Prices of Macroeconomic Uncertainties*

with Tenuous Beliefs

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

December 20, 2016

Abstract

A decision maker expresses ambiguity about statistical models in the following ways. He has a family of structured parametric probability models but suspects that their parameters vary over time in unknown ways that he does not describe probabilistically. He expresses a further suspicion that all of these parametric models are misspecified by entertaining alternative unstructured probability distributions that he represents only as positive martingales and that he restricts to be statistically close to the structured parametric models. Because he is averse to ambiguity, he uses a max-min criterion to evaluate alternative plans. We characterize equilibrium uncertainty prices by confronting a decision maker with a portfolio choice problem. We offer a quantitative illustration for structured parametric models that focus uncertainty on macroeconomic growth and its persistence. There emerge nonlinearities in marginal valuations that induce time variation in market prices uncertainty. Prices of uncertainty fluctuate because the 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 assistance with estimation, Victor Zhorin for his assistance with calculating Chernoff entropy and Yiran Fan with assistance in formulating and solving the ode's for the decision problems. We thank Fernando Alvarez, Ben Brooks, Stavros Panageas, Doron Ravid, and Bálint Szőke for critical comments on an earlier draft. 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 takes a different approach by assuming that the agents inside the model experience various forms of model uncertainty. 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 worry that their parametric model is incorrect. Thus, we put the agents inside our model into what they view as a complicated setting in which outcomes are sensitive to their subjective beliefs and in which learning is very difficult. We draw liberally from literatures on decision theory, robust control theory, and the econometrics of misspecified models to build a tractable model of how decision makers' specification concerns affect equilibrium prices and quantities.

To put our approach to work in a concrete setting, 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 the decision makers inside our dynamic economic model evaluate the utility consequences of alternative specifications of state dynamics. We show how these concerns induce variation in asset values and construct a quantitative example that assigns an important role to macroeconomic growth rate uncertainty. Because it has adverse consequences for discounted expected utilities, investors in our model fear growth rate persistence in times of weak growth. In contrast, they fear the absence of persistence when macroeconomic growth is high.

We describe procedures that simplify this specification challenge both for the investor and for us as outside analysts. We model decision making by blending ideas from two seemingly distinct approaches. We start by assuming that a decision maker considers a parametric family of structured models (with either fixed or time varying parameters) using 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 residing within a statistical neighborhood of them.[1] We argue that the Chen and Epstein structure is too confining to include such statistical concerns about model misspecification. Instead 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. The 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 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 the representative investor. These representative investor shadow prices are also uncertainty prices that clear competitive security markets. The negative of an endogenously determined vector of worst-case drift distortions equals a vector of prices that compensate the representative investor for bearing model uncertainty.[3] Time variation in uncertainty prices emerges endogenously since investor 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 implausibly large risk aversions 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

[1]By "structured" we don't mean what econometricians call "structural" models in the tradition of either the Cowles commission or rational expectations. 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).

martingale representation proves to be a tractable way for us to formulate robust decision problems in sections 4, 5 and 8.

Section 6 uses Chernoff entropy, a statistical distance measure applicable to a set of martingales, to quantify difficulties in discriminating between competing specifications of probabilities. We show how to use this measure a) in the spirit of Good (1952), *ex post* to assess plausibility of worst-case models, and b) to calibrate the penalty parameter used to represent 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 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 who stands for "the market". 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 not possible 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 recommended by Epstein and Schneider (2003) will not work in our setting. Section 11 offers concluding remarks.

2 Models and perturbations

This section describes nonnegative martingales that perturb a baseline probability model. Section 3.1 then describes how we use a family of parametric alternatives to a baseline model to form a convex set of martingales 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 progressively measurable process with respect to the filtration $\mathcal{F} = \{\mathcal{F}_t : t \geqslant 0\}$ associated with the Brownian motion W augmented by any information available at date zero. Under this restriction, 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 the 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

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

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] 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. Applications typically use Markov specifications.

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

[7] Sufficient conditions for the stochastic exponential to be a martingale such as Kazamaki's or Novikov's are not convenient here. 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. Thus, 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 discrepancy

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 drift (also called a local mean)

$$\nu_t = \frac{1}{2}M_t^U|U_t|^2.$$

Write the conditional mean of $M^U \log M^U$ in terms of a history of local means[8]

$$E\left[M_t^U \log M_t^U|\mathcal{F}_0\right] = E\left(\int_0^t \nu_\tau d\tau|\mathcal{F}_0\right)$$

[8]In this paper, we simply impose the first equality. There exists a variety of sufficient conditions that justify this equality.

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

To formulate a decision problem that chooses probabilities to minimize expected utility, we will use the representation after the second equality without imposing that M^U is a martingale and then verify that the solution is indeed a martingale. Hansen et al. (2006) justify this approach.[9]

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

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

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

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

The second line is the limit of Abel integral averages, where scaling by δ makes the weights $\delta \exp(-\delta\tau)$ integrate to one. 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:

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

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

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

where baseline probabilities are represented here by a degenerate martingale that is identi-

[9]See their Claims 6.1 and 6.2.

cally one. Formula (8) quantifies how a martingale M^U distorts baseline model probabilities. Following Hansen and Sargent (2001), we call $\Delta(M^U; 1|\mathcal{F})$ discounted entropy relative to a probability represented by the baseline martingale.

In contrast to Hansen and Sargent (2001) and Hansen et al. (2006) who start from a unique baseline model, we start from a convex set $M^S \in \mathcal{M}^o$ of *structured* models represented as martingales with respect to such a baseline model. We shall describe how we form $\mathcal{M}^o$ in subsection 3.1. These structured models 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 express the idea that a decision maker wants to investigate the utility consequences of all models that are statistically close to those in $\mathcal{M}^o$. The scaling parameter θ measures how heavily we will penalize an expected utility-minimizing agent for distorting probabilities.

3.1 Constructing 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 (10) to construct a set of probability models.[10]

Restriction (10) imposes a recursive structure on the decision maker's preferences that allows using dynamic programming to solve Markov decision problems. That is one manifestation of the fact that these preferences satisfy a dynamic consistency property axiomatized by Epstein and Schneider (2003). The following example provides a specification of Ξ in (10) that encompasses our application with uncertainty in macroeconomic growth to be described in section 5. Later in section 9, we revisit restriction (10) and discuss its implica-

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

tions for applications not explored in this paper. As we shall see, in contrast to Chen and Epstein (2002), we use constraint (10) only as an intermediate step and ultimately construct a larger set of statistically similar unstructured models whose utility consequences we want to study.

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 can generate the following set of time-invariant parameter models:*

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

Here the unknown parameter vector $\pi = \begin{bmatrix} \pi^1 & \pi^2 & \ldots & \pi^n \end{bmatrix}' \in \Pi$*, a closed convex subset of* $\mathbb{R}^n$*. We can enlarge this set to include time-varying parameter models:*

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

where the unknown 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 the realizations of* $\widetilde{\pi}_t$ *are restricted to be in* Π*, this same restriction applies to their conditional expectations, and thus*

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

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

Our application employs a special case of constraint (13). In section 4.2 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 this second construction to guide our specification of the set Π of potential parameter values.

3.2 Misspecification of structured models

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, we employ 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 prohibits exploring 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 preference that Maccheroni et al. (2006b) have shown to be dynamically consistent.

To understand how our formulation relates to dynamic variational preferences, notice that structured models represented in terms of their drift distortion processes S_t enter 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 (10) leads to a conditional discrepancy

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

and an associated integrated discounted discrepancy

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

We want a decision maker to care also about the utility consequences of statistically close unstructured models that we describe in terms of the discrepancy measure

$$\Theta\left(M^U|\mathcal{F}_0\right). \tag{14}$$

For any hypothetical state- and date-contingent plan – consumption in the example of section 5 – we follow Hansen and Sargent (2001) in minimizing a discounted expected utility

[11] See Watson and Holmes (2016) and Hansen and Marinacci (2016) for recent discussions of longstanding misspecification challenges confronted by statisticians and economists.

function plus the θ-scaled relative entropy penalty $\Theta(M^U|\mathcal{F}_0)$ over the set of models. The following remark asserts that this procedure induces a dynamically consistent preference ordering over decision processes.

Remark 3.2. *Define what Maccheroni et al. (2006b) call an ambiguity index process:*

$$\Theta_t\left(M^U\right) = \frac{\theta\delta}{2}\int_0^\infty \exp(-\delta\tau) E\left[\left(\frac{M^U_{t+\tau}}{M^U_t}\right)\xi_{t+\tau}(U_{t+\tau})\Big|\mathcal{F}_t\right] d\tau.$$

This process solves the following continuous-time counterpart to equations (11) and (12) of Maccheroni et al.:[12]

$$0 = -\delta\Theta_t\left(M^U\right) + \frac{\theta\delta}{2}\xi_t(U_t).$$

4 Robust planning problem

To illustrate how a robust planner evaluates utility consequences of unstructured models that our relative entropy measure says are difficult to distinguish, we deliberately consider a simple setup with an exogenous consumption process. We deduce shadow prices of uncertainty from martingales that generate worst-case probabilities in a continuous time planning problem.[13]

To construct a set of models, the planner:

1) Begins with a 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).

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

[12]The term $\frac{\theta\delta}{2}\xi_t(U_t)$ in our analysis is γ_t in Maccheroni et al. (2006b). In Hansen and Sargent (2001) and Hansen et al. (2006), their $\gamma_t = \frac{\theta}{2}|U_t|^2$ where their 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.

[13]Richer models would include production, capital accumulation, and distinct classes of decision makers with differential access to financial markets. Before adding such features, we want to understand uncertainty in our simple environment. In doing this, we follow a tradition extending back to Lucas (1978) and Mehra and Prescott (1985).

