 97 (2):1–30.

Hansen, Lars P. and Thomas J. Sargent. 2001. Robust Control and Model Uncertainty. *American Economic Review* 91 (2):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.

———. 2010. Fragile Beliefs and the Price of Uncertainty. *Quantitative Economics* 1 (1):129–162.

Hansen, Lars P., Thomas J. Sargent, and Jr. Tallarini, Thomas D. 1999. Robust Permanent Income and Pricing. *The Review of Economic Studies* 66 (4):873–907.

Hansen, Lars P., Thomas J. Sargent, Gauhar A. Turmuhambetova, and Noah Williams. 2006. Robust Control and Model Misspecification. *Journal of Economic Theory* 128 (1):45–90.

Hansen, Lars P., John C. Heaton, and Nan Li. 2008. Consumption Strikes Back?: Measuring Long Run Risk. *Journal of Political Economy* .

Hansen, Lars Peter and Massimo Marinacci. 2016. Ambiguity Aversion and Model Misspecification: An Economic Perspective. *Statistical Science* 31 (4):511–515.

Hansen, Lars Peter and José A. Scheinkman. 2012. Pricing Growth-rate Risk. *Finance and Stochastics* 16 (1):1–15.

Ilut, Cosmin L. and Martin Schneider. 2014. Ambiguous Business Cycles. *American Economic Review* 104 (8):2368–99.

James, Matthew R. 1992. Asymptotic Analysis of Nonlinear Stochastic Risk-Sensitive Control and Differential Games. *Mathematics of Control, Signals and Systems* 5 (4):401–417.

Karantounias, Anastasios G. 2013. Managing Pessimistic Expectations and Fiscal Policy. *Theoretical Economics* 8 (1).

Lucas, Jr, Robert E. 1978. Asset Prices in an Exchange Economy. *Econometrica* 46 (6):1429–45.

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.

Miao, Jianjun and Alejandro Rivera. 2016. Robust Contracts in Continuous Time. *Econometrica* 84 (4):1405–1440.

Newman, Charles M. and Barton W. Stuck. 1979. Chernoff Bounds for Discriminating between Two Markov Processes. *Stochastics* 2 (1-4):139–153.

Peng, Shige. 2004. *Nonlinear Expectations, Nonlinear Evaluations and Risk Measures.* Stochastic Methods in Finance: Lectures given at the C.I.M.E.-E.M.S. Summer School held in Bressanone/Brixen, Italy, July 6-12, 2003. Berlin, Heidelberg: Springer Berlin Heidelberg.

Sargent, Thomas J. 1999. *The Conquest of American Inflation.* Princeton, New Jersey: Princeton University Press.

Ulrich, Maxim. 2013. Inflation Ambiguity and the Term Structure of U.S. Government Bonds. *Journal of Monetary Economics* 60 (2):295 – 309.

Watson, James and Chris Holmes. 2016. Approximate Models and Robust Decisons. *Statistical Science* 31 (4):465–489.