For step 1, we use the diffusion (1) as a baseline model. Step 3 includes statistically similar models. We will describe two approaches for step 2.

4.1 Revisiting example 3.1

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

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

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 initial 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 add structured models of the form (13), so we represent an initial set of time invariant parameter models in terms of

$$s(x) = \sum_{j=1}^{n} \pi^j \eta^j(x), \;\; \pi \in \Pi, \tag{15}$$

where Π is a convex set of possible parameters that can evolve over time. These configurations can be fixed or change over time. Without restrictions on the prior across the parameter configurations and how they change over time, we are led to consider mixtures in which

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

represents alternative posterior probabilities that can be assigned to parameter configuration in place at a given date.

We depict preferences with an instantaneous utility function $\delta\nu(x)$ and a subjective discount rate δ. Where m is a realized value of a martingale, a value function $mV(x)$ that satisfies the following HJB equation determines a worst-case model:

$$0 = \min_{u,s} -\delta m V(x) + m\nu(x) + m\widehat{\mu}(x) \cdot \frac{\partial V}{\partial x}(x) + m[\sigma(x)u] \cdot \frac{\partial V}{\partial x}(x)$$

$$+\frac{m}{2}\text{trace}\left[\sigma(x)'\frac{\partial^2 V}{\partial x \partial x'}(x)\sigma(x)\right]+\frac{\theta m}{2}|u-s|^2 \tag{16}$$

where minimization over u, s is subject to (15). Here s represents structured models in $\mathcal{M}^o$ and u represents unstructured models that are statistically similar to models in $\mathcal{M}^o$. Because m multiplies all terms on the right side of equation (16), it can be omitted.

The problem on the right side of HJB equation (16) 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

$$u-s=-\frac{1}{\theta}\sigma(x)'\frac{\partial V}{\partial x}(x). \tag{17}$$

Substituting from (17) into HJB equation (16) gives the reduced HJB equation in:

Problem 4.1.

$$\begin{aligned} 0=\min_s & -\delta V(x)+\nu(x)+\widehat{\mu}(x)\cdot\frac{\partial V}{\partial x}(x)+[\sigma(x)s]\cdot\frac{\partial V}{\partial x}(x) \\ & +\frac{1}{2}\text{trace}\left[\sigma(x)'\frac{\partial^2 V}{\partial x\partial x'}(x)\sigma(x)\right]-\frac{1}{2\theta}\left[\frac{\partial V}{\partial x}(x)\right]'\sigma(x)\sigma(x)'\left[\frac{\partial V}{\partial x}(x)\right] \end{aligned} \tag{18}$$

where minimization is subject to (15)*. Given the minimizing $s^*(x)$, we recover the minimizing u from $u^*(x)=s^*(x)-\frac{1}{\theta}\sigma(x)'\frac{\partial V^*}{\partial x}(x)$ where V^* solves HJB equation* (18)*.*

This problem generates two minimizers, namely, s and u. The minimizing s is a structured drift taking the form $s^*(x)=\sum_{j=1}^n \pi^{j*}(x)s^j(x)$ that evidently depends on the state x. The associated minimizing u is a worst-case drift distortion $u^*(x)$ relative to the worst-case structured model 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 satisfying appropriate boundary conditions and as well as conditions that render a verification theorem applicable. In one of the examples that follows, the first derivative of the value function has a kink, but the value function is still a viscosity solution.

We write problem 4.1 as a single agent problem because, for reasons of pedagogical simplicity, the application in this paper is confined to a setting in which the aggregate consumption process is exogenous in the tradition of Lucas (1978). More generally, the planner could solve a resource allocation problem that involves accumulating physical capital and other factors of production and that makes consumption endogenous. Shadow prices for that problem, including prices of uncertainty, would also be prices in a competitive equilibrium. In such settings, a counterpart to problem 4.1 would be posed as a two-player, zero-sum stochastic differential game of a type studied by Fleming and Souganidis (1989).

4.2 Restricting relative entropy

In this section, we describe how, instead of forming a set of structured model according to equation (15), we use another method to construct this set. In particular, we use a restriction on their relative entropies to form a set of structured models. We will use this approach in our quantitative application below. For special cases, including our application, the two approaches coincide.

Recall from section 3 that relative entropy for a stochastic process conditioned on date 0 information is:

$$\varepsilon(M^U) = \lim_{\delta \downarrow 0} \frac{\delta}{2} \int_0^\infty \exp(-\delta t) E\left(M_t^U |U_t|^2 \middle| \mathcal{F}_0 \right) dt.$$

We continue to use the process M^U to change the probability distribution conditioned on $\mathcal{F}_0$. If we can initialize the probability to make the process for X be stationary and ergodic, then the limit on the right side is the unconditional expectation of $\frac{1}{2}|U_t|^2$ under this stationary distribution. Moreover, the mathematical expectation of discounted relative entropy under the stationary distribution implied by M^U equals $\varepsilon(M^U)$ and does not depend on δ. Hence, relative entropy is simply one half the expectation of $|U_t|^2$ under this measure.

Consider a structured model for which $S_t = \eta(X_t)$. We depict relative entropy as the solution to an HJB equation derived by taking a small δ limit of discounted relative entropy, namely,

$$\frac{\mathsf{q}^2}{2} = \frac{|\eta(x)|^2}{2} + \frac{\partial \rho}{\partial x}(x) \cdot [\widehat{\mu}(x) + \sigma(x)\eta(x)] + \frac{1}{2}\text{trace}\left[\sigma(x)' \frac{\partial^2 \rho}{\partial x \partial x'}(x)\sigma(x)\right] \tag{19}$$

where the function ρ is determined only up to a constant and

$$\frac{\mathsf{q}^2}{2} = \varepsilon(M^S).$$

Imposing a constraint on relative entropy by pre-specifying only q produces a family of probabilities that fails to satisfy (10).[14] Therefore, we instead specify $\frac{\partial \rho}{\partial x}$ and a number q,

[14] Moreover, embedding this set in one that is rectangular would yield too large a set in a sense described

then use them to restrict S_t to satisfy a weakened version of (19):

$$\frac{|S_t|^2}{2} + \frac{\partial \rho}{\partial x}(X_t) \cdot [\widehat{\mu}(X_t) + \sigma(X_t) S_t] + \frac{1}{2}\text{trace}\left[\sigma(X_t)' \frac{\partial^2 \rho}{\partial x \partial x'}(X_t)\sigma(X_t)\right] - \frac{\mathsf{q}^2}{2} \leqslant 0. \qquad (20)$$

Inequality (20) is a state-dependent restriction on S_t. That it is quadratic in S_t, makes it tractable to implement.

One way to construct $\frac{\partial \rho}{\partial x}$ and q is to posit an alternative drift configuration $\eta(x)$ and solve (19). Other models also satisfy inequality (20) for the same $\frac{\partial \rho}{\partial x}$ and q. These are sometimes amenable to characterization as we will illustrate. Alternatively, we can specify directly q and impose *a priori* restrictions on $\frac{\partial \rho}{\partial x}$. An extreme example is to impose that $\frac{\partial \rho}{\partial x}(x) = 0$, which is equivalent to restricting:

$$|S_t| \leqslant \mathsf{q}.$$

Our application will lead us naturally to consider state-dependent (in fact quadratic) specifications of ρ, but we will impose other *a priori* restrictions on the function ρ.

We again depict preferences with an instantaneous utility function $\delta\nu(x)$ and a subjective discount rate δ. The robust planner problem that replaces problem 4.1 has HJB equation

Problem 4.2.

$$\begin{aligned} 0 = \min_s & -\delta V(x) + \nu(x) + \widehat{\mu}(x) \cdot \frac{\partial V}{\partial x}(x) + [\sigma(x)s] \cdot \frac{\partial V}{\partial x}(x) \\ & + \frac{1}{2}\text{trace}\left[\sigma(x)' \frac{\partial^2 V}{\partial x \partial x'}(x)\sigma(x)\right] - \frac{1}{2\theta}\left[\frac{\partial V}{\partial x}(x)\right]' \sigma(x)\sigma(x)' \left[\frac{\partial V}{\partial x}(x)\right] \end{aligned} \qquad (21)$$

where minimization over s is subject to

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

We can recover a minimizing u from a minimizing $s^*(x)$ via $u^*(x) = s^*(x) - \frac{1}{\theta}\sigma(x)' \frac{\partial V^*}{\partial x}(x)$, where V^* solves HJB equation (21).

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

in section 10.

5 Uncertainty about Macroeconomic Growth

To prepare the way for the quantitative illustration in section 7, this section applies our setup within a particular macro-finance setting. We start with a baseline parametric model, then form a family of parametric structured probability models for a representative investor's consumption process. We deduce the pertinent version of HJB equation (16) that describes the value function attained by worst-case drift distortions S and U. The baseline model is

$$\begin{aligned} dY_t &= .01\left(\widehat{\alpha}_y + \widehat{\beta} Z_t\right) dt + .01\sigma_y \cdot dW_t \\ dZ_t &= \left(\widehat{\alpha}_z - \widehat{\kappa} Z_t\right) dt + \sigma_z \cdot dW_t. \end{aligned} \tag{22}$$

We scale by .01 because Y is typically expressed in logarithms and we want to work with growth rates. 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 $\widehat{\alpha}_z - \widehat{\kappa} z = -\widehat{\kappa}(z - \bar{z})$. The term $\bar{z}$ is also the mean of Z_t in the stationary distribution under the baseline model.

We focus on the following collection of structured parametric models:

$$\begin{aligned} dY_t &= .01\left(\alpha_y + \beta Z_t\right) dt + .01\sigma_y \cdot dW_t^S \\ dZ_t &= \left(\alpha_z - \kappa Z_t\right) dt + \sigma_z \cdot dW_t^S, \end{aligned} \tag{23}$$

where W^S is a Brownian motion and (6) continues to describe the relationship between the processes W and W^S. By design, this collection nests the baseline model (22) Here $(\alpha_y, \beta, \alpha_z, \kappa)$ are parameters distinguishing structured models (23) from the baseline model, and (σ_y, σ_z) are parameters common to models (22) and (23).

We represent members of a parametric class defined by (23) 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 (23) to deduce the following restrictions on η_1:

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

where

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

Recall that relative entropy $\frac{\mathsf{q}^2}{2}$ emerges as part of the solution to HJB equation (19) appropriately specialized:

$$\frac{|\eta(z)|^2}{2} + \frac{d\rho}{dz}(z)[-\widehat{\kappa}(z - \bar{z}) + \sigma_z \cdot \eta(z)] + \frac{|\sigma_z|^2}{2}\frac{d^2\rho}{dz^2}(z) - \frac{\mathsf{q}^2}{2} = 0. \tag{24}$$

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

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

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

$$\frac{|S_t|^2}{2} + [\rho_1 + \rho_2(Z_t - \bar{z})]\left[-\widehat{\kappa}(Z_t - \bar{z}) + \sigma_z \cdot S_t\right] + \frac{|\sigma_z|^2}{2}\rho_2 - \frac{\mathsf{q}^2}{2} \leqslant 0. \tag{25}$$

Suppose that $Y = \log C$, where C is consumption, δ is a subjective rate of discount and instantaneous utility $\nu(x) = \delta y$. Let $r = \sigma s$. For this problem, we seek a value function $V(x) = y + \widehat{V}(z)$ The HJB equation used by the robust planner is

$$\begin{aligned} 0 = \min_r & -\delta \widehat{V}(z) + .01[\widehat{\alpha}_y + \widehat{\beta} z + r_1] + [-\widehat{\kappa}(z - \bar{z}) + r_2]\frac{d\widehat{V}}{dz}(z) \\ & + \frac{1}{2}|\sigma_z|^2 \frac{d^2\widehat{V}}{dz^2}(z) - \frac{1}{2\theta}\begin{bmatrix} .01 & \frac{d\widehat{V}}{dz}(z) \end{bmatrix} \sigma\sigma' \begin{bmatrix} .01 \\ \frac{d\widehat{V}}{dz}(z) \end{bmatrix} \end{aligned} \tag{26}$$

where the minimization is subject to

$$\frac{1}{2}r'\Lambda r + [\rho_1 + \rho_2(z - \bar{z})]\left[-\widehat{\kappa}(z - \bar{z}) + r_2\right] + \frac{|\sigma_z|^2}{2}\rho_2 - \frac{\mathsf{q}^2}{2} \leqslant 0 \tag{27}$$

and $\Lambda = (\sigma')^{-1}\sigma^{-1}$. A worst-case structured model induces a worst-case unstructured model via equation (17). (In the portfolio problem of section 8, we will also maximize over portfolio weights and the consumption process C.)

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

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

That this component is linear pushes the solution to the boundary of the convex constraint set, an ellipsoid, for each z. Figure 5 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 $\frac{d\widehat{V}}{dz}(z) > 0$, which will be 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 $\frac{dr_1}{dr_2} = -100\frac{d\widehat{V}}{dz}(z)$ and depends on z, as figure 5 indicates.

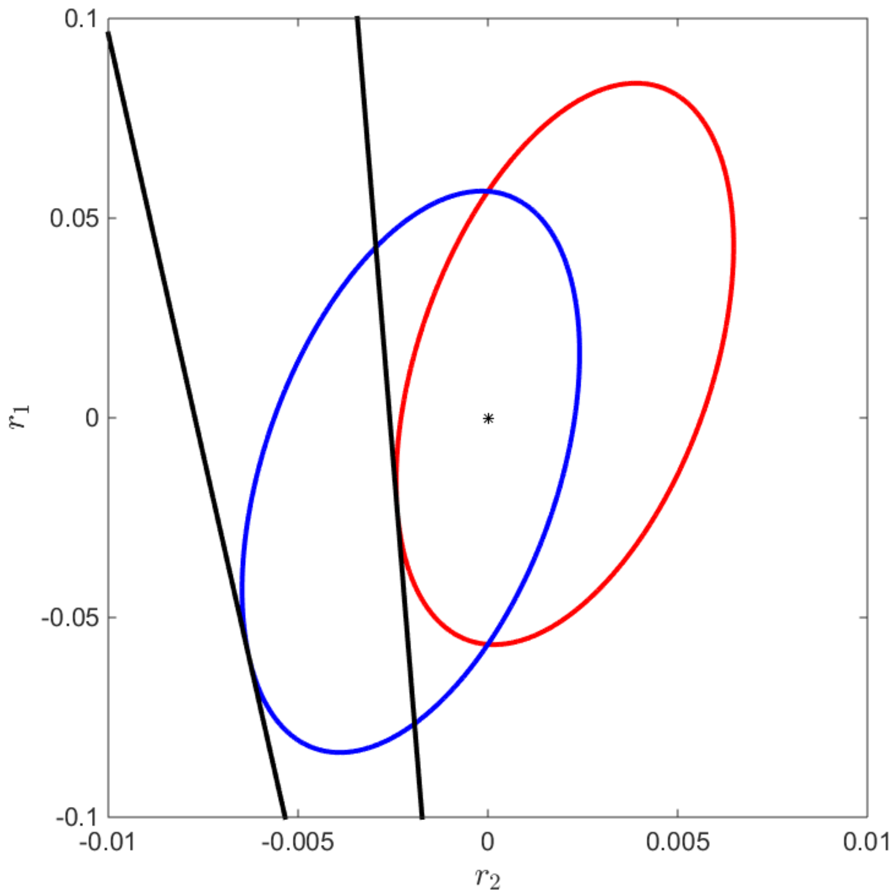

Figure 1: Parameter contours for (r_1, r_2) holding relative entropy fixed. The the upper right contour depicted in red is for z equal to the .1 quantile of the stationary distribution under the baseline model and the lower left contour is for z at the .9 quantile. The dot depicts the $(0, 0)$ point corresponding to the baseline model. The tangency points denote worst-case structured models. This figure assumes that $\mathsf{q} = .1$ and a Chernoff half life of 60.

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 could solve the equation (24) 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 giving us a one dimensional curve for η_0 and η_1. As we will see, this restriction allows for nonlinear specifications of S_t, and such models will emerge endogenously from our analysis. These nonlinear restrictions will also have relative entropy $\frac{\mathsf{q}^2}{2}$. We can link the resulting nonlinear model back to 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. By using this construction, our relative entropy restriction (25) 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 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 (25) *implies that*

$$S_t = 0$$

when $Z_t = \bar{z}$ *with this specification. 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) *for the alternative choices of* η_1*. This illustrates how imposing this* Π *in the HJB equation of problem 4.1 is equivalent to imposing the restricted version* ρ *in the HJB equation of problem 4.2.*

6 Chernoff entropy

Good (1952) suggests that to evaluate a max-min expected utility approach one should verify that the associated worst-case model is plausible.[15] We implement that suggestion by using entropy to measure how far a worst-case model is from a set of structured models, then applying Good's idea to help us calibrate the penalty parameter θ in HJB equation (16).

Chernoff entropy emerges from studying how, by disguising probability distortions of a baseline 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. In this section, we characterize Chernoff entropy by extending a construction of Chernoff (1952). In the spirit of Anderson et al. (2003), we use Chernoff (1952) entropy to measure a distortion M^H to a baseline model. Anderson et al. (2003) use Chernoff entropy measured as a local rate to draw direct connections between magnitudes of market prices of uncertainty and statistical discrimination. This local rate is state dependent and for diffusion models proportional to the local drift in relative entropy. Important distinctions arise when we measure statistical discrepancy globally as did Newman and Stuck (1979). In this section, we characterize the global version of Chernoff entropy and show how to compute it.

6.1 Bounding mistake probabilities

Think of a pairwise model selection problem that statistically compares the baseline model (1) with a model generated by a martingale M^U whose logarithm evolves according to

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

Consider a statistical model selection rule based on a data history of length t that takes the form $\log M_t^U \geqslant \mathsf{h}$, where M_t^U is the likelihood ratio associated with the alternative model for a sample size t. This selection rule might incorrectly choose the alternative model when the baseline model governs the data. We can bound the probability of this outcome by using an argument from large deviations theory that starts from

$$\mathbf{1}_{\left\{\log M_t^U \geqslant \mathsf{h}\right\}} = \mathbf{1}_{\left\{-\gamma \mathsf{h} + \gamma \log M_t^U \geqslant 0\right\}} = \mathbf{1}_{\left\{\exp(-\gamma \mathsf{h})(M_t^U)^\gamma \geqslant 1\right\}} \leqslant \exp(-\gamma \mathsf{h})(M_t^U)^\gamma.$$

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

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

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

for alternative choices of γ. The threshold h does not affect this limit. Furthermore, the limit is often independent of the initial conditioning information. To get the best bound, we compute

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

a limit supremum that is typically negative because mistake probabilities decay with sample size. Chernoff entropy is then

$$\chi(M^U) = -\inf_{0\leqslant\gamma\leqslant 1} \limsup_{t\to\infty} \frac{1}{t} \log E\left[\left(M_t^U\right)^\gamma |\mathcal{F}_0\right]. \tag{28}$$

Setting $\chi(M^U) = 0$ would mean including only alternative models that cannot be distinguished on the basis of histories of infinite length. In effect, that is what is done in papers that extend the rational expectations equilibrium concept to self-confirming equilibria associated with probability models that are wrong off equilibrium paths, i.e., for events that do not occur infinitely often.[16] Because we want to include alternative parametric probability models, we entertain positive values of $\chi(M^U)$. Our decision theory differs from that typically used for self confirming equilibria because our decision maker acknowledges model uncertainty and wants to adjust decisions accordingly.

To interpret $\chi(M^U)$, consider the following argument. If the decay rate of mistake probabilities were constant, say $\overline{\chi}$, then mistake probabilities for two sample sizes $T_i, i = 1, 2$, would be

$$\text{mistake probability}_i = \frac{1}{2} \exp\left(-T_i \overline{\chi}\right)$$

for $\overline{\chi} = \chi(M^U)$. We define a 'half-life' as an increase in sample size $T_2 - T_1 > 0$ that

[16] See Sargent (1999) and Fudenberg and Levine (2009).

multiplies a mistake probability by a factor of one half:

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

so the half-life is approximately

$$T_2 - T_1 = \frac{\log 2}{\overline{\chi}}. \tag{29}$$

The preceding back-of-the-envelope calculation justifies the detection error bound computed by Anderson et al. (2003). The bound on the decay rate should be interpreted cautiously because it is constant although the actual decay rate is not. Furthermore, the pairwise comparison oversimplifies the true challenge, which is statistically to discriminate among *multiple* models.

We can make a symmetrical calculation that reverses the roles of the two models and instead conditions on the perturbed model implied by martingale M^U. It is straightforward to show that 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.

6.2 Using Chernoff entropy

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

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

for a given $\gamma \in (0,1)$ as a dominant eigenvalue for a semigroup of linear operators. When this approach is appropriate, 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. Given the restrictions on γ, since M^U is a martingale, $\left(M^U\right)^\gamma$ is a super martingale and its expectation typically decays to zero at an asymptotically exponential rate.

See Appendix C for a discussion of the numerical method we used to evaluate Chernoff entropy for nonlinear Markov specifications in our forthcoming applications.

7 Quantitative example

Our quantitative example builds on the setup of 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 difficult to detect and troublesome predictable components of consumption growth.[17] Relative entropy and Chernoff entropy shape and quantify the doubts that we impute to investors.

7.1 Baseline model

Our example blends parts 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 (22) that approximates responses of consumption to permanent shocks. In contrast to Bansal and Yaron (2004), we introduce no stochastic volatility because we want to focus exclusively on fluctuations in uncertainty prices that are induced by the representative investor's specification concerns.

In constructing a VAR, we follow Hansen et al. (2008) by using additional macroeconomic time series to infer information about long-term consumption growth. We report a calibration of our baseline model (22) deduced 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 cointegrated. Business income is measured as proprietor's income plus corporate profits per capita. Dividends are personal dividend income per capita.[18] We fit a trivariate vector autoregression that imposes cointegration among these three series. Since we presume that all three time series grow, the coefficients in the cointegrating relation are known. In Appendix D we tell how we used the discrete time VAR estimates to deduce the following parameters for the

[17] 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.

[18] 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.

baseline model (22):

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

$$\begin{aligned} \sigma_y &= \begin{bmatrix} .488 \\ 0 \end{bmatrix} \\ \sigma_z &= \begin{bmatrix} .013 \\ .028 \end{bmatrix} \end{aligned} \tag{30}$$

We suppose that $\delta = .002$ and $\nu = y$, 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 Benchmark models and a robust plan

We solve HJB equation (26) for three different configurations of structured models.

7.2.1 Uncertain growth rate responses

We first compute the solution by first focusing on a specification described Example 5.1 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 or implicitly on (β, κ). Figure 2 plots iso-entropy contours for (β, κ) for $\mathsf{q} = .1$ and $\mathsf{q} = .05$.

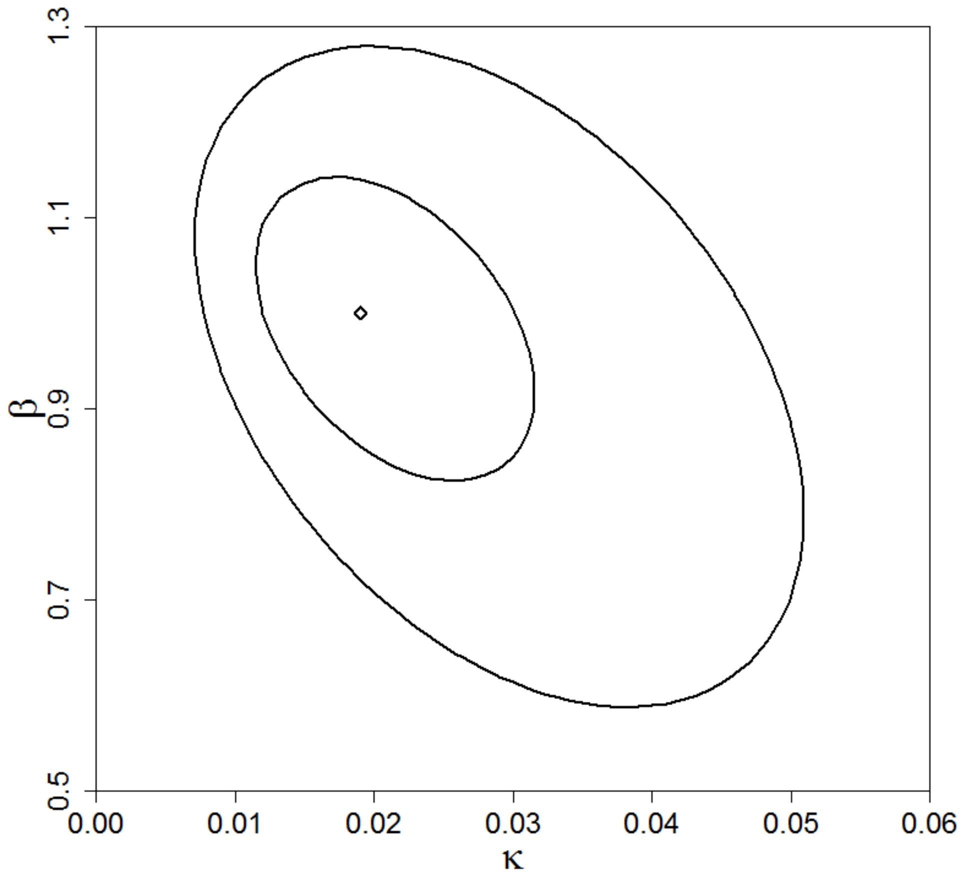

Figure 2: Parameter contours for (β, κ) holding relative entropy fixed. The outer curve $\mathsf{q} = .1$ and the inner curve $\mathsf{q} = .05$. The small diamond depicts the baseline model.

While Figure 2 looks at 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 parameter uncertainty for (β, κ), the set of possible r's differs depending on the state z. As we remarked earlier, the only feasible r when $z = \bar{z}$ is $r = 0$. Figure 1 also reports the 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 have slopes equal to $-100\frac{d\widehat{V}}{dz}$ where the point of tangency is the worst-case structured model. This point occurs at lower drift distortion for the .9 quantile than for the .1 quantile.

Consider next the adjustment for model misspecification. Since

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

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.

The first derivative of the value function under this restriction on structured models is not differentiable at $z = \bar{z}$. We compute the value function and the worst-case models by solving two coupled HJB equations, one for $z < \bar{z}$ and another for $z > \bar{z}$. In effect 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.

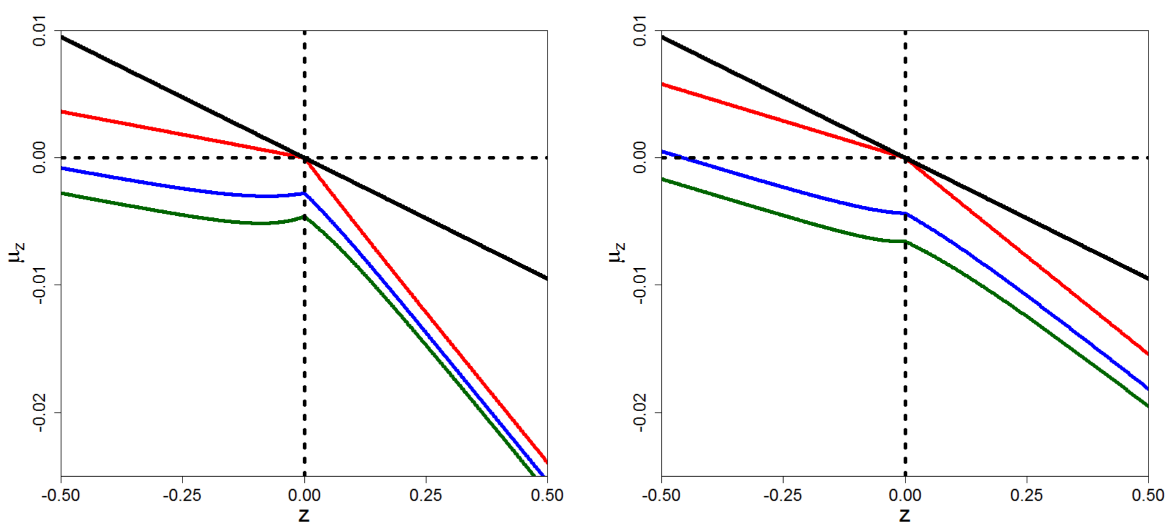

Figure 3: Distorted growth rate drifts. Left panel: larger structured entropy (q = .1). Right panel: smaller structured entropy (q = .05). **red**: worst-case structured model; **blue**: half life 120; and **green**: half life 60.

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 must be 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. Activating a concern for misspecification of the structured models by letting θ be finite shifts the drift as a function of the state downwards, even more so for negative values of z than positive ones.

The impact of the drift for Y is much more modest.

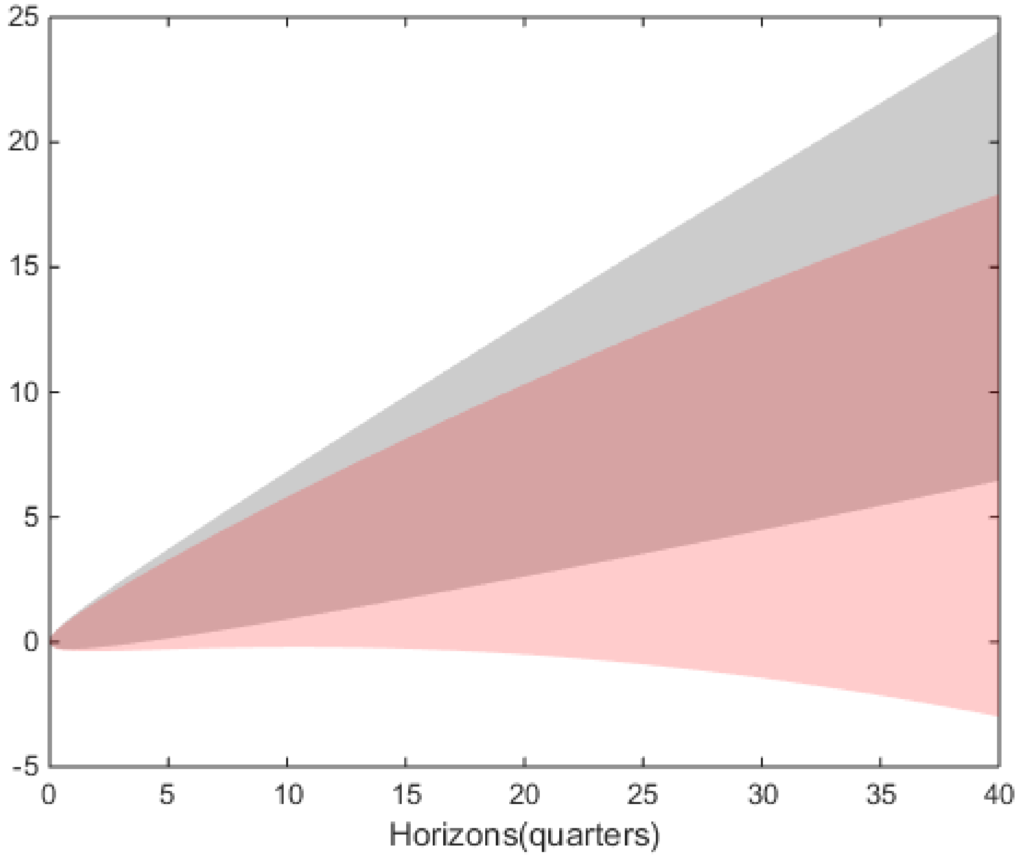

Figure 4: Distribution of $Y_t - Y_0$ under the baseline model and worst-case model for $\mathsf{q} = .1$ and a Chernoff half life of 60 quarters. The gray shaded area depicts the interval between the .1 and .9 deciles for every choice of the horizon under the baseline model. The red shaded area gives the 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 the uncertainty

Previously we restricted

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

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

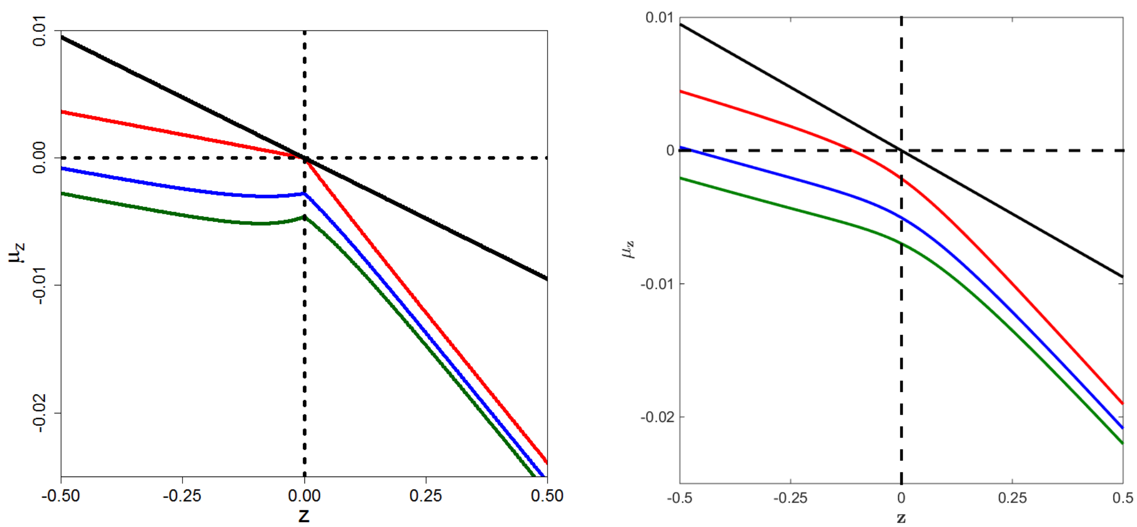

Figure 5: Distorted growth rate drift for Z. Relative entropy $\mathsf{q} = .1$. Left panel: $\rho_2 = \frac{(.01)}{|\sigma_z|^2}$. Right panel: $\rho_2 = \frac{(.01)}{2|\sigma_z|^2}$. **red**: worst-case structured model; **blue**: half life 120; and **green**: half life 60.

We consider the consequences of adding the restriction that $\rho_2 = 0$. In this case, the value function for the robust planner is linear and the minimizing s and u are constant (independent of z). Specifically,

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

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 $|s^*| = \mathsf{q}$. Thus zeroing out ρ_1 and ρ_2 results in worst-case drifts for both Y and Z that are downward parallel shifts of the baseline drifts. It is equivalent to changing the coefficients α_y and α_z in a way that is time invariant and leaving $\kappa = \widehat{\kappa}$ and $\beta = \widehat{\beta}$.

7.3 A bigger set

We now replace (20) with its "on average counterpart." Let

$$\zeta_t(S_t) = \frac{|S_t|^2}{2} + [\rho_1 + \rho_2(Z_t - \bar{z})][-\widehat{\kappa}(Z_t - \bar{z}) + \sigma_z \cdot S_t] + \frac{\rho_2}{2}|\sigma_z|^2 - \frac{\mathsf{q}^2}{2}.$$

We restrict:

$$\lim_{\delta \downarrow 0} \delta E \left[\int_0^\infty \exp(-\delta t) M_t^S \zeta_t(S_t) dt | \mathcal{F}_0 \right] \leqslant 0. \tag{31}$$

Notice that by Ito's Lemma, at date t the drift of $\rho(Z)$ under the $S = \eta(Z)$ implied evolution is:

$$[\rho_1 + \rho_2(Z_t - \bar{z})][-\widehat{\kappa}(Z_t - \bar{z}) + \sigma_z \cdot S_t] + \frac{\rho_2}{2}|\sigma_z|^2$$

Under stationarity and an additional technical restriction, this drift has mean zero.[19] Therefore, under restriction (31)

$$\lim_{\delta \downarrow 0} \delta E \left(\int_0^\infty \exp(-\delta t) M_t^S \left[\frac{|S_t|^2}{2} \right] | \mathcal{F}_0 \right) \leqslant \frac{\mathsf{q}^2}{2}. \tag{32}$$

The second set of martingales that we want to explore is

$$\mathcal{M}^2 = \left\{ M^S \in \mathcal{M}^o : S \text{ satisfies inequality (32)} \right\},$$

where we no longer prespecify a $\frac{d\rho}{dz}$ that we use to restrict S. We instead prespecify only q.

7.4 Commitment to a worst-case structured model

Partly to make contact with an alternative formulation proposed by Hansen and Sargent (2016), we now alter timing protocols. Instead of (S_t, U_t) being chosen simultaneously at

[19] A sufficient condition entails requiring that ρ be in the domain of the generator for the associated Markov process for an appropriate function space. See Ethier and Kurtz (1986) and Hansen and Scheinkman (1995). This particular argument was suggested to us by Yiran Fan.

each instant as depicted in HJB equation (16), a decision maker now confronts a single structured model that has been chosen by a statistician in charge of choosing our S process. The decision maker chooses an U process because he does not trust the worst-case structured model chosen by "the statistician" .

Until now, in choosing (S, U) our decision maker has used discounted relative entropy. In choosing a worst-case structured model, the statistician uses the $\delta \downarrow 0$ measure and imposes the restriction (31) at a given q. Relative entropy neighborhoods of interior martingales are included in this set. Taking the $\delta \downarrow 0$ limit eliminates the dependence on conditioning information for a convex set of martingales M^S.

We start with a date-zero perspective. The statistician uses the same instantaneous utility function as the decision maker and takes a process of instantaneous utilities as given. The statistician then uses a martingale relative to the baseline model to construct a structured model by solving a continuous-time analogue of a control problem posed by Petersen et al. (2000).

The decision maker discounts at rate $\delta > 0$ and accepts the statistician's model as a structured. Because he doubts it he makes a robustness adjustment of the type suggested by Hansen and Sargent (2001). This problem is reminiscent of Brunnermeier and Parker (2005), who formulate two-agent decision problems in which one agent chooses beliefs using an undiscounted utility function while the other agent takes those beliefs as fixed when evaluating alternative plans. Theirs, however, is not intended to be a model of concerns about robustness.

An equilibrium for our robust planner game is particularly easy to compute because the instantaneous utility is specified *a priori*. This allows us to solve the statistician problem first and after that the decision maker's problem.[20]

7.4.1 Statistician Problem

We solve the statistician problem first taking as given a Lagrange multiplier ℓ. Since we study a limiting version of this problem in which δ declines to zero, we separate two objects. The first is the limiting objective function that will be constant and that we denote by $\overline{\varsigma}(\ell)$. The second is the function used to characterize the worst-case model for the statistician, which we denote by $\varsigma(z, \ell)$ and which is determined only up to an additive constant. The

[20]In a model with production, this two-step approach would no long apply.

statistician's value functions $(\bar{\varsigma}, \varsigma)$ solves:

$$\bar{\varsigma}(\ell) = \min_s .01 \left(\widehat{\alpha}_y + \widehat{\beta} z + \sigma_y \cdot s\right) + \frac{d\varsigma}{dz}(z,\ell)\left[-\widehat{\kappa}(z - \bar{z}) + \sigma_z \cdot s\right]$$
$$+ \frac{1}{2}|\sigma_z|^2 \frac{d^2\varsigma}{dz^2}(z,\ell) + \ell\left(\frac{1}{2} s \cdot s - \frac{\mathsf{q}^2}{2}\right)$$

This problem is a limiting version of a constraint robust control problem of a type suggested by Hansen and Sargent (2001); it is a constraint counterpart of an infinite horizon control problem investigated by Fleming and McEneaney (1995). The function ς is linear in z as a function of the multiplier:

$$\varsigma(z,\ell) = \varsigma_1(z - \bar{z}) = .01\left(\frac{\widehat{\beta}}{\widehat{\kappa}}\right)(z - \bar{z}),$$

and the undediscounted objective $\bar{\varsigma}(\ell)$ is:

$$\bar{\varsigma}(\ell) = .01\left(\widehat{\alpha}_y + \widehat{\beta}\bar{z}\right) - \frac{(.01)^2}{\ell}\begin{bmatrix} 1 & \left(\frac{\widehat{\beta}}{\widehat{\kappa}}\right)\end{bmatrix} \sigma'\sigma \begin{bmatrix} 1 \\ \left(\frac{\widehat{\beta}}{\widehat{\kappa}}\right)\end{bmatrix} - \frac{\ell}{2}\mathsf{q}^2.$$

We set the multiplier ℓ to satisfy the relative entropy constraint (31) by maximizing:

$$\max_\ell \bar{\varsigma}(\ell).$$

The implied drift adjustment used to represent the statistician's structured model is

$$s^* = -\frac{.01}{\ell^*}\sigma'\begin{bmatrix} 1 \\ \frac{\widehat{\beta}}{\widehat{\kappa}}\end{bmatrix},$$

Since ς is linear in z, the statistician's worst-case structured model alters the probability distribution for W by adding a constant to the drift vector for the composite X process. This leads us to express the local dynamics for the structured model as:

$$\widehat{\alpha}_y + \widehat{\beta} z + \sigma_y \cdot s^* = \alpha_y^* + \widehat{\beta} z$$
$$\widehat{\alpha}_z - \widehat{\kappa} z + \sigma_z \cdot s^* = \alpha_z^* - \widehat{\kappa} z.$$

Evidently, here the worst-case structured model remains within our parametric class. Note that this structured model is not a time-varying or state-dependent coefficient model, in contrast to the situation found under the distinct section 7.2 setting.

7.4.2 Robust Control Problem

At date zero, the decision maker accepts the statistician's model as a structured, but because he doubts it, he makes a robustness adjustment of the type suggested by Hansen and Sargent (2001). Where the value function is $m[y + \psi(z)]$, write the decision maker's HJB equation as[21]

$$\begin{aligned}
0 = &\min_g -\delta\psi(z) + .01\sigma_y \cdot g + \sigma_z \cdot g \frac{d\psi}{dz}(z) + \frac{\theta}{2} g \cdot g \\
&+ .01\left(\alpha_y^* + \beta^* z\right) + \frac{d\psi}{dz}(z)\left(\alpha_z^* - \kappa^* z\right) + \frac{1}{2}|\sigma_z|^2 \frac{d^2\psi}{dz^2}(z) \\
= &-\delta\psi(z) + .01\left(\alpha_y^* + \beta^* z\right) + \frac{d\psi}{dz}(z)\left(\alpha_z^* - \kappa^* z\right) + \frac{1}{2}|\sigma_z|^2 \frac{d^2\psi}{dz^2}(z) \\
&- \frac{1}{2\theta}\begin{bmatrix} .01 & \frac{d\psi}{dz}(z) \end{bmatrix} \sigma\sigma' \begin{bmatrix} .01 \\ \frac{d\psi}{dz}(z) \end{bmatrix}.
\end{aligned}$$

where we think of $g = u - s$. In this case we can say more. It is straightforward to show that $\psi(z) = \psi_0 + \psi_1(z - \bar{z})$, where in particular ψ_1 solves:

$$0 = -\delta\psi_1 + .01\widehat{\beta} - \psi_1\widehat{\kappa}$$

which implies

$$\psi_1 = (.01)\frac{\widehat{\beta}}{\delta + \widehat{\kappa}}.$$

7.4.3 Results

The following worst-case model for the decision maker emerges from our statistician-decision maker game:

$$u^* = \eta^*(z) = s^* - \frac{1}{\theta}\sigma' \begin{bmatrix} .01 \\ \frac{d\psi}{dz}(z) \end{bmatrix} = -\frac{.01}{\ell^*}\sigma' \begin{bmatrix} 1 \\ \frac{\widehat{\beta}}{\widehat{\kappa}} \end{bmatrix} - \frac{.01}{\theta}\sigma' \begin{bmatrix} 1 \\ \frac{\widehat{\beta}}{\delta + \widehat{\kappa}} \end{bmatrix}$$

[21] As in earlier decision problems, we can omit m from the HJB equation because it multiplies all terms.

In light of how we have formulated this game, it is not surprising that the second term takes a form found by Hansen et al. (2006). Dynamic consistency prevails in the sense that if we ask the players to re-assess their choices at some date $t > 0$, each player would remain content with its original choice. We study values of θ chosen to match Chernoff half lives of 60 quarters and 120 quarters, respectively.

In this formulation, the minimizing drift distortion no longer shows state dependence. We would obtain very similar results by removing the statistician's problem of choosing a worst-case structured model from consideration and focusing on the robust control problem only. The implication of a constant distortion for both components is no doubt special and dependent on the modeling assumptions that we have made. Our aim has been simply to illustrate the possible impact of adding some commitment to the analysis. In this simple example, doing that removes state dependence of the worst-case drift.

8 Robust portfolio choice and pricing

In this section, we describe equilibrium prices that make a representative investor willing to bear risks accurately described by baseline model (1) in spite of his concerns about model misspecification. We construct equilibrium prices by appropriately extracting 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 faces a continuous-time Merton portfolio problem in which individual wealth K evolves as

$$dK_t = -C_t dt + K_t \iota(Z_t) dt + K_t A_t \cdot dW_t + K_t \pi(Z_t) \cdot A_t dt, \tag{33}$$

where $A_t = a$ is a vector of chosen risk exposures, $\iota(x)$ is the instantaneous risk free rate expressed, and $\pi(z)$ is the 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 (33), the vector of risk prices $\pi(z)$, and the quadratic function ρ and relative entropy $\frac{\mathsf{q}^2}{2}$ that define the alternative structured models that concern the representative investor.

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

$$\begin{aligned} 0 = \max_{a,c} \min_{u,s} & -\delta m\widetilde{V}(z) - \delta m \log k - \delta m \log \delta + \delta m \log c - \frac{mc}{k} + m\iota(z) \\ & + m\pi(z) \cdot a + ma \cdot u - \frac{m|a|^2}{2} + m\frac{d\widetilde{V}}{dz}(z)\left[-\widehat{\kappa}(z - \bar{z}) + \sigma_z \cdot u\right] \\ & + \frac{m}{2}|\sigma_z|^2 \frac{d^2\widetilde{V}}{dz^2}(z) + \left(\frac{m\theta}{2}\right)|u - s|^2 \end{aligned} \tag{34}$$

subject to

$$\frac{|s|^2}{2} + \frac{d\rho}{dz}(z)[-\widehat{\kappa}(z - \bar{z}) + \sigma_z \cdot s] + \frac{|\sigma_z|^2}{2}\frac{d^2\rho}{dz^2}(z) - \frac{\mathsf{q}^2}{2} = 0. \tag{35}$$

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. First-order conditions for a and u are

$$\pi(z) + u^* - a^* = 0 \tag{36a}$$

$$a^* + \theta(u^* - s^*) + \frac{d\widetilde{V}}{dz}(z)\sigma_z = 0. \tag{36b}$$

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

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

subject to (35). 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 pertinent two-person zero-sum game.[22]

[22]If we were to use a timing protocol that allows the maximizing player to take account of the impact of its decisions on the minimizing agent, we would obtain the same equilibrium decision rules described in the text.

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 that a competitive equilibrium awards for bearing model uncertainty. To compute a vector $\pi(x)$ 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 (36b) and the solution for s^*,

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

where η^* is the worst-case drift from the robust planning problem provided that we show that $\widetilde{V} = \widehat{V}$, where $\widehat{V}$ is the value function for the robust planning problem. Thus, from (36a), $\pi = \pi^*$, where

$$\pi^*(z) = (.01)\sigma_y - \eta^*(z). \tag{37}$$

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

$$\iota^*(z) = \delta + .01(\widehat{\alpha}_y + \widehat{\beta} z) + .01\sigma_y \cdot \eta^*(z) - \frac{.0001}{2}\sigma_y \cdot \sigma_y. \tag{38}$$

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 $\widetilde{\upsilon} = \upsilon$.

8.3 Local uncertainty prices

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

$$d\log Sdf_t = -\delta dt - .01\left(\widehat{\alpha}_y + \widehat{\beta} Z_t\right) dt - .01\sigma_y \cdot dW_t + U_t^* \cdot dW_t - \frac{1}{2}|U_t^*|^2 dt. \tag{39}$$

The components of the vector $\pi^*(Z_t)$ given by (37) equal minus the local exposures to the Brownian shocks. These are usually interpreted as local "risk prices," but we shall reinterpret them. Motivated by the decomposition

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

we prefer 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. This will change in important ways over longer horizons.

8.4 Uncertainty prices over alternative investment horizons

In the context of our quantitative models, we now report the shock-price elasticities that Borovička et al. (2014) showed are horizon-dependent uncertainty prices of risk exposures. Shock price elasticities describe the dependence of logarithms of expected returns on an investment horizon. The logarithm of the expected return from a consumption payoff at date t consists of two terms:

$$\log E\left(\frac{C_t}{C_0}\middle| X_0 = x\right) - \log E\left[Sdf_t\left(\frac{Y_t}{Y_0}\right)\middle| X_0 = x\right]. \tag{40}$$

where $\log C_t = Y_t$. The first term is the expected payoff and the second is the cost of purchasing that payoff. Notice that in our example we imposed a unitary elasticity of substitution:

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

so the second term features the martingale computed to implement robustness.

To compute an elasticity, we change locally the exposure of consumption to the underlying Brownian motion compute the consequences for the expected return. From a mathematical perspective, an important inputs into this calculation are Malliavin derivatives. These derivatives measure how a shock at given date effects the consumption and the stochastic discount factor processes. Both Sd_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 on Sd_t and C_t. Computing the derivative of the logarithm of the expected return given in (40) 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 the two dimensional vectors of Malliavin derivatives (with respect to the two dimensional Brownian increment at date t for consumption and the worst-case martingale.

Using a convenient formula familiar from other forms of differentiation

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

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

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

and thus

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

Similarly,

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

Therefore, the term structure of prices that interests us is given by

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

The first term is the familiar risk price for consumption-based asset pricing. It is state independent and contributes a (small) term that is independent of the horizon. In contrast, in the second term the equilibrium drift distortion provides state dependent component; its expectation under the distorted probability measure gives a time- and state-dependent contribution to the term structure of uncertainty prices.[23]

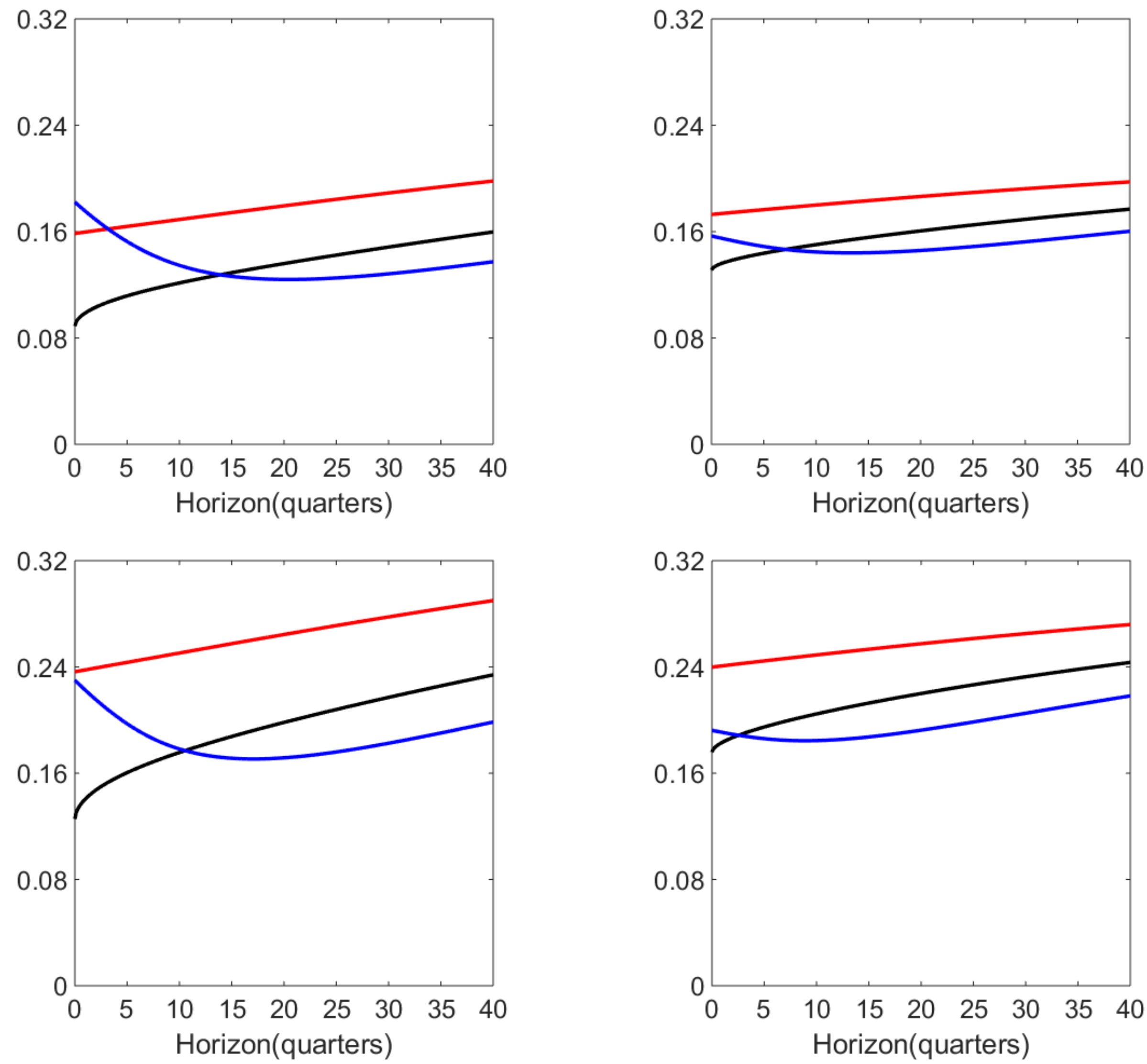

Figure 6: Shock price elasticities for alternative horizons and deciles for the specification with (β, κ) uncertainty. Left panels: larger baseline entropy (q = .1). Right panels: smaller structured entropy (q = .05). Top panels: first shock. Bottom panels: second shock. **Black**: median of the Z stationary distribution **red**: .1 decile; and **blue**: .9 decile.

Notice that although the price elasticity is initially smaller for the median specifica-

[23]There are other horizon dependent elasticities that we could compute. For instance, we might look at the impact of a shock at date zero on C_t and M_t^{U*} and trace out the impact of changing the horizon but keeping the date of the shock fixed.

tion of z than for the .9 quantile, this inequality is eventually reversed as the horizon is increased. (The blue curve and black curve cross.) The uncertainty price for positive z initially diminishes because the probability measure implied by the martingale has reduced persistence for the positive states. Under this probability, the growth rate state variable is expected to spends less time positive region. This is reflected in the smaller prices for the .9 quantile than for 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. This horizon dependence is thus an important aspect of how concerns about misspecification and ambiguity aversion influence valuations of assets.

We have designed our quantitative examples to investigate a particular mechanism for generating fluctuations in uncertainty prices from statistically plausible amounts of uncertainty. 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 a consequence, we do not expect to track closely the high frequency movements in financial markets. 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 models that concerns our decision maker so vast and some of the 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 for the decision maker to learn some things, but ambiguity would still add a source of variation to valuations. Even if we were to begin with a family of time invariant models, for at least two reasons, confining $\mathcal{M}^o$ to time-invariant parameter models would be too restrictive. One is that time invariance excludes learning from new information. Another is that the passage of time alters what a decision maker

[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.

cares about.

The following example illustrates how 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 of 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 date zero information equals unity. The law of motion for 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

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

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 set of priors (i.e., a robust Bayesian) would want to evaluate the utility consequences of sets 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 priors. 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 examine the 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. The set of probabilities implied by martingales in $\mathcal{M}^o$ defined in (10) satisfies this property. 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. Because we use martingales in $\mathcal{M}$ to represent alternative probabilities, the time separability of specification (10) implies a rectangular family of probabilities.[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 the 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 will typically not be 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

$$\Xi_t = \left\{ \pi_t^1 S_t^1 + (1 - \pi_t^1) S_t^2, \ \underline{\pi}_t^1 \leqslant \pi_t \leqslant \overline{\pi}_t^1, \ \pi_t \text{ is } \mathcal{F}_t \text{ measurable.} \right\} \tag{41}$$

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

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

$\{S_t : t \geqslant 0\}$, we allow different date zero priors at each calendar date t. Doing that intertemporally disconnects restrictions on local transition probabilities.[26]

As we have illustrated with our calculations, the representative decision maker's set of structured models can have an important impact on how uncertainty prices change with the state of the macroeconomy. While we have assumed that the set of models of interest to the decision maker in our quantitative application makes learning particularly difficult, had we instead restricted that set of models enough, learning would be possible. For time invariant parameter models with unknown parameters, endowing the decision maker with a family of conjugate priors could make it tractable to construct a rectangular set of models recursively by repeatedly applying Bayes law.

The failure of Epstein and Schneider's procedure to yield a unique prior capable of justifying their dynamically consistent preference ordering creates a tension with the useful concept called admissibility that is widely applied in statistical decision theory. An admissible decision rule cannot be dominated under all possible probability specifications entertained by the decision maker. Verifying optimality against a unique worst-case model 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 can render the resulting decision rule inadmissible. Good (1952)'s recommendation for assessing max-min decision making is then unworkable.[27]

10 Relative entropy and rectangularity

We use an entropy-penalty approach to express the decision maker's conern about what we can think of as "local" forms of misspecification that take the form of alternative structured models. But our decision maker's concern that all structured models might be misspecified leads him to want explore the utility consequences of unstructured probability models that are not rectangular even though they are statistically close, as measured by relative entropy.

An alternative approach would be first to construct a set that includes relative entropy

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

[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. However, 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.

neighborhoods of all martingales in $\mathcal{M}^o$. For instance, we could start with a set

$$\overline{\mathcal{M}} = \left\{ M^U \in \mathcal{M} : \Theta(M^U | \mathcal{F}_0) < \epsilon \right\} \tag{42}$$

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

To show this, it suffices to look at relative entropy neighborhood 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 that date and that satisfies

$$E\left(|\overline{U}_\tau|^2 \mid \mathcal{F}_0 \right) < \infty. \tag{43}$$

Form a stochastic process

$$U_t^h = \begin{cases} 0 & 0 \leqslant t < \tau \\ \overline{U}_\tau & \tau \leqslant t < \tau + h \\ 0 & t \geqslant \tau + h. \end{cases} \tag{44}$$

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 u 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* (42) *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

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

the smallest rectangular set that embeds the set described by (42). That set is too big to pose a meaningful robust decision problem.

11 Concluding remarks

This paper formulates and applies a tractable model of the effects on equilibrium prices of exposures to macroeconomic uncertainties. Our analysis uses 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 dismissing 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 design in light of available empirical evidence.

Specifically, we have produced a model of the log stochastic discount factor whose uncertainty prices reflect a robust planner's worst-case drift distortions U^*. 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 history-dependent 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.

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, including some with substantial growth rate dependence and others with little such dependence. The investor uses observations on consumption growth to update a Bayesian prior 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 their 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 the 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 different way, namely, by specifying a set of alternative models to capture concerns about the baseline model's specification of persistence in consumption growth.

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. With a rectangular augmentation of a set of probabilities, the implied worst case structured model will typically not be the outcome from applying Bayes' rule to a single prior. That prevents following Good's advice in 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 broader 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 had interesting 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:

$$\frac{|s - \tilde{s}(x)|^2}{2} \leqslant \frac{\Upsilon(x)}{2}. \tag{45}$$

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

Next we construct an implicit parametrization for our example in which

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

For this $\rho(x)$, $\tilde{s}$ is affine in $z - \bar{z}$ and Υ is quadratic:

$$\Upsilon(z) = \Upsilon_0 + 2\Upsilon_1(z - \bar{z}) + \Upsilon_2(z - \bar{z})^2.$$

Since $\Upsilon(z) \geqslant 0$ for all z, $\Upsilon_0 \geqslant 0$ and $\Upsilon_2 \geqslant 0$. Moreover

$$|\Upsilon_1| \leqslant \sqrt{\Upsilon_0 \Upsilon_2}.$$

Write

$$\tilde{\eta}(z) = \eta(z) - \tilde{s}(z) = \tilde{\eta}_0 + \tilde{\eta}_1(z - \bar{z}).$$

We seek a family of η's or equivalently an $\tilde{\eta}$ that satisfy constraint (45) with equality for all z. We obtain this by solving:

$$\begin{aligned} \tilde{\eta}_0 \cdot \tilde{\eta}_0 &= \Upsilon_0 \\ \tilde{\eta}_0 \cdot \tilde{\eta}_1 &= \Upsilon_1 \\ \tilde{\eta}_1 \cdot \tilde{\eta}_1 &= \Upsilon_2. \end{aligned} \tag{46}$$

The first and third equations define ellipsoids, each with one free parameter. Suppose first that either Υ_0 or Υ_2 is zero but not both. Then Υ_1 is also zero. Either the first or the third equation gives rise to a nondegenerate ellipse in $\mathbb{R}^2$ representable as a closed curve.

Suppose next that both Υ_0 or Υ_2 are strictly positive and that the first and third equations define non degenerate ellipsoids that restrict $\tilde{\eta}_0$ and $\tilde{\eta}_1$, respectively. For any given solution $\tilde{\eta}_0$ on its respective ellipsoid, the inner product with the alternative restricted $\tilde{\eta}_1$'s have inner products that fill out the interval:

$$\left[-\sqrt{\Upsilon_1 \Upsilon_2}, \sqrt{\Upsilon_1 \Upsilon_2}\right].$$

Thus for every choice of $\tilde{\eta}_0$ on the ellipsoid defined by the first equation in (46), we can find an $\tilde{\eta}_2$ that satisfies the other two equations. This gives us a one parameter family of solutions representable as a closed curve in $\mathbb{R}^4$. Moreover, using

$$s - \tilde{s}(z) = \tilde{\eta}_0 + \tilde{\eta}_1(z - \bar{z}),$$

this construction also defines a closed curve of $(s - \tilde{s})$'s holding z fixed for appropriately restricted choices of $\tilde{\eta}_0$ and $\tilde{\eta}_1$. Call this curve one.

Now consider the constraint set (45) for this special example. The $(s - \tilde{s}(z))$'s that satisfy this constraint with equality for a given z form an ellipsoid. Call this curve two. All points on the constructed closed curve satisfy constraint (45) with equality. Since curve two is an ellipsoid, it is a simple closed curve, meaning a closed curve that does not cross itself. Thus, curve one necessarily traces out the entire ellipsoid. As a consequence, we can view the implied restricted choices of $\tilde{\eta}_0$ and $\tilde{\eta}_1$ as a one-dimensional parameterization of the structured models.

B Computing relative entropy

In this appendix we show how to compute relative entropies for parametric models of the form (23). Recall that relative entropy $\frac{\mathsf{q}^2}{2}$ emerges as part of the solution to HJB equation (19) 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

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

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 entropy

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

$$U_t = \eta(Z_t),$$

these alternative models are Markovian.

Given the 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 $\left(M_t^H\right)^\gamma g(Z_t)$ for $0 \leqslant \gamma \leqslant 1$ at $t = 0$:

$$\begin{aligned}[\mathbb{G}(\gamma)g](z) \doteq & \frac{(-\gamma + \gamma^2)}{2}|\eta(z)|^2 g(z) + \gamma g(z)'\sigma \cdot \eta(z) \\ & - g'(z)\kappa z + \frac{g''(z)}{2}|\sigma|^2,\end{aligned}$$

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

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

whose eigenfunction $e(z,\gamma)$ is the exponential of a quadratic function of z. We compute Chernoff entropy numerically by solving:

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

We solve for $\lambda(\gamma)$ numerically using a finite-difference approach. With this approach, we form an $n \times n$ matrix. the largest eigenvalue of this matrix gives $-\eta(\gamma)$ for a prespecified γ. The matrix is formed by using two-sided symmetric approximations except at the edges where we use corresponding one-sided derivatives. In our calculations we used a grid for z

over the interval $[-2.5, 2.5]$ with grid increments equal to .01. Thus $n = 501$.

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 boundary 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:

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

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

$$\frac{1}{2}d'\Lambda d - \rho_2\widehat{\kappa} + \rho_2 d_2 = 0.$$

When we view this relation as a quadratic equation in d_1 given d_2, there will be two solutions. We pick the solution that makes $d_1(z-\bar{z})$ the smallest which will differ depending 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:

$$(-\delta - \widehat{\kappa} + d_2)\phi + .01(\widehat{\beta} + d_1) = 0$$
$$\Lambda d + \begin{bmatrix} 0 \\ \rho_2 \end{bmatrix} \propto \begin{bmatrix} .01 \\ \phi \end{bmatrix}.$$

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 (26). By taking ratios of this latter condition, we obtain an equation in d and ϕ. Solving the resulting three equations determines (d_1^-, d_2^-, ϕ^-) and (d_1^+, d_2^+, ϕ^+) 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 $\bar{z} = 0$.

References

Anderson, Evan W., Lars Peter 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.

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.

Brunnermeier, Markus K. and Jonathan A. Parker. 2005. Optimal Expectations. *American Economic Review* 95 (4):1092–1118.

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.

Ethier, Stewart N. and Thomas G. Kurtz. 1986. *Markov processes : characterization and convergence*. Wiley series in probability and mathematical statistics. New York, Chichester: J. Wiley & Sons.

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

Fleming, Wendell H. and William M. McEneaney. 1995. Risk-Sensitive Control on an Infinite Time Horizon. *SIAM Journal on Control and Optimization* 33 (6):1881–1915.

Fudenberg, Drew and David K. Levine. 2009. Self-confirming Equilibrium and the Lucas Critique. *Journal of Economic Theory* 144 (6):2354–2371.

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

Hansen, L. P., J. C. Heaton, and N. Li. 2008. Consumption Strikes Back?: Measuring Long Run Risk. *Journal of Political Economy* .

Hansen, Lars P. and Thomas J. Sargent. 2016. Sets of Models and Prices of Uncertainty. NBER Working Papers 22000, National Bureau of Economic Research, Inc.

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

Hansen, Lars Peter and Massimo Marinacci. 2016. Ambiguity Aversion and Model Misspecification: An Economic Perspective. *Statistical Science* Forthcoming.

Hansen, Lars Peter and Thomas Sargent. 2010. Fragile beliefs and the price of uncertainty. *Quantitative Economics* 1 (1):129–162.

Hansen, Lars Peter 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.

Hansen, Lars Peter and José Alexandre Scheinkman. 1995. Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes. *Econometrica* 63 (4):767–804.

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

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.

Mehra, Rajnish and Edward C. Prescott. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15 (2):145–161.

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

Newman, C. M. and B. 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.

Petersen, I.R., M.R. James, and P. Dupuis. 2000. Minimax optimal control of stochastic uncertain systems with relative entropy constraints. *Automatic Control, IEEE Transactions on* 45 (3):398–412.

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

Watson, James and Chris Holmes. 2016. Approximate Models and Robust Decisons. *Statistical Science* Forthcoming.